Supplementary Materials for

The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle

Virginia M. Howick, Andrew J. C. Russell, Tallulah Andrews, Haynes Heaton, Adam J. Reid, Kedar Natarajan, Hellen Butungi, Tom Metcalf, Lisa H. Verzier, Julian C. Rayner, Matthew Berriman, Jeremy K. Herren, Oliver Billker, Martin Hemberg, Arthur M. Talman, Mara K. N. Lawniczak*

*Corresponding author. Email: mara@sanger.ac.uk

Published 23 August 2019, *Science* **365**, eaaw2619 (2019)
DOI: 10.1126/science.aaw2619

This PDF file includes:

Other supplementary material for this manuscript includes the following:

Data S1 to S4 (Excel format)

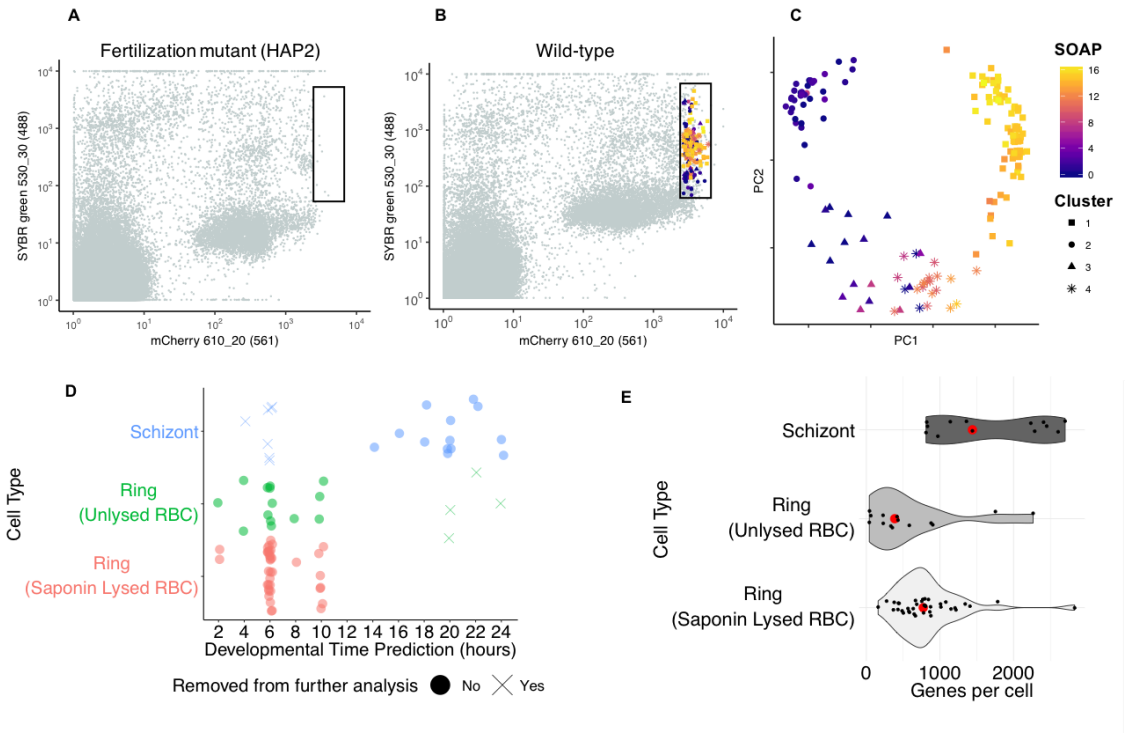

Fig. S1. Isolation of ookinete and ring-stage parasites.

(A-C): FACS gating for ookinete isolation from the blood bolus 18 and 24 h after feeding was determined using an mCherry expressing HAP2 mutant as a negative control. Parasites were expressing mCherry and stained with SYBR green. **(A)** FACS plot of the blood bolus contents from the HAP2 (PBANKA_1212600) mutant control. **(B)** FACS plot of the wild type blood bolus. Cells that were sequenced are coloured by expression of the ookinete marker gene SOAP (PBANKA_1037800) (*63*) and shaped by the SC3 cluster assignment. **(C)** A PCA of profiled cells coloured by expression of SOAP and shaped by SC3 cluster. Clusters 1 and 4 were maintained in the data set and likely represent mature ookinetes (cluster 1) and developing retorts (cluster 4). Cluster 3 may correspond to non-replicated zygotes: cells are expressing Pbs25 (PBANKA_0515000) (*64*), but not SOAP. Cluster 2 may correspond to unfertilised parasites as there was inconsistent expression of canonical marker genes and generally lower levels of SYBR green. **(D-E):** To evaluate the protocol for saponin-mediated lysis of ring-stage infected red blood cells, a 96-well Smart-seq2 plate was sorted consisting of ¼ schizonts, ¼ unlysed rings, and ½ lysed rings. **(D)** The lysed and unlysed cells were collected from the same mouse controlling for life cycle stage. However, in order to confirm that we had sorted rings rather than early trophozoites, we used Spearman's rank-order correlation to correlate each cell to published bulk time course data (*22*). The maximum r value for each cell was then chosen and the cell was assigned that time point. The jitter plot shows the predicted life cycle stage for each sorted sample, confirming similar predicted stages among lysed and unlysed rings. Data points are coloured according to the sorted population they belong to. Cells were then filtered

according to their life cycle stage prediction, so as to remove cells that clearly did not belong to that cell type. Specifically, cells from the schizont culture that were predicted to be <10 h and cells from the ring culture that were predicted to be >15 h were removed. Cells that were removed from further analysis are shown on the plot. **(E)** A violin plot showing the distribution of genes per cell per stage as well as the recovery rate for each stage. Lysing the red blood cell of ring-stage parasites increased the recovery rate and yielded a similar median number of genes per cell (red point) when compared to rings from un-lysed red blood cells.

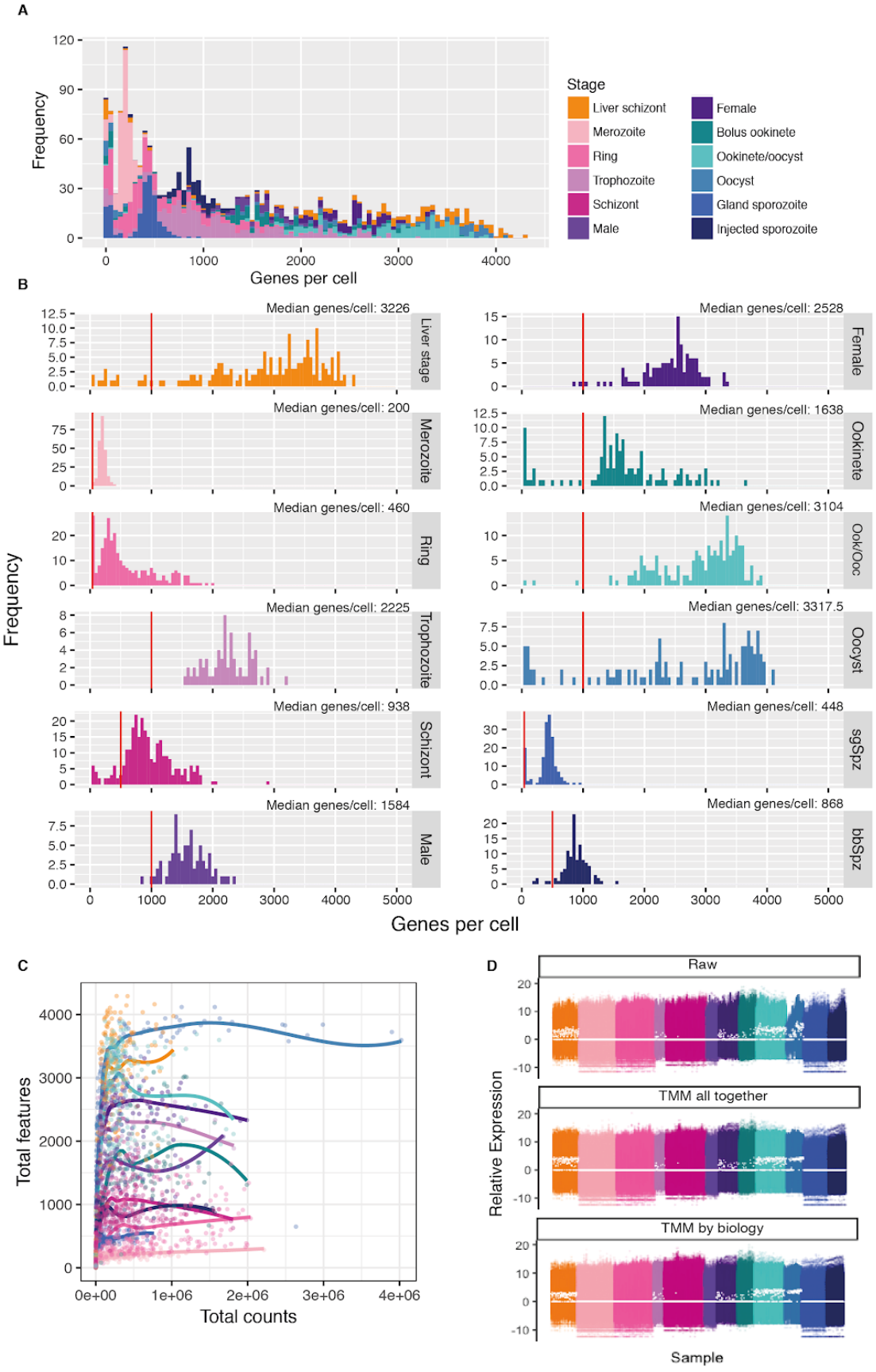
A
Frequency
Genes per cell
Stage
Liver schizont
Merozoite
Ring
Trophozoite
Schizont
Male
Female
Bolus ookinete
Ookinete/oocyst
Oocyst
Gland sporozoite
Injected sporozoite
B
Frequency
Genes per cell
Median genes/cell: 3226
Liver stage
Median genes/cell: 200
Merozoite
Median genes/cell: 460
Ring
Median genes/cell: 2225
Trophozoite
Median genes/cell: 938
Schizont
Median genes/cell: 1584
Male
Median genes/cell: 2528
Female
Median genes/cell: 1638
Ookinete
Median genes/cell: 3104
Ook/Ooc
Median genes/cell: 3317.5
Oocyst
Median genes/cell: 448
sgSpz
Median genes/cell: 868
bbSpz
C
Total features
Total counts
D
Raw
TMM all together
TMM by biology
Relative Expression
Sample

Fig. S2. Quality control and normalization.

(A) The number of genes detected per cell was dependent on life stage ($p<0.001$, ANOVA: Genes detected ~ parasite stage). **(B)** Poor quality cells were identified on a per stage basis based on the distribution of the number of genes detected per cell (cells to the left of each red line on histograms were excluded). The maximum number of genes per cell was found in oocysts, at day 4 post mosquito infection, with a median of 3318 genes per cell, while the minimum number of genes detected was found in merozoites with a median of 200 genes per cell. This is consistent with smaller cells having less total mRNA (*65*), and likely reflects biological differences in the mRNA content of each life stage. Interestingly, we detect almost twice the number of genes in sporozoites taken directly from injected saliva (n genes=868) relative to sporozoites that were isolated from salivary glands (n genes=448), suggesting a potential upregulation of gene expression at this transition. The numbers and proportions of cells that passed QC (based on reads and genes per cell) in each isolation method are reported in table S1. **(C)** Total counts per cell vs. total genes detected in all Smart-seq2 cells. Each stage is fit with a loess function. Sequencing saturation was likely reached because deeper sequencing did not result in an increase in the number of genes detected, irrespective of life stage. **(D)** Relative Expression plots for all cells across all genes. The top panel shows the relative expression of the raw counts data. The second panel shows the relative expression whereby cells have been TMM-normalised as one set. The bottom panel shows the same plot but for cells that have been subsetted into their respective groups of stages: Liver schizonts, IDC, gametocytes, ookinetes/oocysts, sporozoites, normalised by TMM, and then re-pooled. Based on the plots, normalisation method does not have a large effect on the relative expression; cell expression patterns have been smoothed in both methods. Unless otherwise stated, the grouped normalised data were used for further analyses.

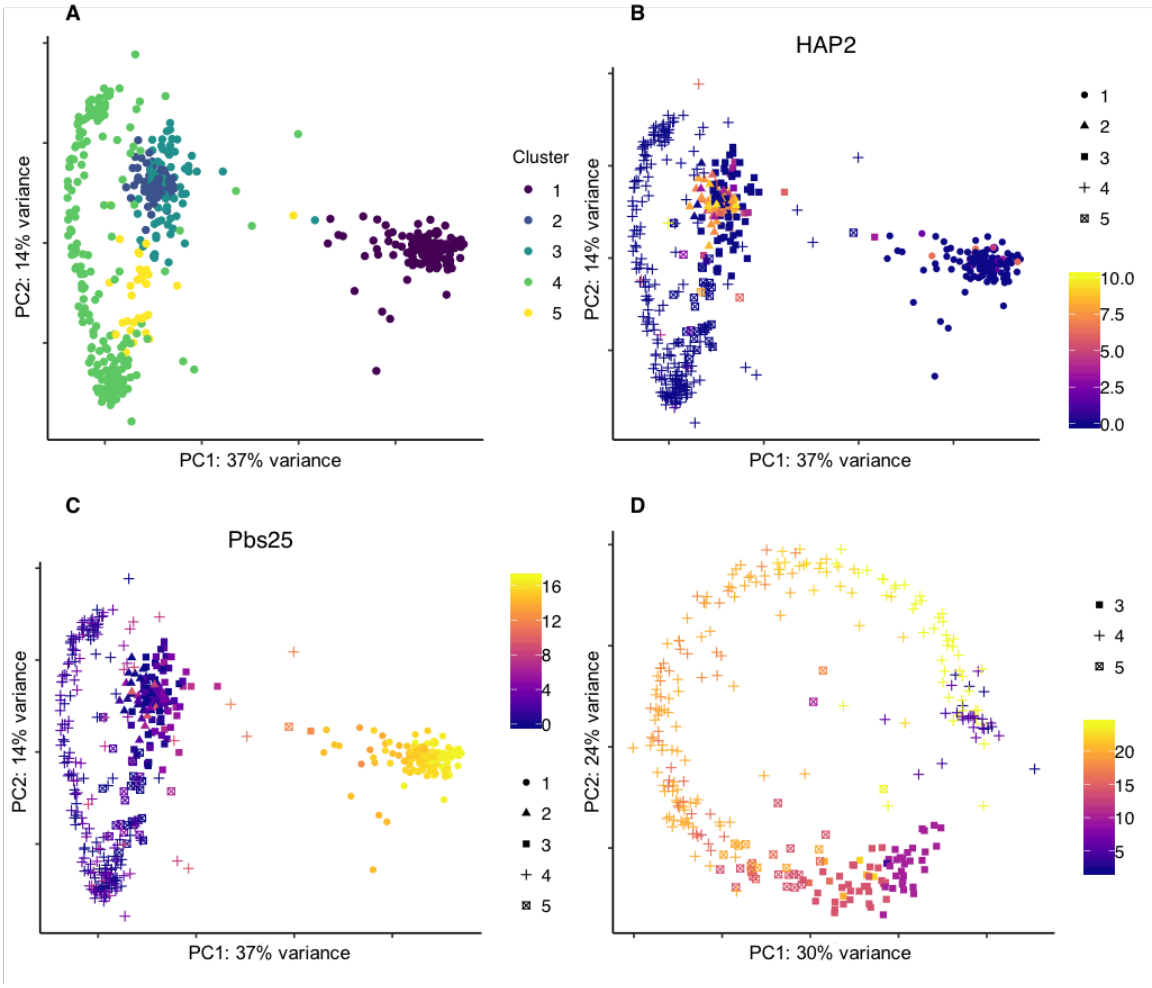

Fig. S3. Classification of blood-stage parasites.

Late-stage blood-stage parasites (trophozoites, schizonts and gametocytes) were purified from an overnight (20 h) culture of infected blood. Classification of these parasites was determined using single-cell consensus clustering in SC3 (*42*). (**A**) PCA of all late-stage parasites coloured by SC3 cluster assignment. (**B**) The PCA coloured by expression of HAP2 (PBANKA_1212600), an established male marker gene (*32*), supporting assignment of cluster 2 cells as male gametocytes. (**C**) The PCA coloured by expression of Pbs25 (PBANKA_0515000), an established female marker gene (*64*), supporting assignment of cluster 1 cells as female gametocytes. (**D**) PCA of asexual parasites (primarily trophozoites and schizonts). Points are shaped by SC3 cluster and coloured by the Spearman correlation with bulk time-course data from (*22*). A good correspondence between SC3 clusters and the bulk data is observed with trophozoites classified as 8-16 h, and schizonts greater than 16 h. A few parasites that cluster with schizonts had the strongest correlation with ring stages, which could be a potential contamination of a small number of rings in the purification process.

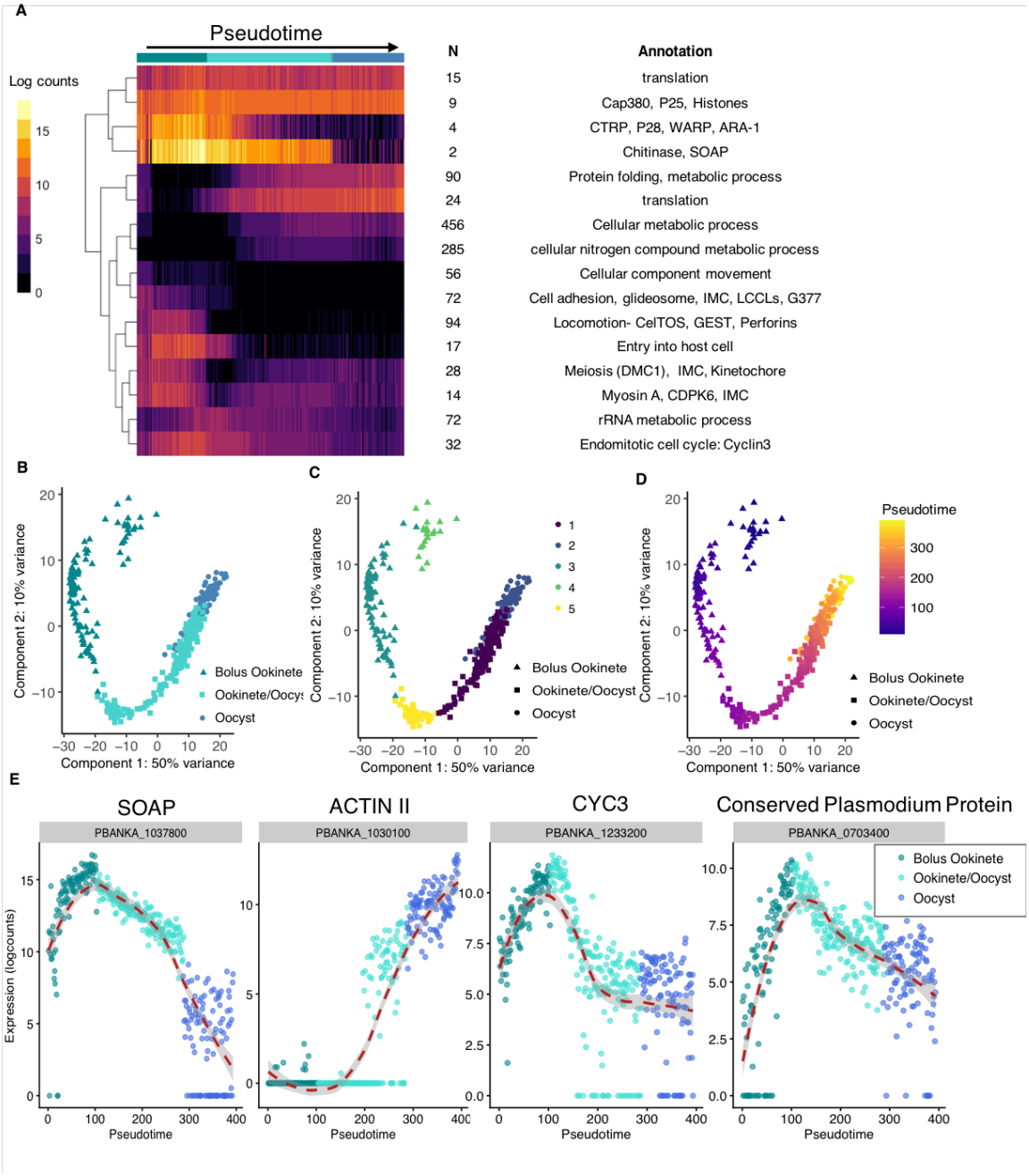

Fig. S4. The ookinete to oocyst transition.

In order to understand fine-scale changes in transcription over developmental time, we ordered cells from three experimental time-points post infectious blood feed (18/24 h bolus ookinetes, the ookinete/oocyst transition period (48 h), and early oocysts (4 days)) in pseudotime using SLICER (*43*). We identified 1270 genes that were differentially expressed over pseudotime among the 393 cells included (file S1). (**A**) A heatmap of the mean expression of each cluster of differentially expressed genes over pseudotime with manual annotation shows fine-scale patterns of expression over development. (**B**) PCA of stages represented by these cells. (**C**) The PCA coloured by their cluster assignments from

SC3. Clusters three, four and five were classified as ookinetes in further analysis, while clusters one and two were classified as oocysts based on expression of known marker genes. (**D**) The pseudotime ordering overlaid on the PCA. The pseudotime ordering matched both the different collection groups and the SC3 clusters. (**E**) Expression of four genes over pseudotime. SOAP (PBANKA_1037800) is an ookinete marker gene (*63*). Actin II (PBANKA_1030100) is an oocyst marker gene (*66*). We found other genes such as cyclin-3 (PBANKA_1233200) (*67*) and PBANKA_0703400 expressed most highly at the transition between ookinete and oocyst stages. Cyclin-3 is known to be essential for oocyst development and we hypothesize that genes with a similar pattern of expression such as PBANKA_0703400 may be essential for the ookinete to oocyst transition.

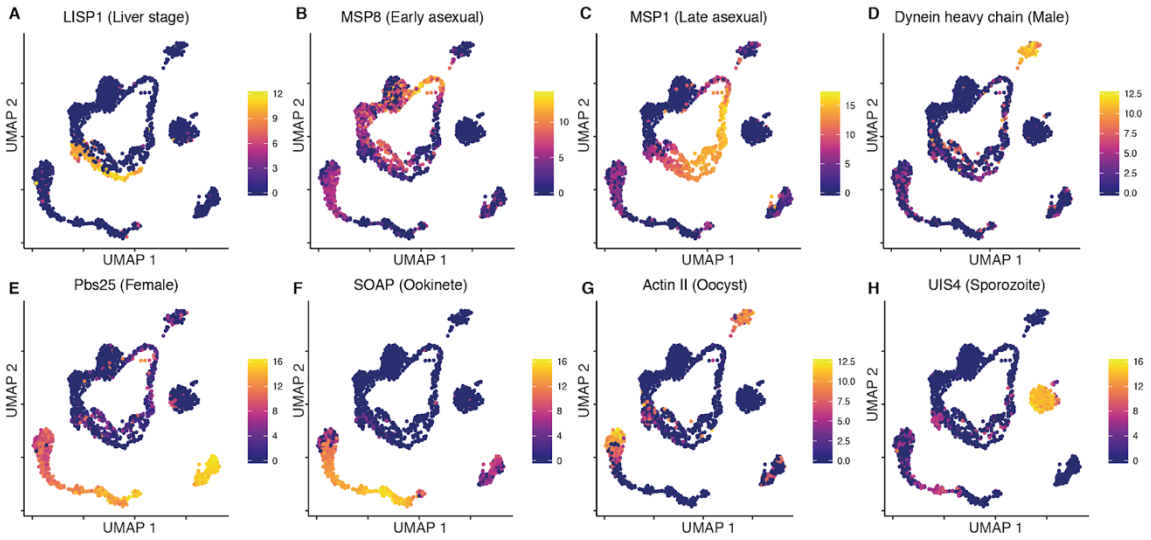

Fig. S5. Marker gene expression.

Expression of known marker genes corresponded with isolation method and stage classification. **(A)** LISP1 (PBANKA_1024600) (*68*) was highly expressed in liver stages. (**B)** MSP8 (PBANKA_1102200) (*69*) was highly expressed in early blood stage asexual parasites. **(C)** MSP1 (PBANKA_0831000) (*70*) was highly expressed in late blood stage asexual parasites. **(D)** Dynein heavy chain (PBANKA_0416100) (*71*) was highly expressed in male gametocytes. **(E)** Pbs25 (PBANKA_0515000) (*64*) was highly expressed in female gametocytes. (**F)** SOAP (PBANKA_1037800) (*63*) was highly expressed in ookinetes. **(G)** Actin II (PBANKA_1030100) (*66*) was highly expressed in oocysts. **(H)** UIS4 (PBANKA_0501200) (*72*) was highly expressed in sporozoites.

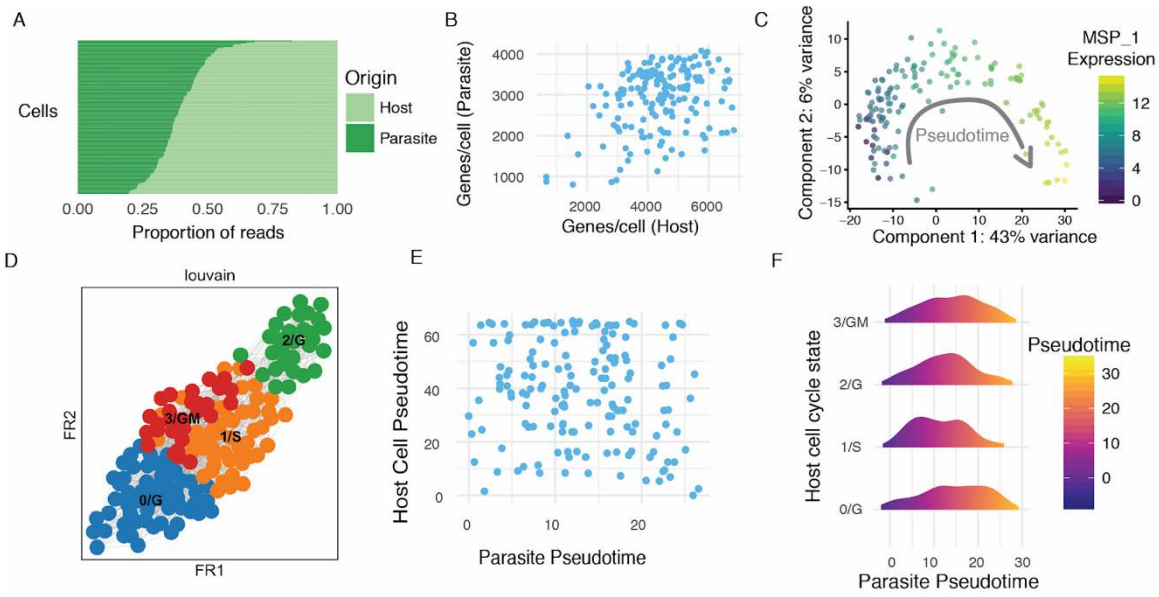

Fig. S6. Dual scRNA-seq of host and parasite transcriptomes during exo-erythrocytic schizogony.

Transcriptomes were generated from HeLa cells containing mCherry liver stage parasites 44 h post infection and selected based on fluorescence **(A)** Proportions of reads mapping to either the host (*H. sapiens*) or the parasite (*P. berghei*) genomes. **(B)** Number of genes per cell identified in host and parasite transcriptomes. **(C)** PCA of parasite transcriptomes identifies a developmental progression of the liver stages that corresponds to MSP-1 (PBANKA_0831000) expression, a known marker of progressing exo-erythrocytic schizogony (*73*). **(D)** Force directed graph (*74*) of host transcriptomes, with different louvain clusters identified as different cell-cycle states. **(E)** Pseudotime analysis for both host and parasite transcriptomes were computed using SLICER (*43*) and plotted against one another showing no correlation of developmental state between the host cell and the parasite cell that resided in it. **(F)** Cell-cycle state of the host cell also did not correspond to a particular pseudotime of the parasite indicating a decoupling between host cell-cycle state and parasite developmental progression.

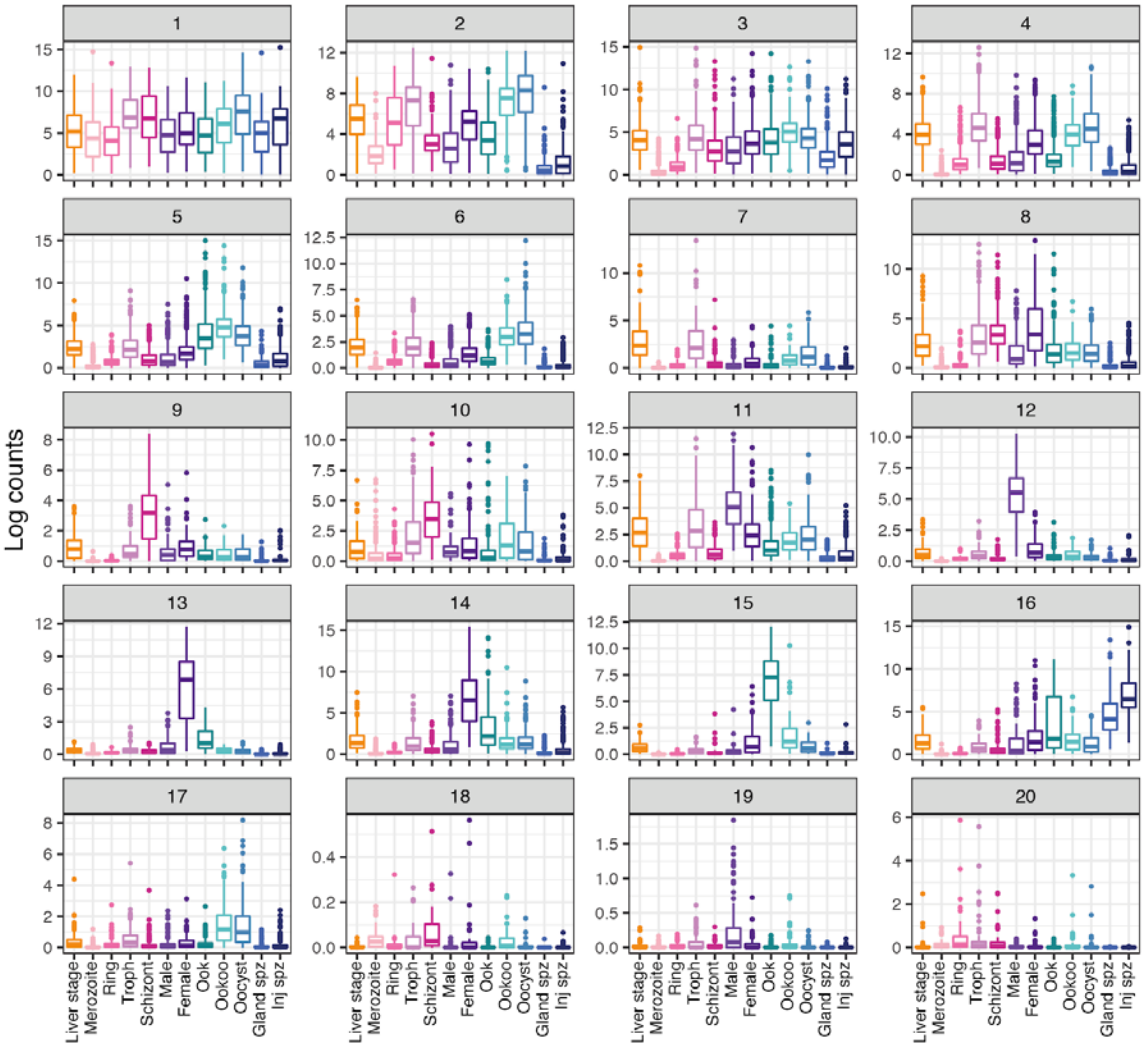

Fig. S7. The mean expression by stage across the genes in each cluster.

The lower and upper hinges of the box-plot correspond to the first and third quartiles. Overall level and stage-specific expression patterns varied across the clusters.

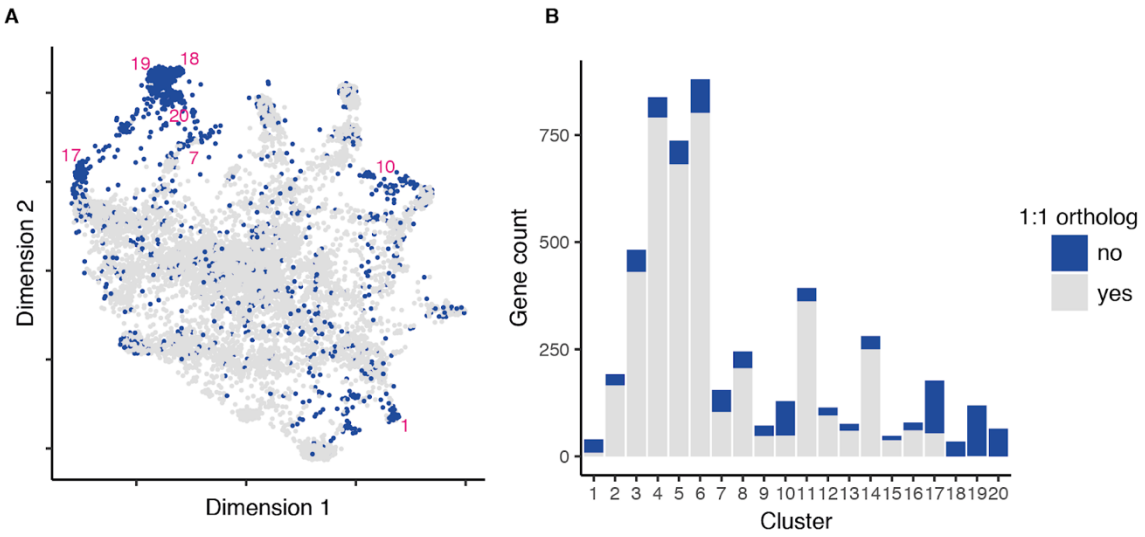

Fig. S8. Conservation across kNN graph.

(A) The kNN graph with non-orthologous (*P. berghei* to *P. falciparum*) genes highlighted in blue. **(B)** A barplot of the gene counts for each cluster colored by orthology. Clusters 1, 7, 10, 17-20 were significantly enriched for genes that have no ortholog with *P. falciparum* (Fisher's exact test, bonferroni p <0.001). Clusters 18, 19, and 20 were composed almost completely of genes from multigene families containing no orthologs in *P. falciparu*m.

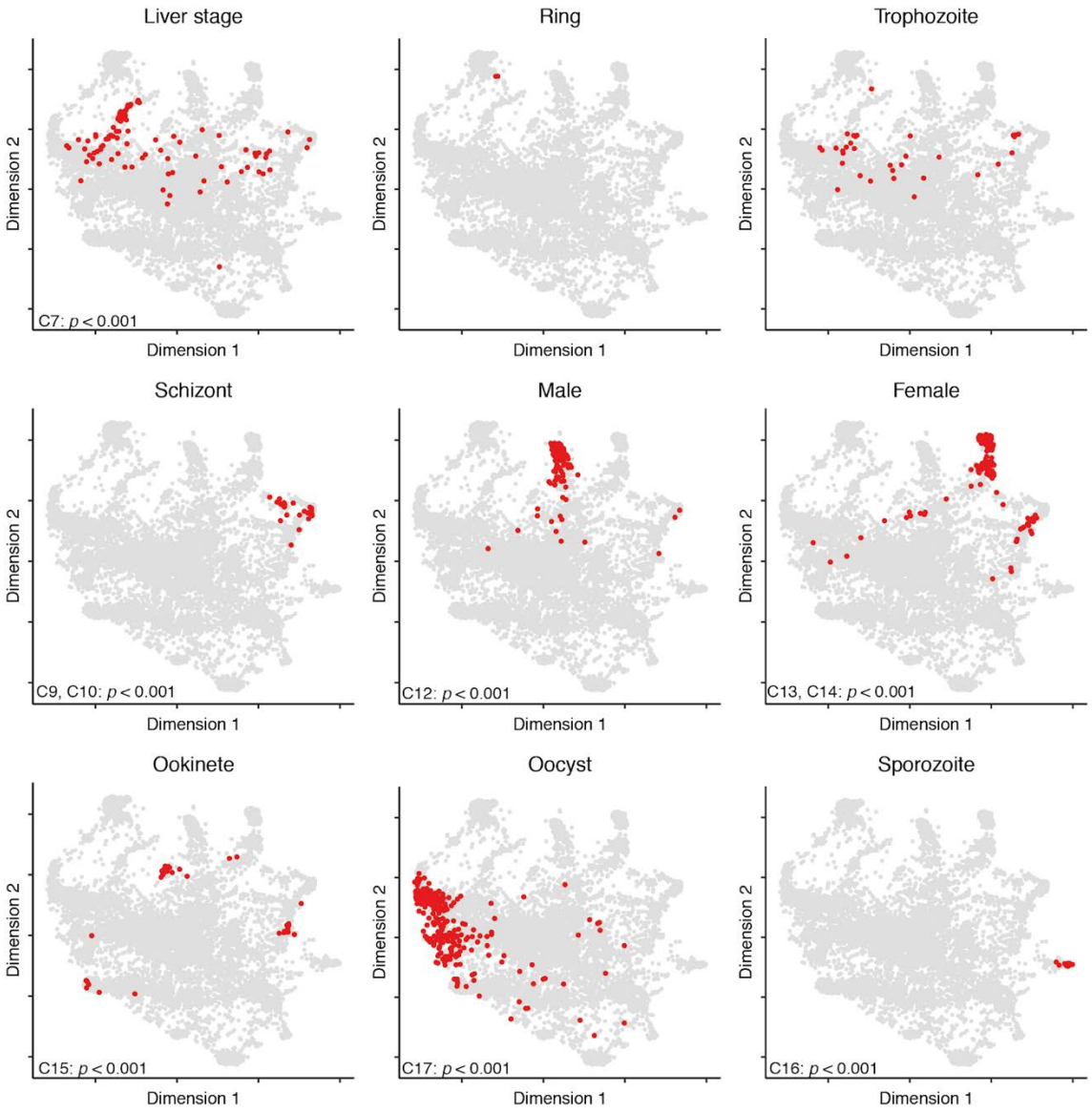

Fig. S9. Unique core expression genes by stage across the kNN graph.

For each stage, a unique core set of genes was defined by identifying the genes for each stage expressed in more than fifty percent of 60 randomly selected cells from that stage. Core genes unique to that stage were considered the 'unique core'. These genes are highlighted in red on the kNN graph for each stage, except for merozoites as there were no unique core genes for this stage. We see a clear pattern of unique core genes associated with specific clusters for most stages (see inset p-values), and a general pattern that unique core genes matched the graph spectral clustering shown in Fig. 2A. Enrichment of core unique genes was calculated for each cluster using a Fisher's exact test.

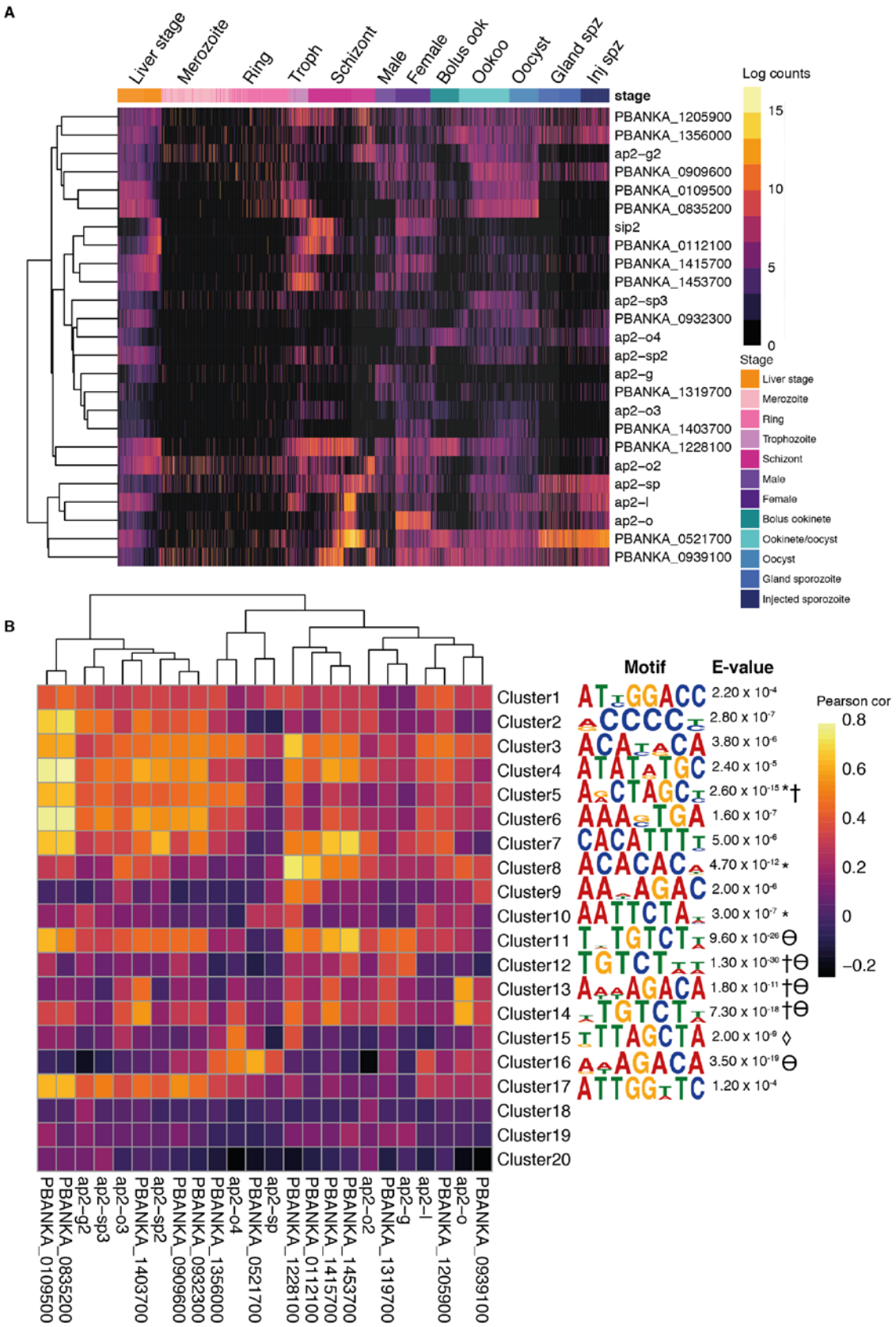
A
Liver stage
Merozoite
Ring
Troph
Schizont
Male
Female
Bolus ook
Ookoo
Oocyst
Gland spz
Inj spz
stage
PBANKA_1205900
PBANKA_1356000
ap2−g2
PBANKA_0909600
PBANKA_0109500
PBANKA_0835200
sip2
PBANKA_0112100
PBANKA_1415700
PBANKA_1453700
ap2−sp3
PBANKA_0932300
ap2−o4
ap2−sp2
ap2−g
PBANKA_1319700
ap2−o3
PBANKA_1403700
PBANKA_1228100
ap2−o2
ap2−sp
ap2−l
ap2−o
PBANKA_0521700
PBANKA_0939100
Log counts
15
10
5
0
Stage
Liver stage
Merozoite
Ring
Trophozoite
Schizont
Male
Female
Bolus ookinete
Ookinete/oocyst
Oocyst
Gland sporozoite
Injected sporozoite
B
Motif
E-value
Cluster1
2.20 x 10-4
Cluster2
2.80 x 10-7
Cluster3
3.80 x 10-6
Cluster4
2.40 x 10-5
Cluster5
2.60 x 10-15 *†
Cluster6
1.60 x 10-7
Cluster7
5.00 x 10-6
Cluster8
4.70 x 10-12 *
Cluster9
2.00 x 10-6
Cluster10
3.00 x 10-7 *
Cluster11
9.60 x 10-26 Θ
Cluster12
1.30 x 10-30 †Θ
Cluster13
1.80 x 10-11 †Θ
Cluster14
7.30 x 10-18 †Θ
Cluster15
2.00 x 10-9 ◊
Cluster16
3.50 x 10-18 Θ
Cluster17
1.20 x 10-4
Cluster18
Cluster19
Cluster20
Pearson cor
0.8
0.6
0.4
0.2
0
−0.2
PBANKA_0109500
PBANKA_0835200
ap2−g2
ap2−sp3
ap2−o3
PBANKA_1403700
ap2−sp2
PBANKA_0909600
PBANKA_0932300
PBANKA_1356000
ap2−o4
PBANKA_0521700
ap2−sp
PBANKA_1228100
PBANKA_0112100
PBANKA_1415700
PBANKA_1453700
ap2−o2
PBANKA_1319700
ap2−g
ap2−l
PBANKA_1205900
ap2−o
PBANKA_0939100

Fig. S10. Expression of AP2 transcription factors across all cells in the data set.

(A) A heatmap of log expression counts of 25 AP2 transcription factors across all cells in the data set. Cells are ordered along a developmental trajectory based on life cycle sequence and pseudotime. **(B)** Heatmap of Pearson's correlations between AP2 gene expression and gene clusters from Fig. 2. The column on the right shows the most significantly enriched DNA motif in the upstream region of the genes included in that cluster with the corresponding e-value (see also file S1). Symbols denote motifs previously identified in other studies using *P. berghei* (†, □) or *P. falciparum* (*, Θ) that also match at e-value <0.01 those found here. Motifs in cluster 5 and clusters 12-14 (†) were also found in (*12*) using bulk RNA-seq of AP2 TF KO mutants in the IDC, gametocytes, and ookinetes. The most significant motif of cluster 15 matches the AP2-O motif (denoted as □) discovered using Chip-seq (*53*). Motifs marked with * correspond to those found in (*52*) using protein binding microarrays in the IDC (*52*), whereas Θ denotes motifs found in (*13*) using cDNA microarrays across the parasite lifecycle. Clusters 5, 8, and 10 had matching motifs in the *P. falciparum* IDC (*52*), whereas clusters 11-14 and 16 all matched the [A/T]AGACA motif (or its reverse complement) discovered in the sexual stages in (*13*). The similar timing of expression of genes containing these motifs strongly suggests the motifs and the function of the genes they regulate are conserved across species. The correspondence between our data and these various methods of motif discovery suggests that our analysis can identify as-yet unknown TF binding motifs in *Plasmodium*.

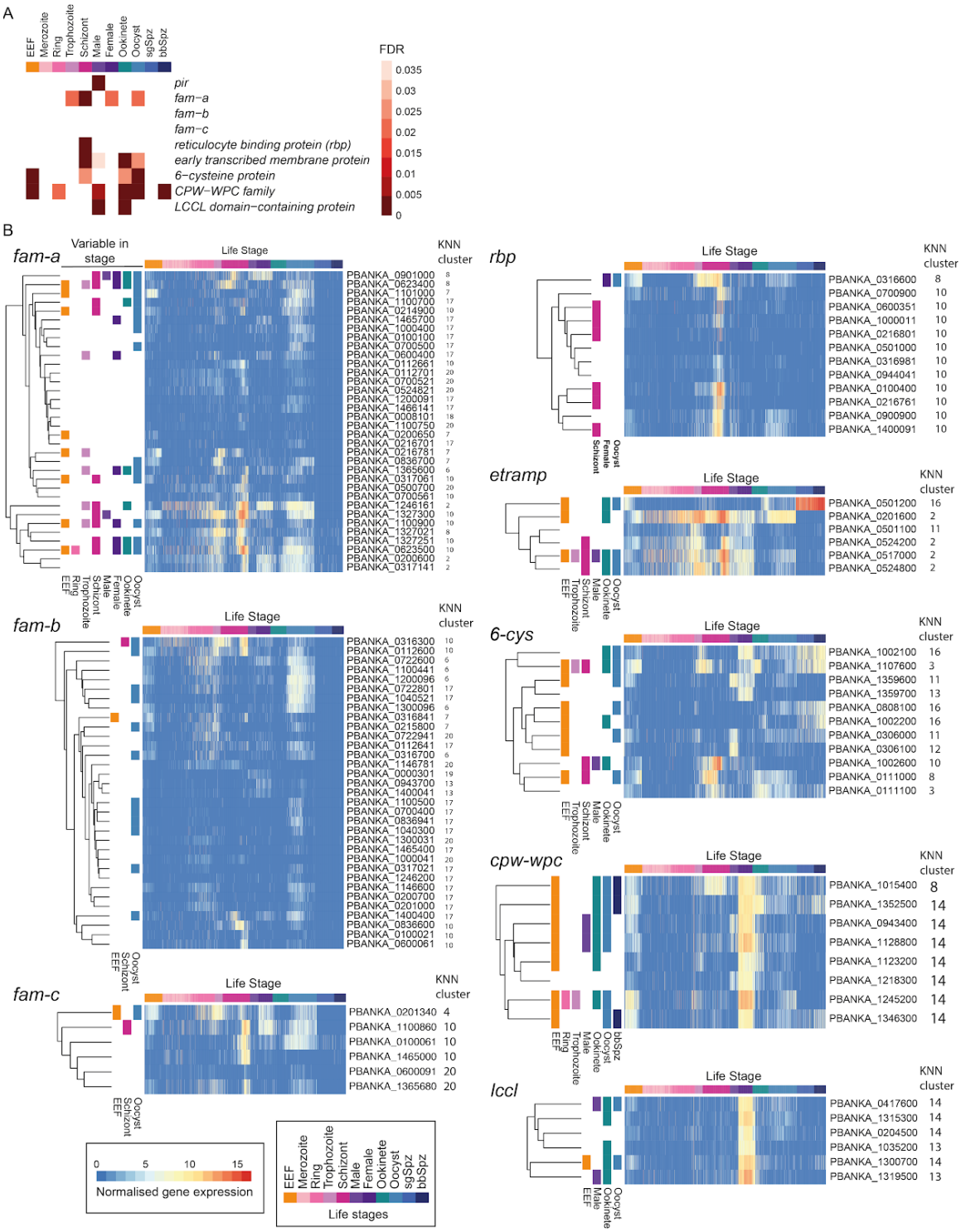
A
EEF
Merozoite
Ring
Trophozoite
Schizont
Male
Female
Ookinete
Oocyst
SgSpz
bbSpz
FDR
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
pir
fam−a
fam−b
fam−c
reticulocyte binding protein (rbp)
early transcribed membrane protein
6−cysteine protein
CPW−WPC family
LCCL domain−containing protein
B
fam-a
Variable in stage
Life Stage
KNN cluster
PBANKA_0901000 8
PBANKA_0623400 8
PBANKA_1101000 7
PBANKA_1100700 17
PBANKA_0214900 10
PBANKA_1465700 17
PBANKA_1000400 17
PBANKA_0100100 17
PBANKA_0700500 17
PBANKA_0600400 17
PBANKA_0112661 10
PBANKA_0112701 20
PBANKA_0700521 20
PBANKA_0524821 20
PBANKA_1200091 17
PBANKA_1466141 17
PBANKA_0008101 18
PBANKA_1100750 20
PBANKA_0200650 7
PBANKA_0216701 17
PBANKA_0216781 7
PBANKA_0836700 7
PBANKA_1365600 6
PBANKA_0317061 10
PBANKA_0500700 20
PBANKA_0700561 10
PBANKA_1246161 2
PBANKA_1327300 10
PBANKA_1100900 10
PBANKA_1327021 8
PBANKA_1327251 10
PBANKA_0623500 10
PBANKA_0200600 2
PBANKA_0317141 2
Oocyst
Ookinete
Male
Female
Schizont
Trophozoite
Ring
EEF
fam-b
Life Stage
KNN cluster
PBANKA_0316300 10
PBANKA_0112600 10
PBANKA_0722600 6
PBANKA_1100441 6
PBANKA_1200096 6
PBANKA_0722801 17
PBANKA_1040521 17
PBANKA_1300096 6
PBANKA_0316841 7
PBANKA_0215800 7
PBANKA_0722941 20
PBANKA_0112641 17
PBANKA_0316700 6
PBANKA_1146781 20
PBANKA_0000301 19
PBANKA_0943700 13
PBANKA_1400041 13
PBANKA_1100500 17
PBANKA_0700400 17
PBANKA_0836941 17
PBANKA_1040300 17
PBANKA_1300031 20
PBANKA_1465400 17
PBANKA_1000041 20
PBANKA_0317021 17
PBANKA_1246200 17
PBANKA_1146600 17
PBANKA_0200700 17
PBANKA_0201000 17
PBANKA_1400400 17
PBANKA_0836600 10
PBANKA_0100021 10
PBANKA_0600061 10
Oocyst
Schizont
EEF
fam-c
Life Stage
KNN cluster
PBANKA_0201340 4
PBANKA_1100860 10
PBANKA_0100061 10
PBANKA_1465000 10
PBANKA_0600091 20
PBANKA_1365680 20
Oocyst
Schizont
EEF
0 5 10 15
Normalised gene expression
EEF
Merozoite
Ring
Trophozoite
Schizont
Male
Female
Ookinete
Oocyst
SgSpz
bbSpz
Life stages
rbp
Life Stage
KNN cluster
PBANKA_0316600 8
PBANKA_0700900 10
PBANKA_0600351 10
PBANKA_1000011 10
PBANKA_0216801 10
PBANKA_0501000 10
PBANKA_0316981 10
PBANKA_0944041 10
PBANKA_0100400 10
PBANKA_0216761 10
PBANKA_0900900 10
PBANKA_1400091 10
Oocyst
Female
Schizont
etramp
Life Stage
KNN cluster
PBANKA_0501200 16
PBANKA_0201600 2
PBANKA_0501100 11
PBANKA_0524200 2
PBANKA_0517000 2
PBANKA_0524800 2
EEF
Trophozoite
Schizont
Male
Ookinete
Oocyst
6-cys
Life Stage
KNN cluster
PBANKA_1002100 16
PBANKA_1107600 3
PBANKA_1359600 11
PBANKA_1359700 13
PBANKA_0808100 16
PBANKA_1002200 16
PBANKA_0306000 11
PBANKA_0306100 12
PBANKA_1002600 10
PBANKA_0111000 8
PBANKA_0111100 3
EEF
Trophozoite
Schizont
Male
Ookinete
Oocyst
cpw-wpc
Life Stage
KNN cluster
PBANKA_1015400 8
PBANKA_1352500 14
PBANKA_0943400 14
PBANKA_1128800 14
PBANKA_1123200 14
PBANKA_1218300 14
PBANKA_1245200 14
PBANKA_1346300 14
EEF
Ring
Trophozoite
Male
Ookinete
Oocyst
bbSpz
lccl
Life Stage
KNN cluster
PBANKA_0417600 14
PBANKA_1315300 14
PBANKA_0204500 14
PBANKA_1035200 13
PBANKA_1300700 14
PBANKA_1319500 13
EEF
Male
Ookinete
Oocyst

Fig. S11. Multigene families show development-independent variable expression between cells in most life stages.

(A) A heatmap shows which stages were enriched for variability in expression of each multigene family. Variable genes were identified in each stage using Smart-seq2 data and controlling for development by regressing out pseudotime in liver stage EEFs, merozoites, rings, trophozoites, schizonts, ookinetes and oocysts. The hypergeometric test was used to determine enrichment of each gene family amongst variable genes in each stage and the resulting *p*-values were adjusted using the False Discovery Rate (FDR). **(B)** Each heatmap shows the TMM-normalised expression levels of members of a gene family over the life cycle of *P. berghei* (except for the *pir* gene family that is explored in fig S12). The data were filtered to show only those genes with at least 10 read counts in at least 10 cells. Genes were clustered based on their expression pattern. Genes found to vary in expression level between cells of the same stage independent of development are highlighted to the left of each heatmap. The presence of a turquoise square, for example, indicates that that gene was variably expressed between ookinetes. The numbers to the right of each gene represent the cluster assignment from the kNN gene graph in Fig. 2A.

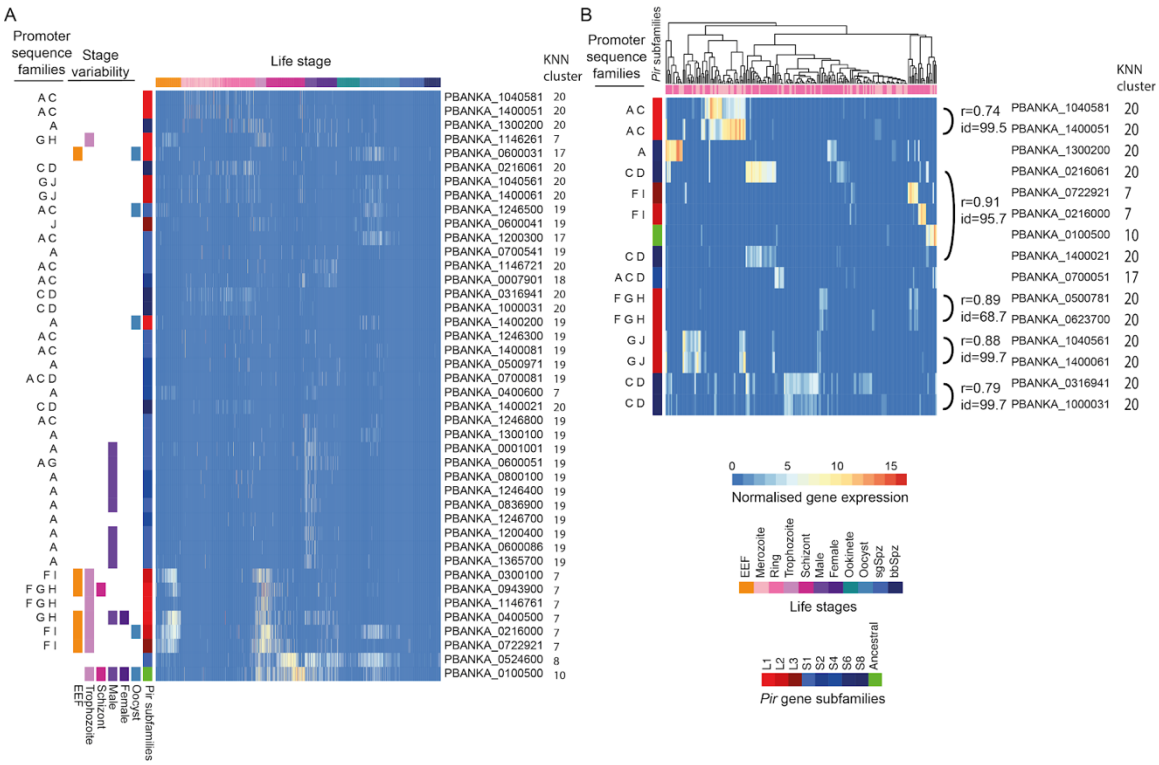

Fig. S12. The *pir* gene family shows distinct patterns of expression and promoter architecture in several parts of the life cycle.

(A) The heatmap shows the TMM-normalised expression levels of members of the *pir* gene family over the life cycle of *P. berghei*. The data were filtered to show only those genes expressed with at least 10 read counts in at least 10 cells over the whole data set. The 42 genes of the 137 annotated *pir* genes that passed this threshold are shown in the heatmap. Genes were clustered based on their expression pattern. The presence of an orange square, for example, indicates that that gene was variably expressed between liver stage EEFs. The subfamily classification of each *pir* gene is indicated in red colours (L-type *pir* genes), blue colours (S-type *pir* genes) and green (ancestral *pir* gene) to the left of the heatmap. Families of sequence identified in the promoter regions of each gene are indicated with letters A-J. One cluster of *pir* genes variably expressed in both EEFs and trophozoites tended to have 'F' sequences in their promoters. Another cluster was variably expressed in male gametocytes and tended to have promoters with 'A' sequences. **(B)** This heatmap shows merozoites and rings and *pir* genes expressed with at least 10 read counts in at least 10 of these cells. Both genes and cells were clustered based on their expression levels. Five pairs of genes show evidence of co-expression (Pearson r > 0.7). Each pair tended to have high sequence identity (BLAST sequence identity > 95%), the same promoter architecture (identical pattern of upstream sequence families) with each member of a pair residing on different chromosomes. The numbers to the right of each gene represent the cluster assignment from the kNN gene graph in Fig. 2A.

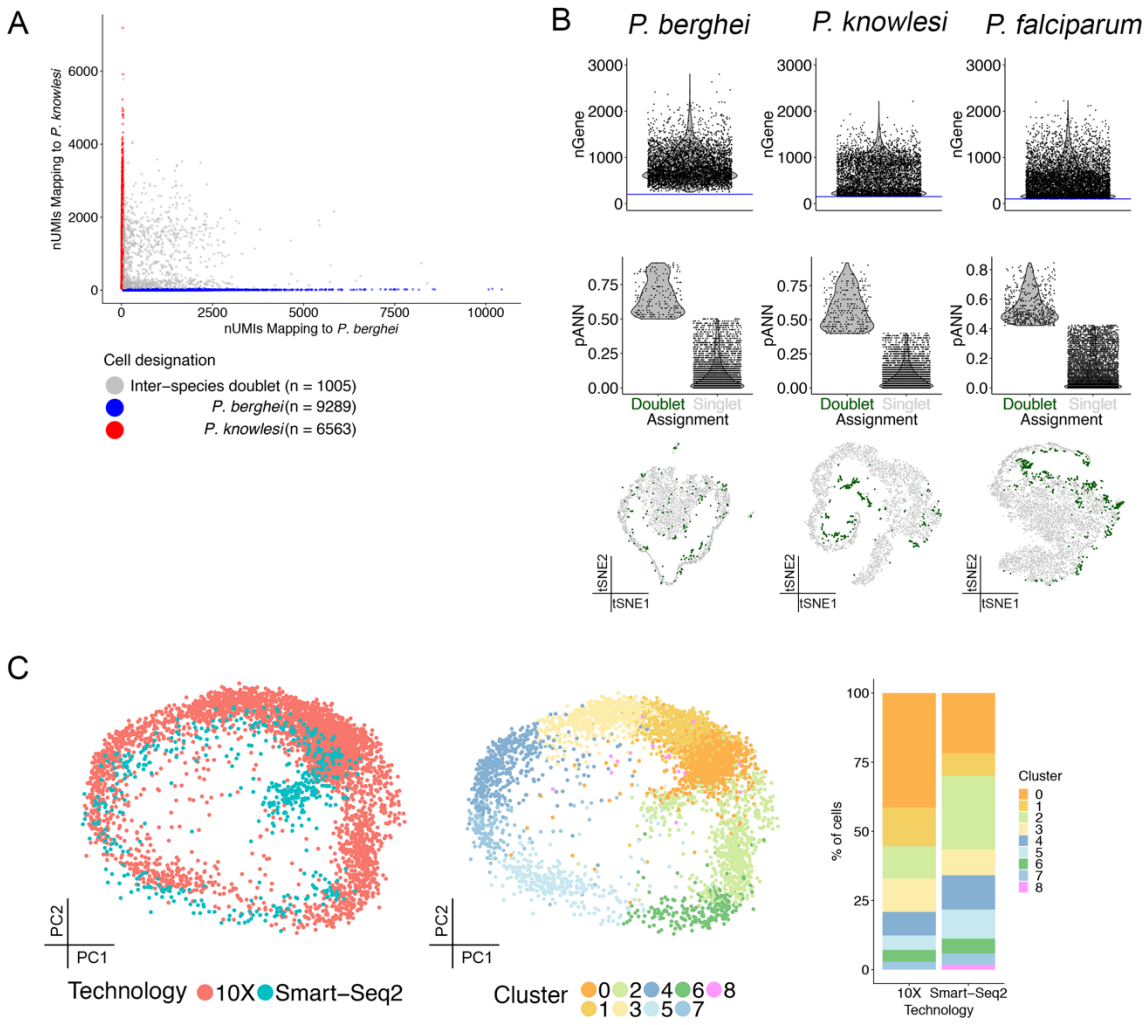

Fig. S13. Comparison of Smart-seq2 data with 10x data and technical assessment

(A) A 'barnyard' plot showing the doublet rate between the *P. knowlesi* and *P. berghei* mixed species experiment. **(B)** For each species, Top: a violin plot showing the number of genes detected per cell (nGene) with a line representing the QC threshold used; Middle: A violin plot showing the proportion of artificial nearest neighbors (pANN) as calculated by DoubletFinder (*41*), with cells split by their final assignment; Bottom: A tSNE plot showing doublets highlighted in green against singlets in grey, showing the distribution of doublets among stages. **(C)** PCA plot overlaying the Smart-seq2 to the 10x data using a CCA corrected distance matrix coloured by technology (left) and cluster (middle). The proportional representation of these clusters captured by each technology is displayed on the right.

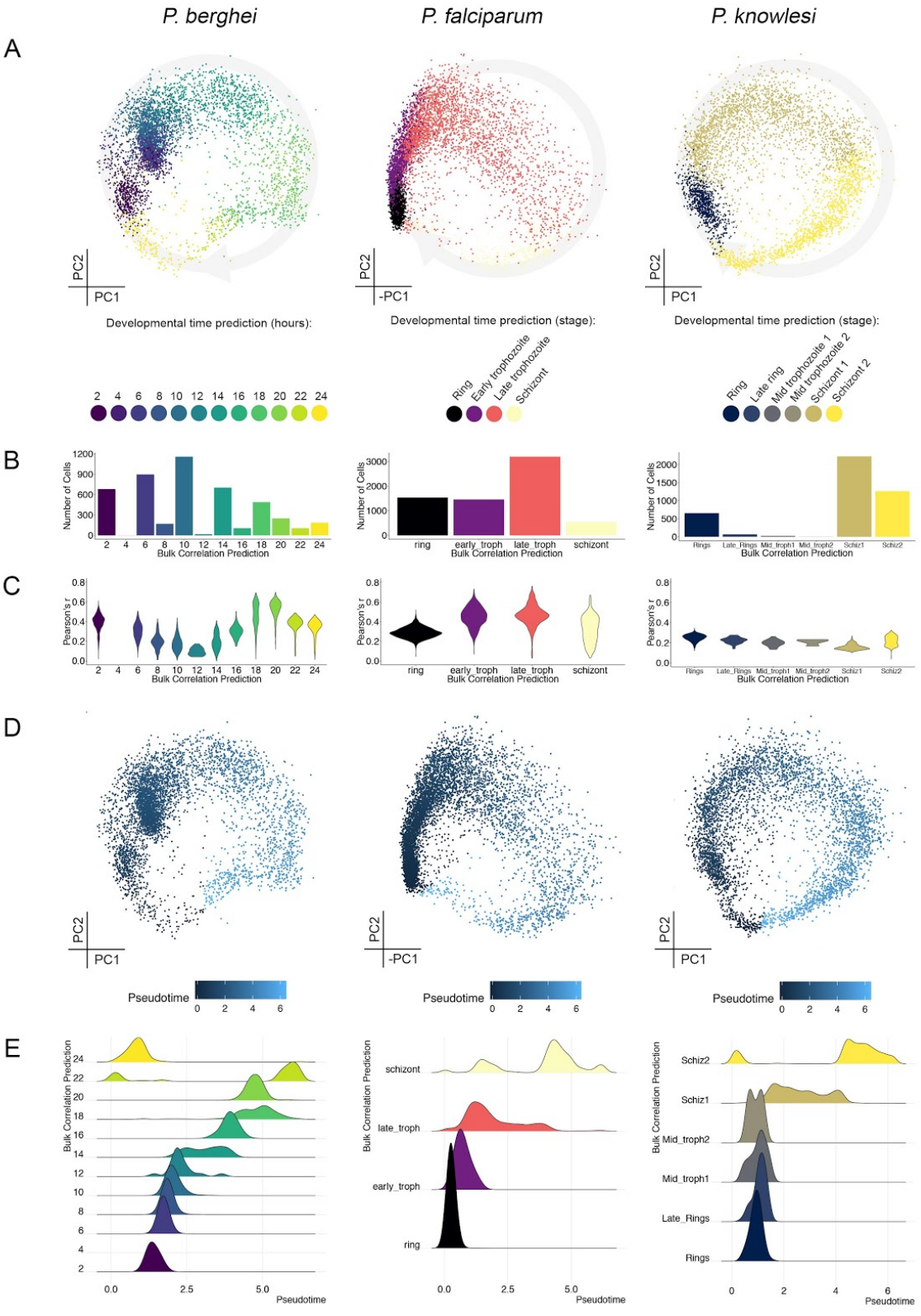
P. berghei
P. falciparum
P. knowlesi
A
PC2
PC1
-PC1
Developmental time prediction (hours):
2 4 6 8 10 12 14 16 18 20 22 24
Developmental time prediction (stage):
Ring
Early trophozoite
Late trophozoite
Schizont
Developmental time prediction (stage):
Ring
Late ring
Mid trophozoite 1
Mid trophozoite 2
Schizont 1
Schizont 2
B
Number of Cells
Bulk Correlation Prediction
ring
early_troph
late_troph
schizont
Rings
Late_Rings
Mid_troph1
Mid_troph2
Schiz1
Schiz2
C
Pearson's r
D
Pseudotime
E
Pseudotime

Fig. S14. Correlation of 10x data sets with bulk expression data

(A) PCA plots for each species, coloured by their predicted life cycle stage according to maximum correlation with the following bulk time course reference data sets: *P. berghei* microarray (*22*); *P. falciparum* RNA-seq (*75*); *P. knowlesi* microarray (*76*). **(B)** Distribution of cells according to predicted life cycle stage based on bulk data. The temporal density of sampling over the IDC is different between bulk data sets so this results in a different number of bins for each species. **(C)** Violin plots showing the Pearson correlation coefficient, r, for each species by predicted life cycle stage. **(D)** PCA plots for each species, coloured by pseudotime value, calculated using the custom clock approach. **(E)** Density ridgeline plots showing the relationship between pseudotime and predicted life cycle stage.

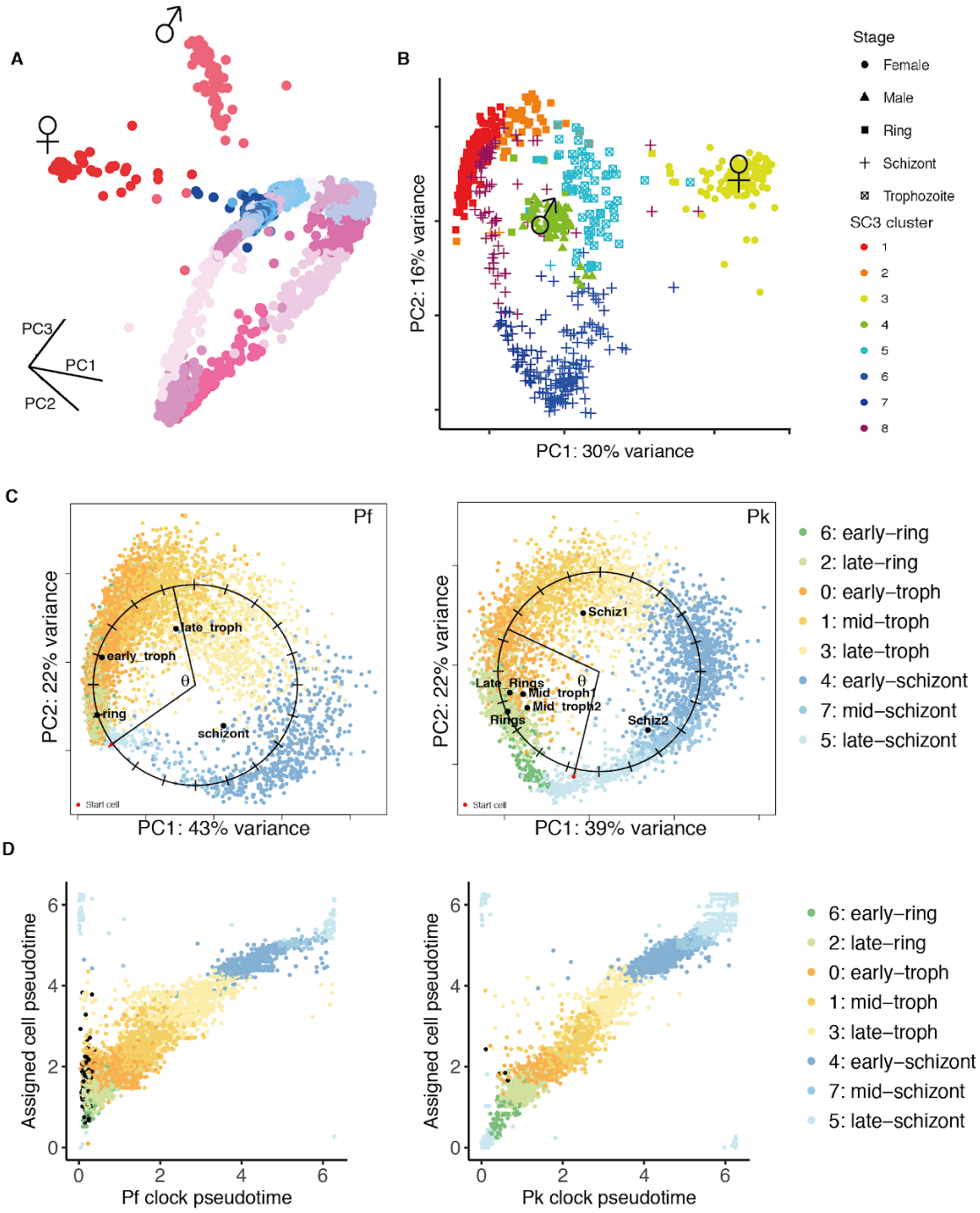

Fig. S15. scmap of 10x data

(A) In order to identify male and female gametocytes, clustering of 10x data was initially done with the Seurat shared nearest neighbor modularity optimization method, which identified 26 clusters, two of which corresponded to mature male and female gametocytes based on known markers. Cells are shown on a PCA and colored by cluster assignment. **(B)** SC3 clustering of blood-stage parasite data (rings, trophozoites, schizonts, and mature gametocytes). Cells were re-clustered in SC3 into eight clusters that were then used for the cluster assignment shown in Fig. 3A. **(C)** 'Clock' pseudotime was calculated independently

for both *P. falciparum* (left) and *P. knowlesi* (right) and the mean coordinates of the bulk prediction were mapped onto the PCA (black points). **(D)** The correspondence between the independent pseudotime calculation from (B) and the pseudotime of the assigned *P. berghei* cell based on scmap. Black points represent “unassigned” cells based on a cosine similarity of <0.3. There was a strong correlation between pseudotime for each species and the assigned cell pseudotime further supporting the cell assignments from scmap.

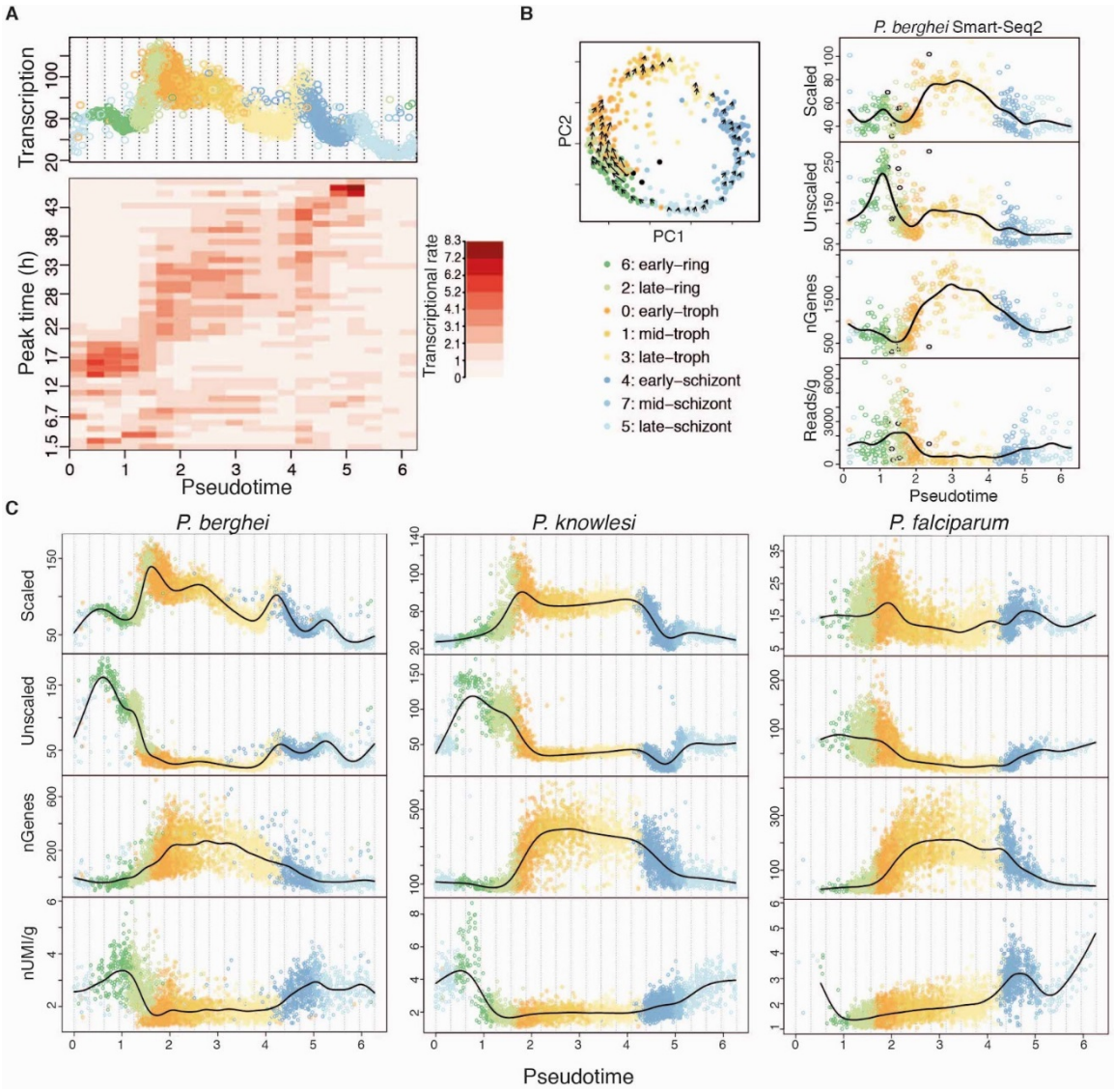

Fig. S16. RNA velocity varies across the IDC.

(A) A heatmap showing the average scaled transcriptional rate of genes in *P. berghei* as measured by RNA velocity (*20*) in the IDC. Genes are ordered and binned along the vertical axis by their peak time in *P. falciparum* in (*21*). The top panel shows the scaled transcriptional rate of each cell over pseudotime in the *P. berghei* 10x data. The groups of genes from (*21*) also had high transcriptional rates at corresponding points in the life cycle in *P. berghei* based on RNA velocity. **(B)** RNA velocity of the Smart-seq2 IDC cells. Cells were mapped to the 10x *P. berghei* reference index with scmap-cell to assign each cell a cluster and a pseudotime value allowing us to directly compare the two data sets. Cells are colored by their assigned cell cluster and black points represent "unassigned" cells based on cosine similarity. The left plot is a PCA of the 548 Smart-seq2 IDC cells with the arrows representing the local average velocity. The right panel displays the same cells over pseudotime showing the relative increase in transcription across all transcripts unscaled or scaled by gene, followed by the number of genes detected and the number of reads per gene in each cell over pseudotime. **(C)** The three species 10x data sets over assigned cell

pseudotime based on the *P. berghei* 10x reference index. Cells are colored by their assigned cell cluster. Each panel shows the relative increase in transcription (scaled and unscaled) as well as the number of genes detected and the number of UMIs per gene. We observed a similar pattern of transcriptional dynamics in both the Smart-seq2 and 10x *P. berghei* data.

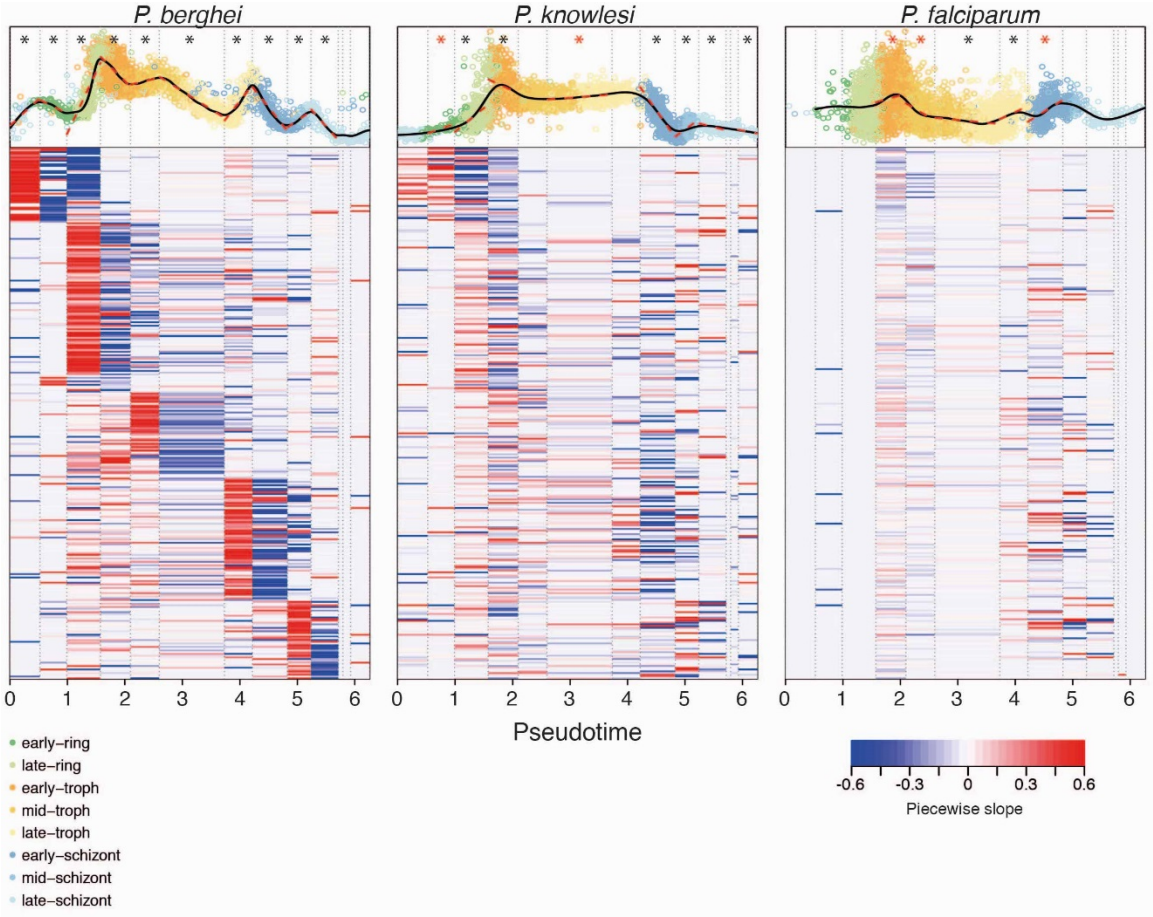

Fig. S17. Comparison of transcriptional dynamics through the IDC across species.

Peaks and troughs in the *P. berghei* data were algorithmically identified (represented as dashed grey lines in in top plots). Stars and dashed red lines indicate significant ($p < 0.001$) increases or decreases in transcriptional rate within these windows in each species. A red star represents a significant change that was in the opposite direction as *P. berghei*. The heat maps represent the 306 genes that passed the RNA velocity quality control thresholds and were one to one orthologs across all three species. The increase/decrease in transcriptional rate of each of these genes through the IDC was evaluated using piecewise-regression between the peaks and troughs in each species independently (top panels). Those found to have significant slopes (FDR 5%) are shown in the heatmap. Non-significant slopes were set to 0 for visualization, and slopes >0.6 and <-0.6 were assigned the most intense colours (see supplementary file 4 for exact values). We found conserved patterns of expression between late-rings and early-trophs as well as in early schizonts between *P. berghei* and *P. knowlesi* (table S2). The increase in rate as well as the expression of corresponding genes seen in *P. berghei* and *P. knowlesi* late-rings is slightly delayed in *P. falciparum* (one window later) (table S2).

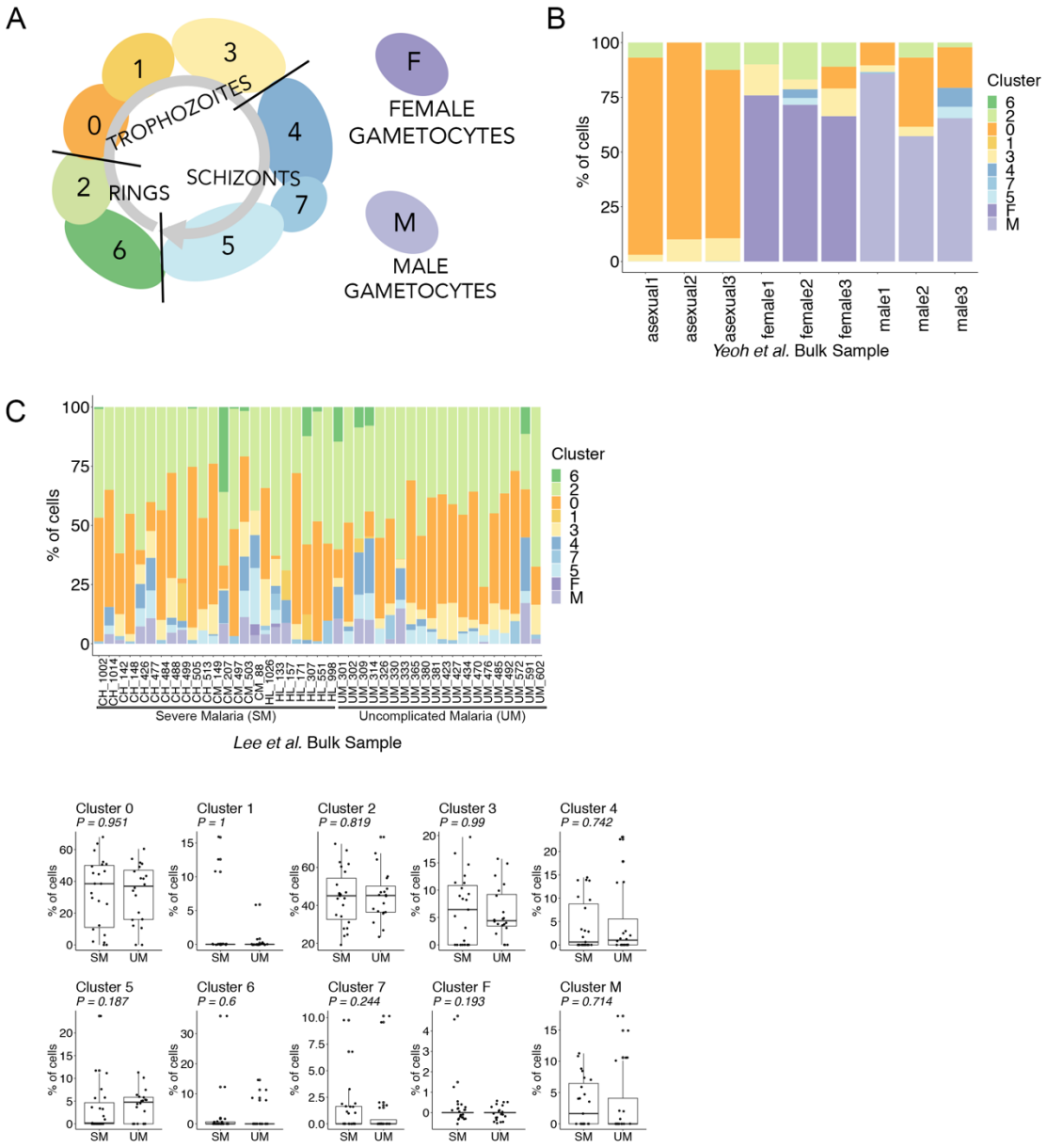

Fig. S18. Deconvolution of bulk transcriptomic samples using scRNA-seq data.

(A) A diagram of the *P. berghei* 10x data from fig. S13, which is used as a reference to call markers for each cell cluster. **(B)** The *Yeoh et al.* study (*23*) used the 820cl1m1cl1 *P. berghei* parasite line (*77*), which has RFP females and GFP males, to examine each sex separately by RNA-seq. Asexual samples were obtained straight from a mouse and only circulating forms are expected to be present. Using BSeq-SC (*62*) to estimate cell type proportions present in the bulk data, only circulating forms are observed in the asexual samples, as well as a majority of females in the female samples, and a majority of males in the male samples. Additional cells called in the gametocyte samples are hypothesised to be developing gametocytes, technical noise in our method, and/or true asexual cells that are

present in the sorting gate. **(C)** We tested whether the *P. berghei* cell atlas could be used to deconvolve bulk RNA-seq data from *P. falciparum* samples collected from patients with uncomplicated or severe malaria (*24*). We find that the cell population in the bulk data is composed primarily of early stage circulating forms as expected (top). We also observe cell type composition is independent of the severity of malaria (bottom) as described by the authors in this study. The Malaria Cell Atlas provides a robust index by which to deconvolve bulk transcriptomic data into contributing cell populations.

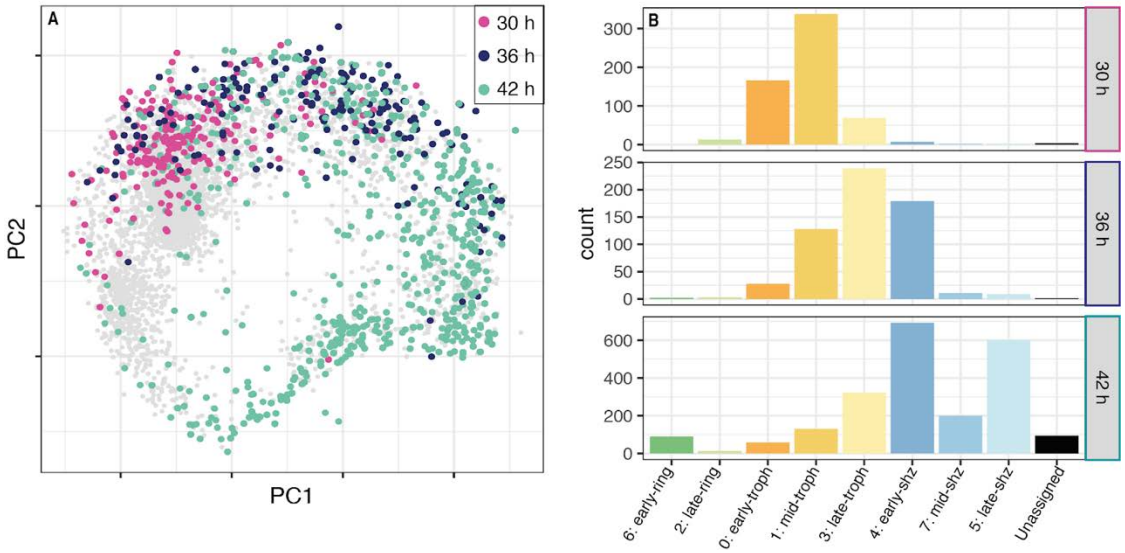

Fig. S19. scmap of *P. falciparum* time course Drop-seq data

(A). A PCA of the 10x *P. berghei* IDC data (grey points) overlaid with cell assignments of a Drop-seq data time course (*25*) using scmap-cell. **(B).** The distribution of cluster assignments based on scmap-cell across the three time points from (*25*) The time points correspond with the cluster assignments from the *P. berghei* data with the early time points having more early stages and the late time point having more late stages. It is interesting to note how the synchronous culture still captures a broad range of stages, especially in the late time point.

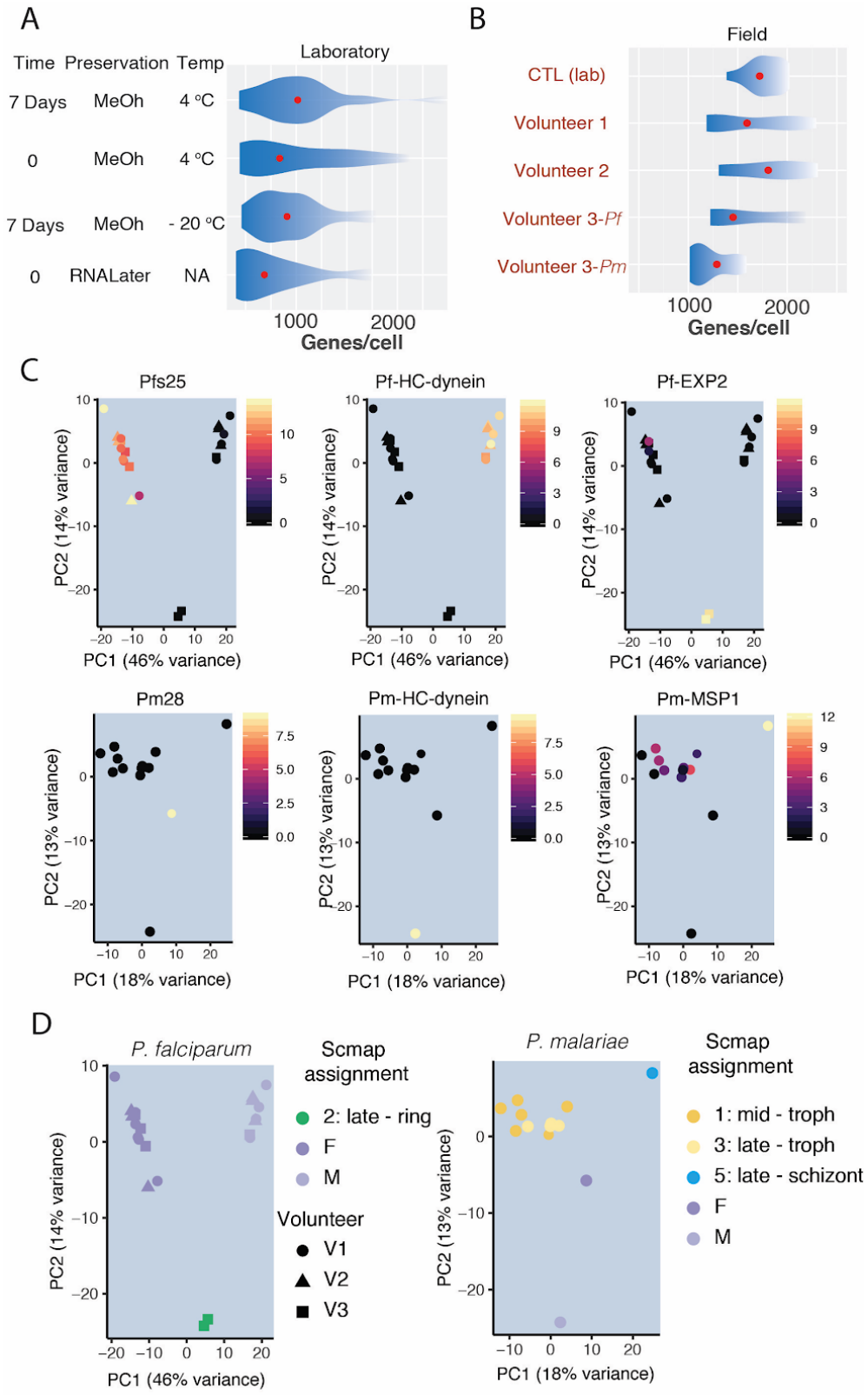
A
Time Preservation Temp
7 Days MeOh 4 °C
0 MeOh 4 °C
7 Days MeOh - 20 °C
0 RNALater NA
Laboratory
1000 2000
Genes/cell
B
Field
CTL (lab)
Volunteer 1
Volunteer 2
Volunteer 3-Pf
Volunteer 3-Pm
1000 2000
Genes/cell
C
Pfs25
Pf-HC-dynein
Pf-EXP2
Pm28
Pm-HC-dynein
Pm-MSP1
PC1 (46% variance)
PC2 (14% variance)
PC1 (18% variance)
PC2 (13% variance)
D
P. falciparum
P. malariae
Scmap assignment
2: late - ring
F
M
Volunteer
V1
V2
V3
1: mid - troph
3: late - troph
5: late - schizont

Fig. S20. Preservation and analysis of single cell transcriptomes from natural carriers of *Plasmodium*.

(A) *P. falciparum* trophozoites and schizonts were preserved in 80% methanol and kept at different temperatures for varying amounts of time before analysis by Smart-seq2; this revealed that transcriptomes from methanol fixed parasites are of equivalent quality to parasites fixed briefly with RNALater in the lab. **(B)** We methanol preserved parasites collected from naturally-infected carriers in Mbita, Kenya and established that even after three weeks of preservation, high quality single cell transcriptomes were possible to retrieve using Smart-seq2. **(C)** PCA of 22 *P. falciparum* and 13 *P. malariae* wild single cell transcriptomes showing expression of known *P. falciparum* (*78*) or putative *P. malariae* male (HC-dynein PF3D7_0905300 and PmUG01_07016800), female (Pfs25 (PF3D7_1031000) and Pm28 (PmUG01_10042100)), early ring (EXP2 (PF3D7_1471100) and PmUG01_12045800), and putative late stage markers (MSP-1) **(D)** PCAs of the same cells as in (C) but with their scmap assignment based on the *P. berghei* 10x Malaria Cell Atlas. Life stages can be assigned using known markers as demonstrated in (C) but to place cells in developmental time, the full atlas is required (see Fig. 4B for placement of these cells).

Table S1. Quality control of single-cell transcriptomes by purification method.

Additional thresholding (fig. S2) was performed on the late blood stages after stage assignment with SC3 (fig. S3). Liver-stage, merozoite, ring, and late blood stage time points are represented as hours post invasion. Ookinete, ookinete/oocyst, oocyst, gland and injected sporozoite time points are hours or days post infectious feed.

Stages purified	Time point	# before QC	Genes per cell cut off	Reads per cell cut off	# Passed QC	Success rate	Median genes per cell
Liver stage	44 h	186	1000	2500	159	85%	2977
Merozoite	24 h	249	40	400	229	92%	202
Ring	0 h	282	40	2500	235	83%	392.5
Late blood stages	24 h	560	500	2500	517	92%	1537
Ookinete	18 h & 24 h	129	1000	2500	103	80%	1558
Ookinete/ Oocyst	48 h	187	1000	2500	183	98%	3088
Oocyst	4 days	133	1000	2500	106	80%	2995
Gland spz	26 days	188	40	400	153	81%	419
Injected spz	26 days	108	500	2500	102	94%	859.5
Early blood stages (Field)	NA	106	40	2500	2	2%	630
Late blood stages (Field)	NA	78	500	2500	35	45%	1101 (Pm, n=13) 1355 (Pf, n=22)

Table S2. Alignment of transcriptional dynamics in three *Plasmodium* species.

Twelve peaks and troughs were algorithmically identified in the transcriptional waves through the IDC in *P. berghei*. For each of the 13 segments defined by these peaks and troughs we performed piecewise linear regression on the individual genes for significant increases/decreases in transcriptional rate across each segment in each species (Pb DE genes, 5% FDR). To examine the conservation of the genes involved in the transcriptional waves through the IDC, we identified orthologs of the genes used in the RNA velocity analysis across all three species and counted the number of genes that had a consistent (agreement) or opposite (disagreement) direction with the pattern observed in *P. berghei*. Genes that did not significantly increase/decrease in both species were not counted among the agreement/disagreement. The p-values were calculated using a Binomial test to evaluate if a greater number of genes were in agreement with *P. berghei* than were in disagreement.

Time	Pb DE genes	Pb stage matching Pk	Gene Agreement Pk & Pb	Gene Disagreement Pk & Pb	p-value	Pb stage matching Pf	Gene Agreement Pf & Pb	Gene Disagreement Pf & Pb	p-value
0.26	83	0.26	40	17	0.0006	None	NA	NA	-
0.76	78	0.26	28	9	0.0004	None	NA	NA	-
1.3	271	1.3	155	42	$1x10^{-20}$	None	NA	NA	-
1.8	246	1.8	134	62	$7x10^{-8}$	1.3	149	70	$2x10^{-8}$
2.3	195	2.3	70	42	0.003	1.8	101	77	0.03
3.2	235	4	84	57	0.009	3.2	93	82	0.2
4	208	4	48	34	0.05	4	48	38	0.1
4.5	204	4.5	110	44	$2x10^{-8}$	4.5	118	52	$1x10^{-07}$
5	183	5	60	46	0.07	5	51	38	0.07
5.5	157	5.5	45	45	0.5	5.5	37	29	0.1
5.8	0	None	NA	NA	-	None	NA	NA	-
5.9	0	None	NA	NA	-	None	NA	NA	-
6.1	13	None	NA	NA	-	None	NA	NA	-

Additional data file S1 (separate file)

Information from *P. berghei* Smart-seq2 analyses including summary statistics for each gene, cluster assignment, highly variable genes, marker genes, and unique core assignments (Gene info tab). Additional tabs include the differential expression analyses, gene ontology and motif discovery results.

Additional data file S2 (separate file)

Summary statistics of the down-sampled Smart-seq2 data set. Sixty cells from each stage that were selected for the core unique analysis along with their quality control metrics, the summary statistics for each gene (*e.g.* total transcript counts), and the mean expression for each gene across each of the 10 stages are reported.

Additional Data file S3 (separate file)

One-to-one orthologs across ten *Plasmodium* species. These orthologs were used for mapping cells across species.

Additional Data file S4 (separate file)

Differential expression analysis over IDC development in three species. Slopes and FDR-corrected p-values (qval) for each of the 306 conserved genes across each of the 13 segments. Segments are labelled by their mid-point pseudotime and species are labelled pb (*P. berghei*), pk (*P. knowlesi*), and pf (*P. falciparum*) respectively. In addition, there is a column for each segment indicating if the Pk[Pf] slope was significant and agreed with Pb (1) or not (0). Additional sheets provide slope results for all RNA velocity genes in each species.

References

1. A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, S. A. Teichmann, The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015). doi:10.1016/j.molcel.2015.04.005 Medline

2. X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, G. Guo, Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018). doi:10.1016/j.cell.2018.02.001 Medline

3. C. T. Fincher, O. Wurtzel, T. de Hoog, K. M. Kravarik, P. W. Reddien, Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**, eaaq1736 (2018). doi:10.1126/science.aaq1736 Medline

4. T. D. Otto, U. Böhme, A. P. Jackson, M. Hunt, B. Franke-Fayard, W. A. M. Hoeijmakers, A. A. Religa, L. Robertson, M. Sanders, S. A. Ogun, D. Cunningham, A. Erhart, O. Billker, S. M. Khan, H. G. Stunnenberg, J. Langhorne, A. A. Holder, A. P. Waters, C. I. Newbold, A. Pain, M. Berriman, C. J. Janse, A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol.* **12**, 86 (2014). doi:10.1186/s12915-014-0086-0 Medline

5. A. J. Reid, A. M. Talman, H. M. Bennett, A. R. Gomes, M. J. Sanders, C. J. R. Illingworth, O. Billker, M. Berriman, M. K. N. Lawniczak, Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *eLife* **7**, e33105 (2018). doi:10.7554/eLife.33105 Medline

6. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 1802.03426 [stat.ML] (6 December 2018).

7. K. K. Hanson, S. March, S. Ng, S. N. Bhatia, M. M. Mota, In vitro alterations do not reflect a requirement for host cell cycle progression during Plasmodium liver stage infection. *Eukaryot. Cell* **14**, 96–103 (2015). doi:10.1128/EC.00166-14 Medline

8. J. Shi, J. Malik, Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000). doi:10.1109/34.868688

9. T. Kariu, T. Ishino, K. Yano, Y. Chinzei, M. Yuda, CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. *Mol. Microbiol.* **59**, 1369–1379 (2006). doi:10.1111/j.1365-2958.2005.05024.x Medline

10. M. J. Stewart, R. J. Nawrot, S. Schulman, J. P. Vanderberg, *Plasmodium berghei* sporozoite invasion is blocked in vitro by sporozoite-immobilizing antibodies. *Infect. Immun.* **51**, 859–864 (1986). Medline

11. E. Bushell, A. R. Gomes, T. Sanderson, B. Anar, G. Girling, C. Herd, T. Metcalf, K. Modrzynska, F. Schwach, R. E. Martin, M. W. Mather, G. I. McFadden, L. Parts, G. G. Rutledge, A. B. Vaidya, K. Wengelnik, J. C. Rayner, O. Billker, Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260–272.e8 (2017). doi:10.1016/j.cell.2017.06.030 Medline

12. K. Modrzynska, C. Pfander, L. Chappell, L. Yu, C. Suarez, K. Dundas, A. R. Gomes, D. Goulding, J. C. Rayner, J. Choudhary, O. Billker, A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the *Plasmodium* Life Cycle. *Cell Host Microbe* **21**, 11–22 (2017). doi:10.1016/j.chom.2016.12.003 Medline

13. J. A. Young, J. R. Johnson, C. Benner, S. F. Yan, K. Chen, K. G. Le Roch, Y. Zhou, E. A. Winzeler, In silico discovery of transcription regulatory elements in *Plasmodium falciparum*. *BMC Genomics* **9**, 70 (2008). doi:10.1186/1471-2164-9-70 Medline

14. J. Guizetti, A. Scherf, Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cell. Microbiol.* **15**, 718–726 (2013). doi:10.1111/cmi.12115 Medline

15. N. Rovira-Graells, A. P. Gupta, E. Planet, V. M. Crowley, S. Mok, L. Ribas de Pouplana, P. R. Preiser, Z. Bozdech, A. Cortés, Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.* **22**, 925–938 (2012). doi:10.1101/gr.129692.111 Medline

16. T. Brugat, A. J. Reid, J. Lin, D. Cunningham, I. Tumwine, G. Kushinga, S. McLaughlin, P. Spence, U. Böhme, M. Sanders, S. Conteh, E. Bushell, T. Metcalf, O. Billker, P. E. Duffy, C. Newbold, M. Berriman, J. Langhorne, Antibody-independent mechanisms regulate the establishment of chronic *Plasmodium* infection. *Nat. Microbiol.* **2**, 16276 (2017). doi:10.1038/nmicrobiol.2016.276 Medline

17. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi:10.1038/ncomms14049 Medline

18. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi:10.1038/nbt.4096 Medline

19. V. Y. Kiselev, A. Yiu, M. Hemberg, scmap: Projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018). doi:10.1038/nmeth.4644 Medline

20. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi:10.1038/s41586-018-0414-6 Medline

21. H. J. Painter, N. C. Chung, A. Sebastian, I. Albert, J. D. Storey, M. Llinás, Genome-wide real-time in vivo transcriptional dynamics during Plasmodium falciparum blood-stage development. *Nat. Commun.* **9**, 2656 (2018). doi:10.1038/s41467-018-04966-3 Medline

22. R. Hoo, L. Zhu, A. Amaladoss, S. Mok, O. Natalang, S. A. Lapp, G. Hu, K. Liew, M. R. Galinski, Z. Bozdech, P. R. Preiser, Integrated analysis of the *Plasmodium* species transcriptome. *EBioMedicine* **7**, 255–266 (2016). doi:10.1016/j.ebiom.2016.04.011 Medline

23. L. M. Yeoh, C. D. Goodman, V. Mollard, G. I. McFadden, S. A. Ralph, Comparative transcriptomics of female and male gametocytes in *Plasmodium berghei* and the evolution of sex in alveolates. *BMC Genomics* **18**, 734 (2017). doi:10.1186/s12864-017-4100-0 Medline

24. H. J. Lee, A. Georgiadou, M. Walther, D. Nwakanma, L. B. Stewart, M. Levin, T. D. Otto, D. J. Conway, L. J. Coin, A. J. Cunnington, Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci. Transl. Med.* **10**, eaar3619 (2018). doi:10.1126/scitranslmed.aar3619 Medline

25. A. Poran, C. Nötzel, O. Aly, N. Mencia-Trinchant, C. T. Harris, M. L. Guzman, D. C. Hassane, O. Elemento, B. F. C. Kafsack, Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95–99 (2017). doi:10.1038/nature24280 Medline

26. U. Böhme, T. D. Otto, J. A. Cotton, S. Steinbiss, M. Sanders, S. O. Oyola, A. Nicot, S. Gandon, K. P. Patra, C. Herd, E. Bushell, K. K. Modrzynska, O. Billker, J. M. Vinetz, A. Rivero, C. I. Newbold, M. Berriman, Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.* **28**, 547–560 (2018). doi:10.1101/gr.218123.116 Medline

27. V. M. Howick, A. J. Reid, *vhowick/MalariaCellAtlas: Malaria Cell Atlas* (2019); http://dx.doi.org/10.5281/zenodo.2843883.

28. Malaria Cell Atlas; www.sanger.ac.uk/science/tools/mca/mca/.

29. P.-C. Burda, M. A. Roelli, M. Schaffner, S. M. Khan, C. J. Janse, V. T. Heussler, A *Plasmodium* phospholipase is involved in disruption of the liver stage parasitophorous vacuole membrane. *PLOS Pathog.* **11**, e1004760 (2015). doi:10.1371/journal.ppat.1004760 Medline

30. R. W. Moon, J. Hall, F. Rangkuti, Y. S. Ho, N. Almond, G. H. Mitchell, A. Pain, A. A. Holder, M. J. Blackman, Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 531–536 (2013). doi:10.1073/pnas.1216457110 Medline

31. A. R. Gomes, E. Bushell, F. Schwach, G. Girling, B. Anar, M. A. Quail, C. Herd, C. Pfander, K. Modrzynska, J. C. Rayner, O. Billker, A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. *Cell Host Microbe* **17**, 404–413 (2015). doi:10.1016/j.chom.2015.01.014 Medline

32. Y. Liu, R. Tewari, J. Ning, A. M. Blagborough, S. Garbom, J. Pei, N. V. Grishin, R. E. Steele, R. E. Sinden, W. J. Snell, O. Billker, The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev.* **22**, 1051–1068 (2008). doi:10.1101/gad.1656508 Medline

33. R. Carter, D. H. Chen, Malaria transmission blocked by immunisation with gametes of the malaria parasite. *Nature* **263**, 57–60 (1976). doi:10.1038/263057a0 Medline

34. W. J. Martin, J. Finerty, A. Rosenthal, Isolation of *Plasmodium berghei* (malaria) parasites by ammonium chloride lysis of infected erythrocytes. *Nat. New Biol.* **233**, 260–261 (1971). doi:10.1038/newbio233260a0 Medline

35. I. Siden-Kiamos, A. Ecker, S. Nybäck, C. Louis, R. E. Sinden, O. Billker, *Plasmodium berghei* calcium-dependent protein kinase 3 is required for ookinete gliding motility and mosquito midgut invasion. *Mol. Microbiol.* **60**, 1355–1363 (2006). doi:10.1111/j.1365-2958.2006.05189.x Medline

36. I. Kozarewa, Z. Ning, M. A. Quail, M. J. Sanders, M. Berriman, D. J. Turner, Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* **6**, 291–295 (2009). doi:10.1038/nmeth.1311 Medline

37. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015). doi:10.1038/nmeth.3317 Medline

38. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi:10.1093/bioinformatics/bts635 Medline

39. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi:10.1093/bioinformatics/btu638 Medline

40. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010). doi:10.1186/gb-2010-11-3-r25 Medline

41. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019). doi:10.1016/j.cels.2019.03.003 Medline

42. V. Y. Kiselev *et al.*, SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017). doi:10.1038/nmeth.4236

43. J. D. Welch, A. J. Hartemink, J. F. Prins, SLICER: Inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. *Genome Biol.* **17**, 106 (2016). doi:10.1186/s13059-016-0975-3 Medline

44. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi:10.1038/nbt.2859 Medline

45. F. Pedregosa *et al*., Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

46. W. Saelens, R. Cannoodt, Y. Saeys, A comprehensive evaluation of module detection methods for gene expression data. *Nat. Commun.* **9**, 1090 (2018). doi:10.1038/s41467-018-03424-4 Medline

47. M. Bastian, S. Heymann, M. Jacomy, in *Third International AAAI Conference on Weblogs and Social Media* (2009); www.aaai.org/ocs/index.php/ICWSM/09/paper/viewPaper/154.

48. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE* **9**, e98679 (2014). doi:10.1371/journal.pone.0098679 Medline

49. C. Aurrecoechea, J. Brestelli, B. P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, J. A. Miller, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, C. J. Stoeckert Jr., C. Treatman, H. Wang, PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009). doi:10.1093/nar/gkn814 Medline

50. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000 Res.* **5**, 2122 (2016). Medline

51. T. L. Bailey, DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011). doi:10.1093/bioinformatics/btr261 Medline

52. T. L. Campbell, E. K. De Silva, K. L. Olszewski, O. Elemento, M. Llinás, Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLOS Pathog.* **6**, e1001165 (2010). doi:10.1371/journal.ppat.1001165 Medline

53. I. Kaneko, S. Iwanaga, T. Kato, I. Kobayashi, M. Yuda, Genome-Wide Identification of the Target Genes of AP2-O, a *Plasmodium* AP2-Family Transcription Factor. *PLOS Pathog.* **11**, e1004905 (2015). doi:10.1371/journal.ppat.1004905 Medline

54. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007). doi:10.1186/gb-2007-8-2-r24 Medline

55. T. S. Andrews, M. Hemberg, M3Drop: Dropout-based feature selection for scRNASeq. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty1044 Medline

56. T. Carver, S. R. Harris, M. Berriman, J. Parkhill, J. A. McQuillan, Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469 (2012). doi:10.1093/bioinformatics/btr703 Medline

57. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002). doi:10.1093/nar/30.7.1575 Medline

58. R. C. Edgar, MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004). doi:10.1186/1471-2105-5-113 Medline

59. S. R. Eddy, Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **7**, e1002195 (2011). doi:10.1371/journal.pcbi.1002195 Medline

60. D. J. McCarthy, K. R. Campbell, A. T. L. Lun, Q. F. Wills, Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017). Medline

61. L. Li, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003). doi:10.1101/gr.1224503 Medline

62. M. Baron, A. Veres, S. L. Wolock, A. L. Faust, R. Gaujoux, A. Vetere, J. H. Ryu, B. K. Wagner, S. S. Shen-Orr, A. M. Klein, D. A. Melton, I. Yanai, A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346–360.e4 (2016). doi:10.1016/j.cels.2016.08.011 Medline

63. J. T. Dessens, I. Sidén-Kiamos, J. Mendoza, V. Mahairaki, E. Khater, D. Vlachou, X.-J. Xu, F. C. Kafatos, C. Louis, G. Dimopoulos, R. E. Sinden, SOAP, a novel malaria ookinete protein involved in mosquito midgut invasion and oocyst development. *Mol. Microbiol.* **49**, 319–329 (2003). doi:10.1046/j.1365-2958.2003.03566.x Medline

64. M. del Carmen Rodriguez, P. Gerold, J. Dessens, K. Kurtenbach, R. T. Schwartz, R. E. Sinden, G. Margos, Characterisation and expression of pbs25, a sexual and sporogonic stage specific protein of *Plasmodium berghei*. *Mol. Biochem. Parasitol.* **110**, 147–159 (2000). doi:10.1016/S0166-6851(00)00265-6 Medline

65. S. Marguerat, J. Bähler, Coordinating genome expression with cell size. *Trends Genet.* **28**, 560–565 (2012). doi:10.1016/j.tig.2012.07.003 Medline

66. M. Andreadaki, R. N. Morgan, E. Deligianni, T. W. A. Kooij, J. M. Santos, L. Spanos, K. Matuschewski, C. Louis, G. R. Mair, I. Siden-Kiamos, Genetic crosses and complementation reveal essential functions for the *Plasmodium* stage-specific actin2 in sporogonic development. *Cell. Microbiol.* **16**, 751–767 (2014). doi:10.1111/cmi.12274 Medline

67. M. Roques, R. J. Wall, A. P. Douglass, A. Ramaprasad, D. J. P. Ferguson, M. L. Kaindama, L. Brusini, N. Joshi, Z. Rchiad, D. Brady, D. S. Guttery, S. P. Wheatley, H. Yamano, A. A. Holder, A. Pain, B. Wickstead, R. Tewari, Plasmodium P-Type Cyclin CYC3 Modulates Endomitotic Growth during Oocyst Development in Mosquitoes. *PLOS Pathog.* **11**, e1005273 (2015). doi:10.1371/journal.ppat.1005273 Medline

68. T. Ishino, B. Boisson, Y. Orito, C. Lacroix, E. Bischoff, C. Loussert, C. Janse, R. Ménard, M. Yuda, P. Baldacci, LISP1 is important for the egress of *Plasmodium berghei* parasites from liver cells. *Cell. Microbiol.* **11**, 1329–1339 (2009). doi:10.1111/j.1462-5822.2009.01333.x Medline

69. D. R. Drew, P. R. Sanders, B. S. Crabb, *Plasmodium falciparum* merozoite surface protein 8 is a ring-stage membrane protein that localizes to the parasitophorous vacuole of infected erythrocytes. *Infect. Immun.* **73**, 3912–3922 (2005). doi:10.1128/IAI.73.7.3912-3922.2005 Medline

70. S. Das, N. Hertrich, A. J. Perrin, C. Withers-Martinez, C. R. Collins, M. L. Jones, J. M. Watermeyer, E. T. Fobes, S. R. Martin, H. R. Saibil, G. J. Wright, M. Treeck, C. Epp, M. J. Blackman, Processing of *Plasmodium falciparum* Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. *Cell Host Microbe* **18**, 433–444 (2015). doi:10.1016/j.chom.2015.09.007 Medline

71. A. M. Talman, J. H. Prieto, S. Marques, C. Ubaida-Mohien, M. Lawniczak, M. N. Wass, T. Xu, R. Frank, A. Ecker, R. S. Stanway, S. Krishna, M. J. E. Sternberg, G. K. Christophides, D. R. Graham, R. R. Dinglasan, J. R. Yates 3rd, R. E. Sinden, Proteomic analysis of the *Plasmodium* male gamete reveals the key role for glycolysis in flagellar motility. *Malar. J.* **13**, 315 (2014). doi:10.1186/1475-2875-13-315 Medline

72. A.-K. Mueller, N. Camargo, K. Kaiser, C. Andorfer, U. Frevert, K. Matuschewski, S. H. I. Kappe, *Plasmodium* liver stage developmental arrest by depletion of a protein at the parasite-host interface. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3022–3027 (2005). doi:10.1073/pnas.0408442102 Medline

73. U. L. Rathnapala, C. D. Goodman, G. I. McFadden, A novel genetic technique in *Plasmodium berghei* allows liver stage analysis of genes required for mosquito stage development and demonstrates that de novo heme synthesis is essential for liver stage

development in the malaria parasite. *PLOS Pathog.* **13**, e1006396 (2017). doi:10.1371/journal.ppat.1006396 Medline

74. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi:10.1186/s13059-017-1382-0 Medline

75. M. J. López-Barragán, J. Lemieux, M. Quiñones, K. C. Williamson, A. Molina-Cruz, K. Cui, C. Barillas-Mury, K. Zhao, X. Z. Su, Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* **12**, 587 (2011). doi:10.1186/1471-2164-12-587 Medline

76. S. A. Lapp, S. Mok, L. Zhu, H. Wu, P. R. Preiser, Z. Bozdech, M. R. Galinski, *Plasmodium knowlesi* gene expression differs in ex vivo compared to in vitro blood-stage cultures. *Malar. J.* **14**, 110 (2015). doi:10.1186/s12936-015-0612-8 Medline

77. M. Ponzi, I. Sidén-Kiamos, L. Bertuccini, C. Currà, H. Kroeze, G. Camarda, T. Pace, B. Franke-Fayard, E. C. Laurentino, C. Louis, A. P. Waters, C. J. Janse, P. Alano, Egress of *Plasmodium* berghei gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell. Microbiol.* **11**, 1272–1288 (2009). doi:10.1111/j.1462-5822.2009.01331.x Medline

78. E. Lasonder, S. R. Rijpma, B. C. L. van Schaijk, W. A. M. Hoeijmakers, P. R. Kensche, M. S. Gresnigt, A. Italiaander, M. W. Vos, R. Woestenenk, T. Bousema, G. R. Mair, S. M. Khan, C. J. Janse, R. Bártfai, R. W. Sauerwein, Integrated transcriptomic and proteomic analyses of *P. falciparum* gametocytes: Molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res.* **44**, 6087–6101 (2016). doi:10.1093/nar/gkw536 Medline