SUPPLEMENTAL INFORMATION

The variantome of the malaria parasite *Plasmodium falciparum*

Núria Rovira-Graells, Archna P. Gupta, Evarist Planet, Valerie M. Crowley, Sachel Mok, Lluís Ribas de Pouplana, Peter R. Preiser, Zbynek Bozdech and Alfred Cortés

SUPPLEMENTAL RESULTS

Expression of *var*, *rif*, *stevor* and *pfmc-2tm* genes.

76 of the 138 genes showing variant expression in 3D7-derived parasite lines were members of the large hypervariable gene families *var*, *rif*, *stevor* and *pfmc-2tm*. Some pseudogenes from these gene families also showed variant expression (Supplemental Table S1). Only one of 44 type-B *rif* genes (*rifB*) analyzed showed variant expression, in contrast to 21 of 103 type-A *rif* (*rifA*) (Supplemental Table S1), suggesting that different regulatory mechanisms may apply for these two types of *rif* genes. Functional and transcriptional differences have been previously proposed for these two *rif* subfamilies (Petter et al. 2007; Joannin et al. 2008; Wang et al. 2009).

To identify the predominantly expressed members of these gene families, we analyzed normalized Cy5 signal intensities (rather than the ratio against the reference pool), which provides semi-quantitative estimates of the level of expression of these genes (Methods and Supplemental Fig. S2). As expected from the well-established mutually exclusive expression of *var* genes, recently subcloned parasite lines expressed only one predominant *var* gene, with the exception of the 10G subclone in which no dominant transcript was detected. In contrast, 3D7-A and 3D7-B expressed more than one *var* gene at high levels, as expected for parasites lines that have been maintained in culture for a long time (Dzikowski et al. 2006) (Supplemental Fig. S2A). The dominant *var* genes were different between 3D7-A and 3D7-B.

In contrast to this result, 3D7-A and 3D7-B showed similar patterns of dominant *rif* genes: of the five most highly expressed *rif* in 3D7-A, four were among the five most highly expressed in 3D7-B, suggesting that the dominantly expressed *rif* are remarkably stable (Supplemental Fig. S2B). 3D7-A subclones also showed similar patterns of dominant expressed *rif* genes. This indicates that *rif* genes undergo frequent expression switches, as many *rif* were found to be variantly expressed (Supplemental Table S1), but the dominant expressed genes remain relatively unchanged, as previously suggested (Cabral and Wunderlich 2009). All five dominant *rif* were of the *rifA* type. The most highly expressed *rif* in all 3D7 parasite lines was PFD0070c, which belongs to the divergent *rif* subgroup *rifA2*. The observation that *var2csa* was only expressed at low levels in all parasite lines does not support the proposed possible association between expression of *rifA2* and *var2csa* genes (Wang et al. 2009). All recently subcloned parasite lines expressed several *rif* genes, confirming previous observations that there is no mutually exclusive expression in this gene family (Petter et al. 2007). Gene families *stevor* and *pfmc-2tm* revealed more dynamic expression patterns than *rif*, with several genes expressed at higher levels in 3D7-B than in 3D7-A (Supplemental Fig. S2C-D). There was no evidence for co-activation of *var*, *rif*, *stevor* and *pfmc-2tm* genes located in the same subtelomeric or central cluster, as the predominantly expressed genes from the different families are in separate chromosomal locations (Supplemental Fig. S3).

Genes excluded from the list of variant genes in the D10 comparison.

Only in the case of D10-derived parasite lines, samples from parasites prepared on the same date clustered together and revealed a large number of genes with expression levels that depended on preparation date (Supplemental Fig. S5). All of these genes showed one of two alternative expression patterns, with differences observed in the second half of the life cycle: either expression was higher in parasite lines E3 and G4 (cluster 1), or expression was higher in D10, F1 and G2 (cluster 2) (Supplemental Fig. S5). RNAs from E3 and G4 were obtained from cultures prepared in parallel, as were RNAs from D10, F1 and G2. Furthermore, changes in the expression of some of these genes were not reproducible in independent parasite preparations (not shown). Comparison with expression levels in 7G8 and HB3A parasite lines revealed that D10-F1-G2, and not E3-G4, had altered expression patterns.

Differences in the expression of these genes are attributable to uncontrollable differences in culture conditions between different dates, such as use of erythrocytes from different donors. The specific conditions that resulted in the large transcriptional alteration in D10-F1-G2 remain enigmatic. Gene set enrichment analysis (GSEA) (Subramanian et al. 2005) revealed that genes expressed at higher levels in E3 and G4 were enriched in specific transcription factors and translation initiation factors, whereas genes with opposite expression patterns were enriched in tRNAs, ribosome proteins and mitochondria/apicoplast genes (data not shown).

Genes in clusters 1 and 2 were excluded from the list of D10 variant genes (Supplemental Fig. S5 and Supplemental Table S1), because they showed changes in expression levels attributable to uncontrolled factors, rather than authentic stable clonally transmitted variant expression. Genes for which variant expression could be explained by the anomalous pattern in D10 were also removed from the list of variant genes in the parentals comparison (Supplemental Table S1).

Variant expression of early gametocyte markers and exclusion of genes expressed in gametocytes from the lists of variant genes.

We addressed the possibility that presence of gametocytes (sexual stages) in different proportions between different parasite lines might be responsible for some of the transcript level differences observed. Microscopy inspection of the parasite samples used for microarray analysis revealed presence of early gametocytes only in 7G8-derived parasite lines. Early gametocytes (stages I and II) were identified as trophozoite-like forms that persisted throughout the asexual cycle. Gametocytes at later stages of development were not observed in any of the samples. Consistent with these observations, expression patterns for a large set of gametocyte-specific markers (all gametocyte stages) (Young et al. 2005) did not reveal higher or lower expression of the majority of genes in any of the parasite lines (Supplemental Fig. S6A). However, early gametocytes are characterized by upregulation of a small number of transcripts, and the only six *bona fide* early gametocyte markers (Olivieri et al. 2009) were consistently expressed at higher levels in the 7G8 parental line than in the other parental lines. Furthermore, these six genes were expressed at lower levels in LD10 compared to the other 7G8-derived parasite lines (Supplemental Fig. S6B). Hence, higher expression in 7G8 than in D10 and HB3 parental lines, and lower expression in LD10 than in the

other 7G8-derived parasite lines, constitutes a signature for genes potentially expressed by early gametocytes. Genes showing this signature were excluded from the lists of variant genes in the 7G8 and parentals comparisons, because differences in their transcript levels may be attributable to different gametocyte abundance, rather than clonally variant expression (Supplemental Fig. S6C and Supplemental Table S1).

Some of the early gametocyte markers also showed variant transcript levels in other comparisons (Supplemental Fig. S6B). For instance, three of the six early gametocyte markers were clearly upregulated in 3D7-B compared to other 3D7-derived parasite lines. There are two possible interpretations for this observation: i) different transcript levels for some early gametocyte markers are attributable to expression from gametocytes occurring below the microscopy detection threshold; ii) some of the early gametocyte transcripts are variantly expressed in asexual parasites. While we can not conclusively distinguish between the two possibilities, some observations strongly support the latter: i) apart from parasite lines in the 7G8 and parentals comparisons, no parasite line showed consistent up- or down-regulation of the six markers (Supplemental Fig. S6B), or of the majority of genes showing the early gametocyte signature in the 7G8 and parentals comparisons (Supplemental Fig. S6C); ii) in a gametocyte production assay, 3D7-A completely failed to produce gametocytes, whereas 3D7-B produced a very low amount compared to 7G8 parasite lines (not shown). Consistent with the possibility that some early gametocyte markers may be expressed by asexual parasites, 3D7-A showed significant expression of four of the early gametocyte markers (Supplemental Fig. S7); iii) the majority of early gametocyte markers showed strong stage-specific expression along the time-course in non-7G8-derived parasite lines, whereas relatively stable expression would be expected if these genes were expressed only by gametocytes (Supplemental Fig. S7). In fact, the same genes showed relatively stable expression in 7G8, where gametocytes are the major source of these transcripts. All together, our data suggests that some early gametocyte markers are expressed during asexual stages, with variant expression. The expression level of some of these genes in asexual parasites may have a role in determining the propensity to differentiate into gametocytes. Genes that showed the early gametocyte signature in the 7G8 and parentals comparisons and were variantly expressed in other parasite lines are indicated in Supplemental Table S1.

SUPPLEMENTAL METHODS

Heat-shock experiments.

To measure survival to heat-shock, 3D7-A, 10G and 1.2B parasite lines (Fig. 1A) were sorbitol-synchronized and parasitemia adjusted to 1%. Heat-shock was performed 22 h after sorbitol, when all parasites were at the trophozoite stage, by transferring cultures to an incubator at 41.5ºC for 3 h. Control cultures were maintained in the 37ºC incubator all the time. Parasitemia at the next generation was determined by FACS when all schizonts had burst (typically we measured parasitemia about 60 h after sorbitol treatment, because in addition to reducing culture viability, heat-shock delays cycle progression) (Blair et al. 2002). FACS determination of parasitemia was performed using SYTO 11 essentially as previously described (Urbán et al. 2011).

To study adaptation of parasites to heat-shock, the three parasite lines were subjected to periodic heat-shock for five consecutive generations. Heat-shock was always performed 22 h after sorbitol treatment. At each cycle, 53.5 h after sorbitol treatment an aliquot of each heat-shocked culture was separated and maintained in culture to measure parasitemia later when all schizonts had burst. The rest of the culture was sorbitol-synchronized, parasitemia was adjusted to 1% by diluting with uninfected erythrocytes, and adjusted cultures were split in two dishes, one to serve as control and one for next cycle heat-shock. With this protocol, growth of heat-shocked parasites was compared at each cycle with non-heat-shocked controls of identical initial parasitemia, and cultures were kept synchronized throughout the experiment. We found that parasites cultured in erythrocytes that had been stored for more than 2-3 weeks were more sensitive to heat-shock than parasites cultured in fresh erythrocytes (not shown). Consequently, we avoided using old erythrocytes for the adaptation experiment.

Microarray reannotation and exclusion of probes.

The microarray used in this study, which was designed against *P. falciparum* 3D7 genome (PlasmoDB 4.4 release) (Hu et al. 2007) was reannotated against the 3D7 PlasmoDB 7.0 genome release. The oligonucleotides in the array were blasted against PlasmoDB 7.0 transcripts (excluding non-coding RNAs) using blast 2.2 default parameters. Based on the comparison of blast scores with CGH values (Supplemental Fig. S4A), we selected stringent cut-off blast scores to exclude probes with low first hit blast scores (<100, 70 oligonucleotides) or high second hit blast scores (>70, 258

oligonucleotides), because they may not hybridize efficiently or they may cross-react with other genes, respectively. Furthermore, 424 of the remaining probes were reassigned to a new gene ID (Supplemental Table S5).

Additional probes were excluded for the analysis of parasites genetically different from 3D7, as some probes fall within highly polymorphic regions of the genome. All oligonucleotides in the microarray were blasted against supercontigs of the incomplete genomes of HB3, D10 and 7G8 (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp) (Volkman et al. 2007). Probes for the gene families *var*, *rif*, *stevor* and *pfmc-2tm* were excluded, because both blast and CGH analysis revealed that this microarray is not suitable for the analysis of these hypervariable gene families in non-3D7 parasite lines (Supplemental Fig. S4B-C). Furthermore, for each non-3D7 genetic background we excluded probes with low blast score and low CGH value that met three conditions: i) blast score lower than 100; ii) average CGH result [$\log_2$(Cy5/Cy3)] among the parasite lines of the given genetic background lower than -1, and individual CGH result lower than -0.75 in all or all but one of the parasite lines; iii) low Cy5 intensity in the transcriptional analysis of all parasite lines of the given genetic background (average Cy5 intensity <500). Only probes that fulfilled these three criteria were eliminated. The latter condition (iii) was introduced empirically because probes that fulfilled the other two conditions but hybridized with high intensity, invariably revealed identical expression patterns to other probes for the same gene that were not excluded. Probes with a high second hit blast score in non-3D7 parasites are potentially cross-reactive, but they were not eliminated because we found that, in most cases, high second blast hit scores are attributable to redundancy in the supercontigs of the incomplete genomes. A very small number of probes were excluded based on alignments of individual genes.

By applying these criteria, 34, 61 and 22 probes were excluded from the analysis of HB3, D10 and 7G8 parasites, respectively (Supplemental Table S5). We performed stringent tests for the criteria used to exclude probes, and found that they resulted on correct exclusion of probes even for genes that are polymorphic between parasite isolates, but have very similar paralogs (Supplemental Fig. S12A). Our analysis also revealed that artifacts related with genetic polymorphism are unlikely to be a source of false positives (Supplemental Fig. S12B).

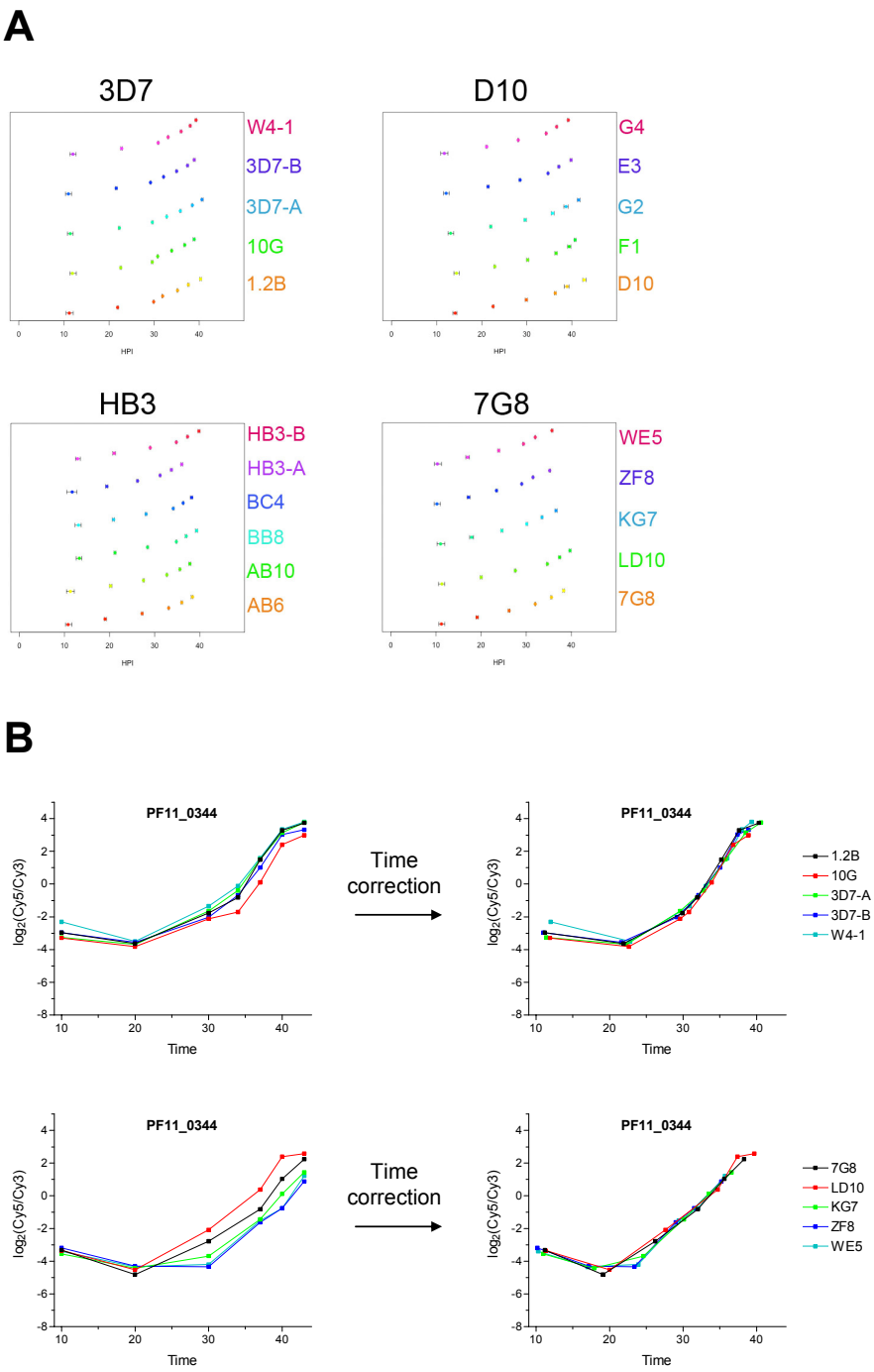
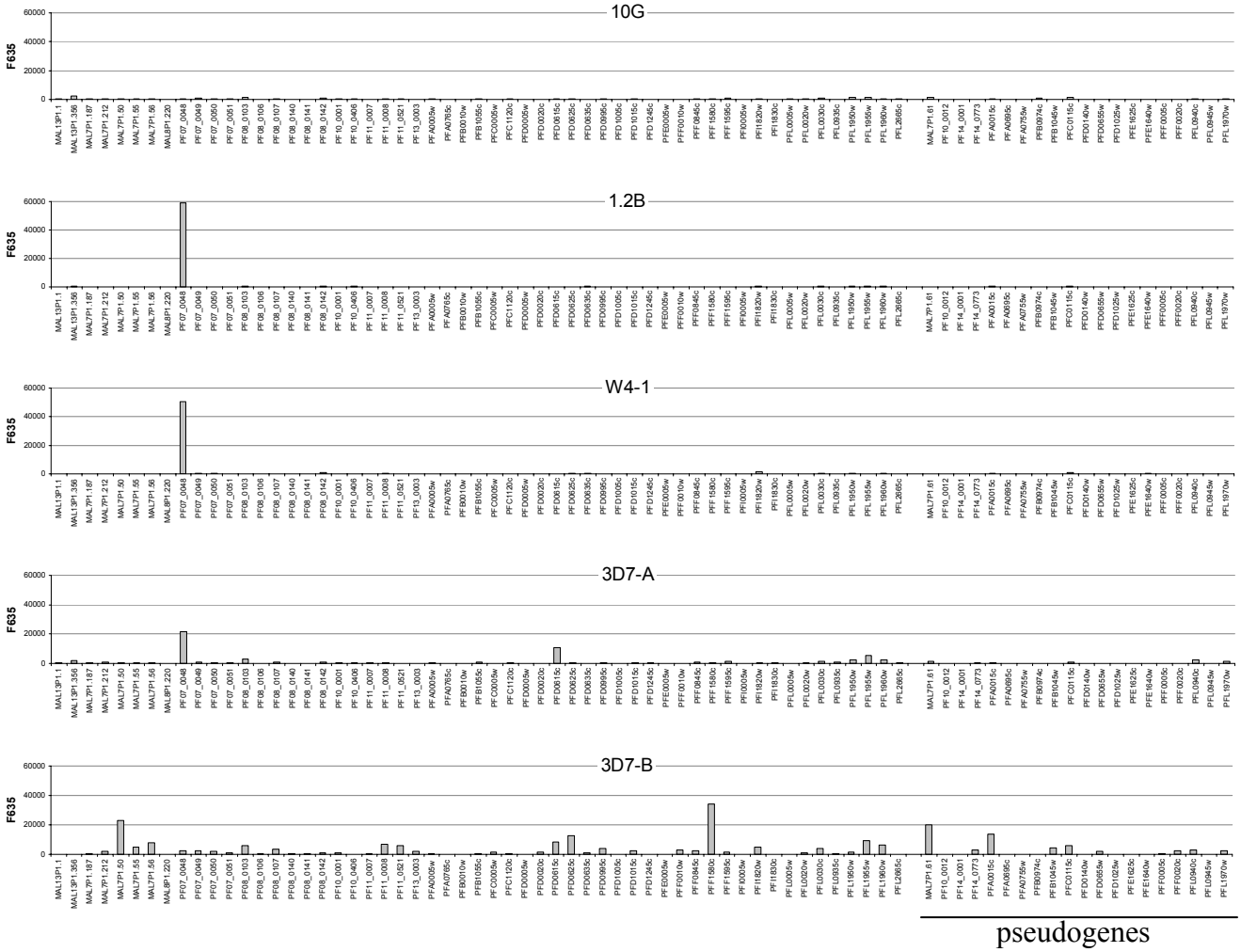


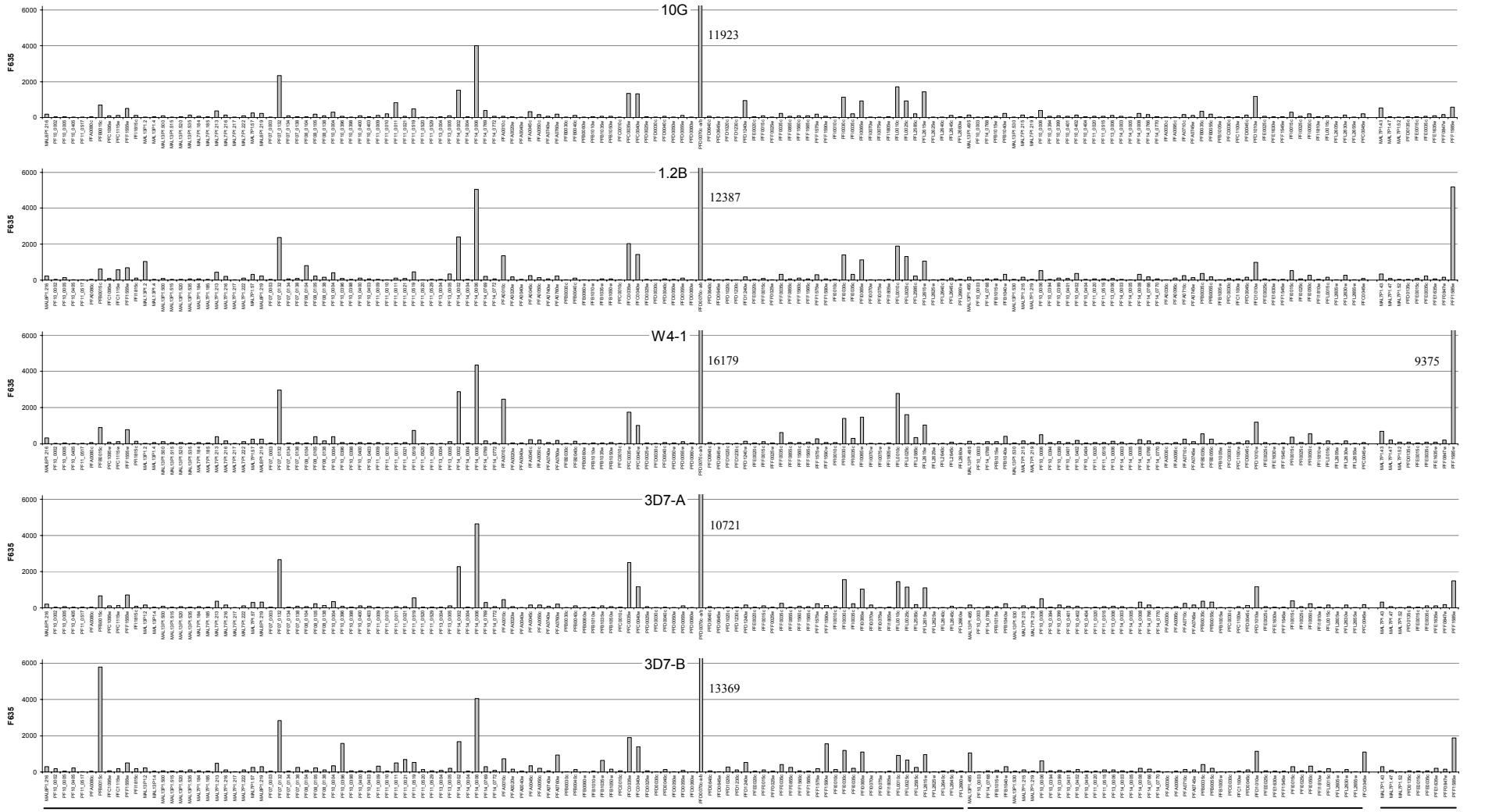
Figure S1. Estimation of parasite age. **(A)** Parasite age (in hours post-invasion) was estimated for all samples against a one hour-resolution time course (Bozdech et al. 2003) (HB3 transcriptome), using a published statistical likelihood-based method (Lemieux et al. 2009). Error bars are 95% confidence intervals. For each parasite line, the first value is for parasites collected 10-15 h post-invasion, followed by parasites collected at 20-25, 30-35, 34-39 (only 3D7 parasites), 37-42, 40-45 and 43-48 h post-invasion. **(B)** Expression levels ($\log_2$ ratio of expression relative to the reference pool) for the gene *ama1*, commonly used as a control for schizont-specific expression, were plotted against experimental times (left panels) or against statistically-estimated times (right panels). Upper and lower panels correspond to the 3D7 and 7G8 comparisons, respectively. In both cases, apparent expression differences attributable to small differences in parasite age are eliminated by using estimated parasite age.

A



var

B


10G
11923
1.2B
12387
W4-1
16179
9375
3D7-A
10721
3D7-B
13369
F635
A-type
B-type
pseudogenes

rif

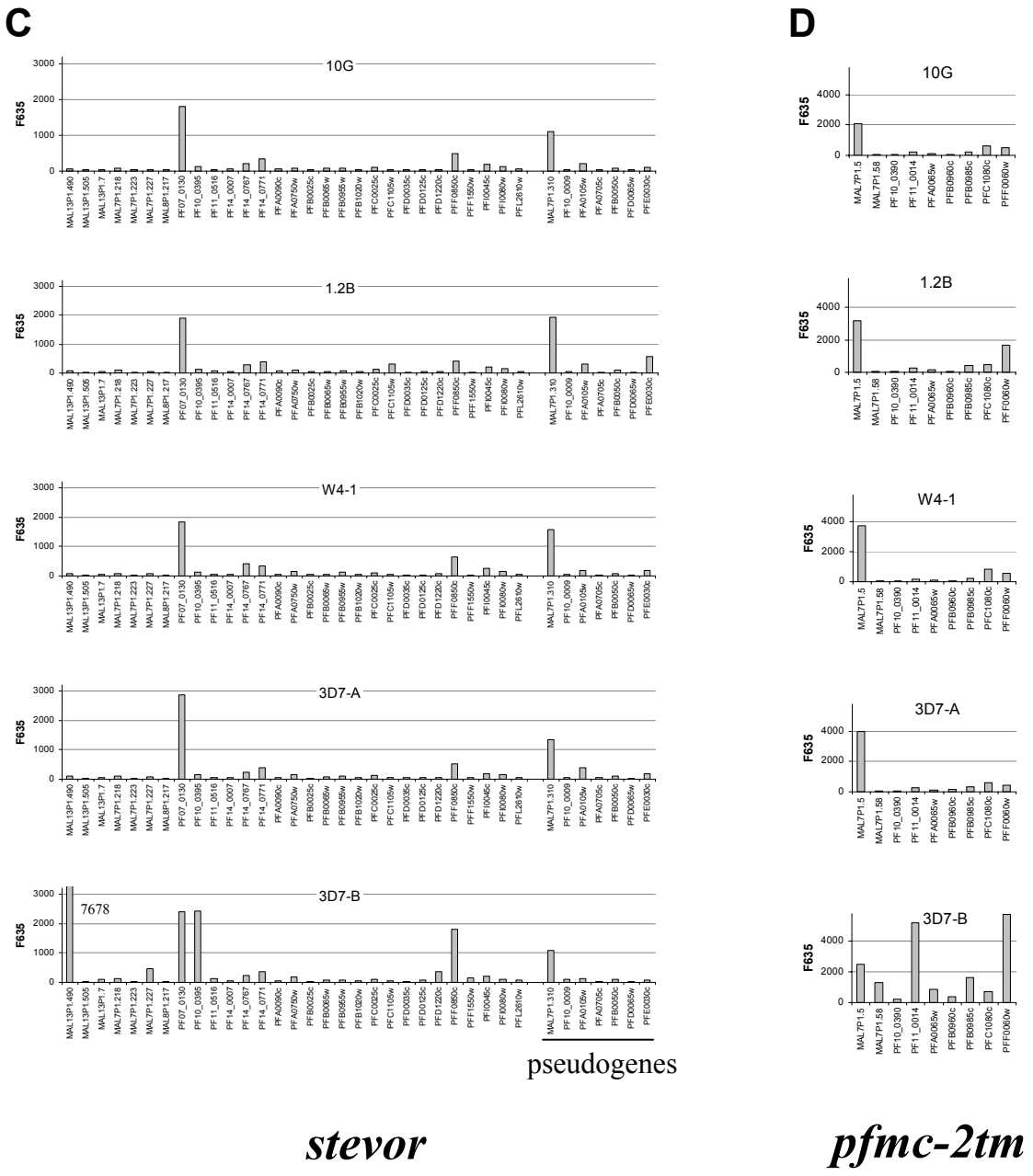


Figure S2. Expression of *var* **(A)**, *rif* **(B)**, *stevor* **(C)** and *pfmc-2tm* **(D)** genes in 3D7-derived parasite lines. For *var* genes, values are normalized Cy5 intensity at 10-15 h post-invasion. For other gene families, values are the average of normalized Cy5 intensity at the two time points with highest expression, because genes from these families have heterogeneous patterns of expression along the asexual life cycle.

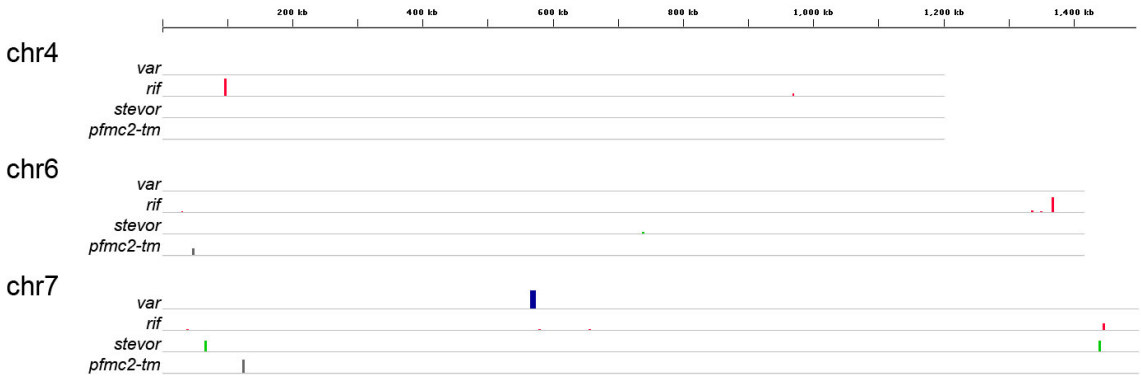


Figure S3. Chromosomal position of expressed *var*, *rif*, *stevor* and *pfmc-2tm* genes in parasite line 1.2B. Values are normalized Cy5 intensity, as in Supplemental Fig S2. Only chromosomes with the most highly expressed genes from these families are shown. There was no co-activation of neighboring genes from these families in 1.2B or in any other 3D7-derived parasite line.

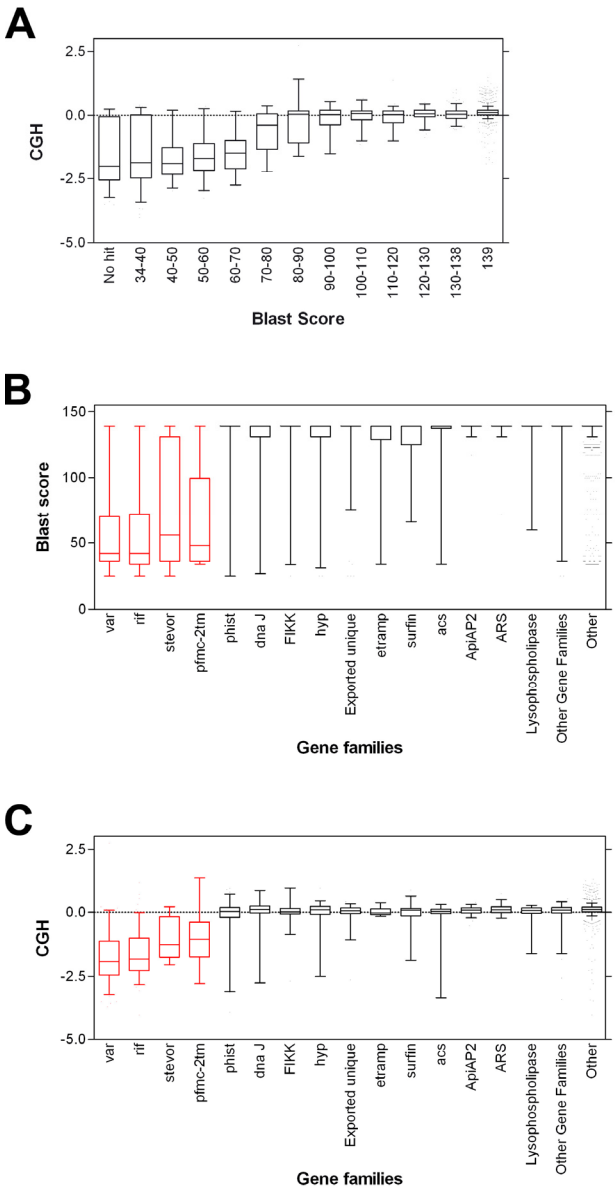


Figure S4. Blast and CGH analysis of the oligonucleotides in the microarray used for this study (Hu et al. 2007). CGH values are the average of the normalized $\log_2$ ratio of intensities [$\log_2$(Cy5/Cy3)] in the six HB3-derived parasite lines. Test gDNA samples (labeled with Cy5) were hybridized against a 3D7-A gDNA reference pool (labeled with Cy3). Blast scores are the score of the first blast hit against supercontigs of the incomplete HB3 genome (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp). Boxes are the 25th and 75th quartiles, and whiskers represent the 5th and 95th percentiles. **(A)** CGH values relative to blast scores. There is a progressive decrease in the hybridization efficiency for probes with a score <100. Probes with a score >70 give a hybridization signal. The higher 75th quartiles observed for probes with no blast hit or a blast score lower than 40, compared with probes with a blast score 40-70, are attributable to probes against regions of the genome that have not been sequenced yet in HB3. **(B)** Blast scores of probes for genes that belong to gene families. Values are shown for gene families of more than ten genes. The four gene families that were excluded from the transcriptional analysis of non-3D7 parasite lines are shown in red. "Other gene families" includes the smaller gene families *clag*, *eba*, *pfRh*, *acbp*, *msp7/msrp*, *msp3/6*, *hrp*, *gpb130/gbph*, *emp3*, *crmp*, alveolins, *ab_hyda/b*

and PFD0075-like, whereas "Other" refers to all genes that are not part of any of the gene families. **(C)** CGH data for probes from the same gene families as in panel B. Parallel analysis using CGH data from D10 and 7G8, and blast scores against their less complete genome sequences, gave similar results (not shown).

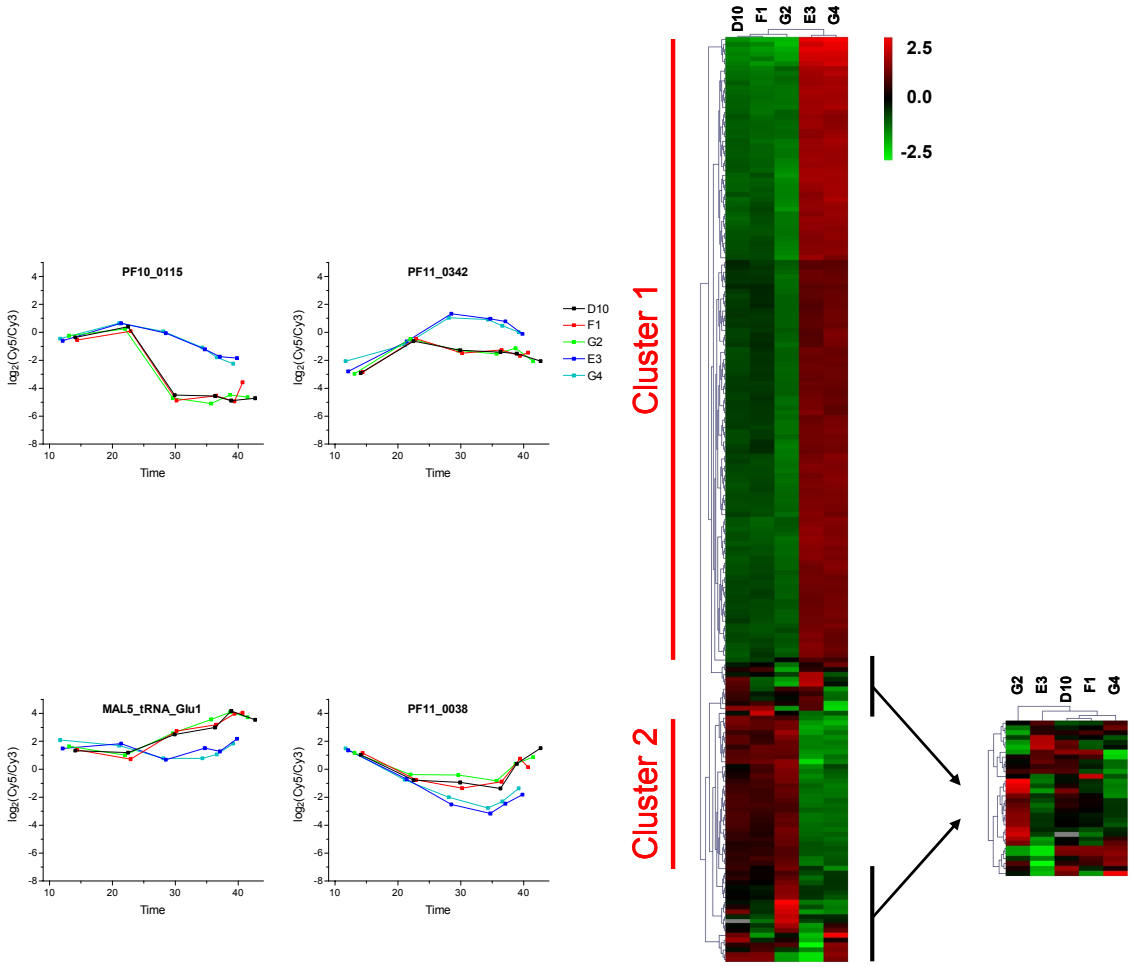


Figure S5. Genes excluded from the list of clonally variant genes in D10. Cluster analysis of genes showing different expression levels among D10-derived parasite lines. Two large clusters of genes with expression patterns that followed parasites preparation date and were not reproducible in independent parasite preparations were excluded from the list of D10 variant genes. Only genes shown in the small heat map were retained. Values in the heatmaps are the $\log_2$ of the expression fold-change relative to the average expression among all D10-derived lines. The plots on the left are representative examples of the expression patterns observed for genes in cluster 1 or cluster 2, showing differences in expression only in the second half of the asexual blood cycle.

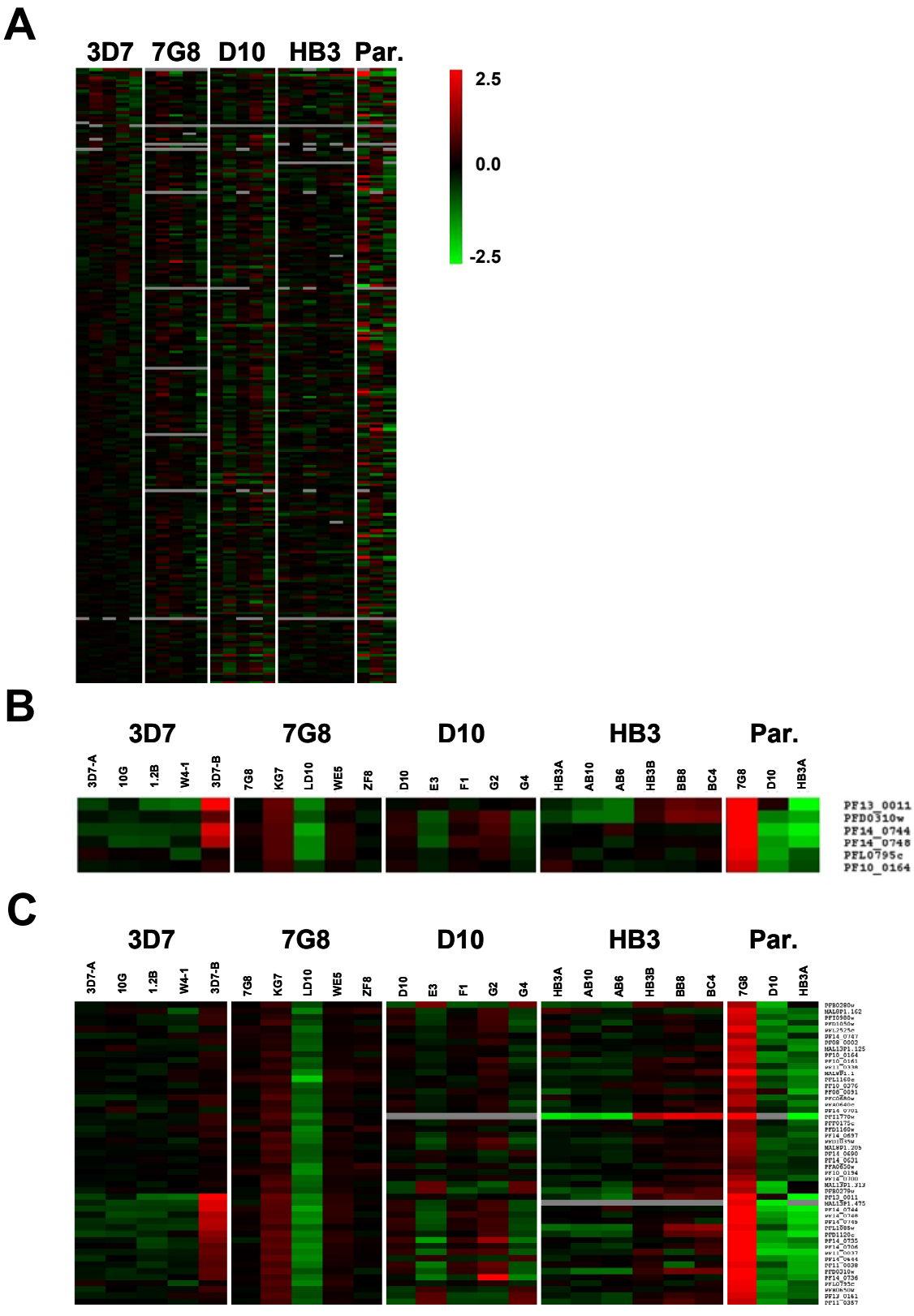


Figure S6. Expression patterns for genes related with gametocyte differentiation. Values are expression fold-change (in $\log_2$ scale) relative to the average of the parasite lines in each comparison. **(A)** Expression of 246 gametocyte-specific genes (all gametocyte stages) according to Young and collaborators (Young et al. 2005). None of the parasite lines showed higher or lower expression of the majority of the genes.

Analysis of two alternative sets of gametocyte-specific genes (Silvestrini et al. 2005; Silvestrini et al. 2010) neither revealed general higher or lower transcript levels in any of the parasite lines. **(B)** Expression patterns for six validated early gametocyte markers. **(C)** Expression patterns for 49 genes that were excluded from the lists of variant genes in the 7G8 and/or parentals comparisons (Supplemental Table S1). Transcriptional variation of these genes in these comparisons may be explained by different proportions of gametocytes among the parasite lines compared. These genes showed higher expression in 7G8 than in other parental lines, and lower expression in LD10 than in other 7G8-derived parasite lines, which is a signature for potentially early gametocyte-specific genes. Approximately half of these 49 genes (24 genes) were identified as early gametocyte-specific in a recent proteomics study (Silvestrini et al. 2010).

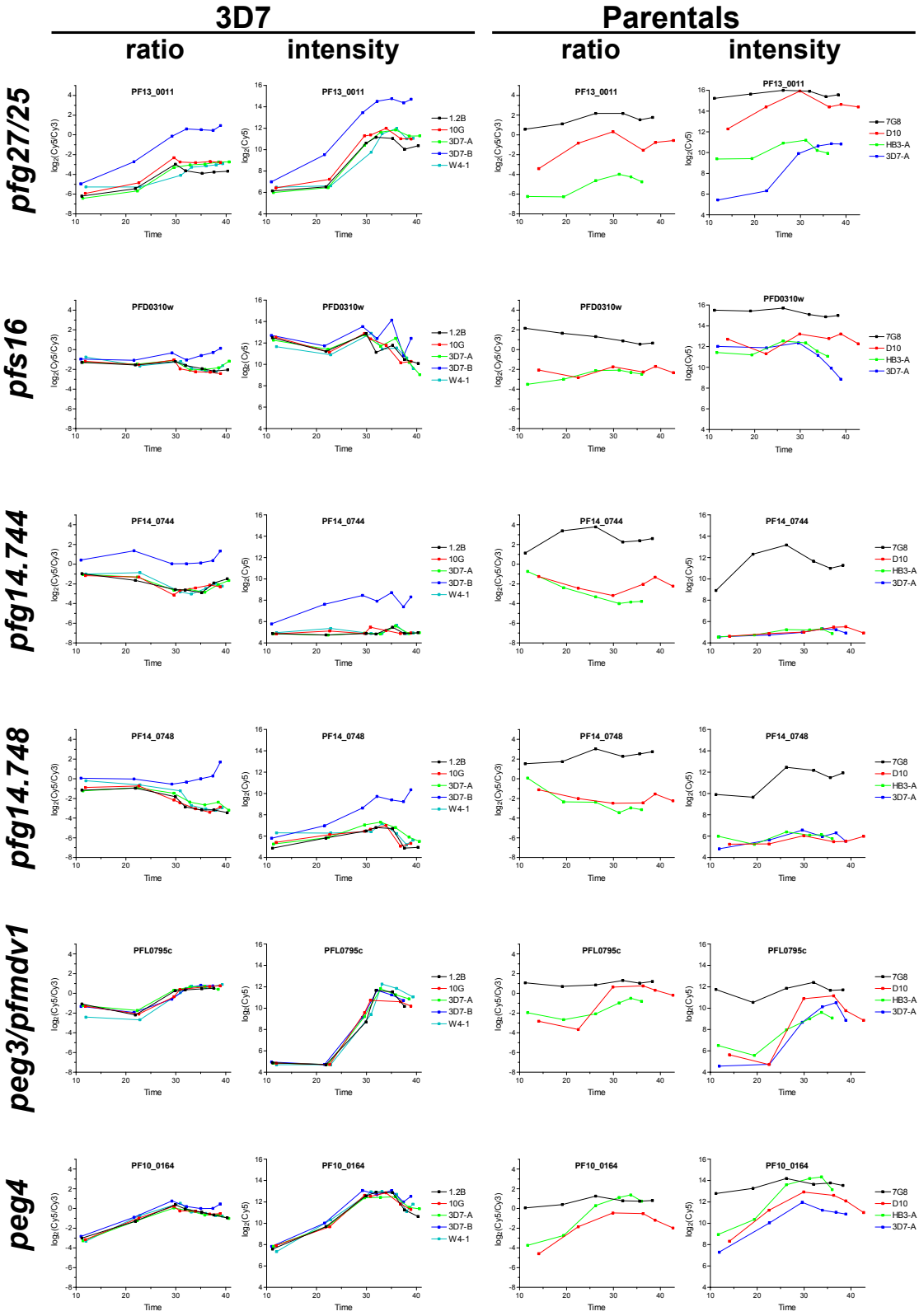
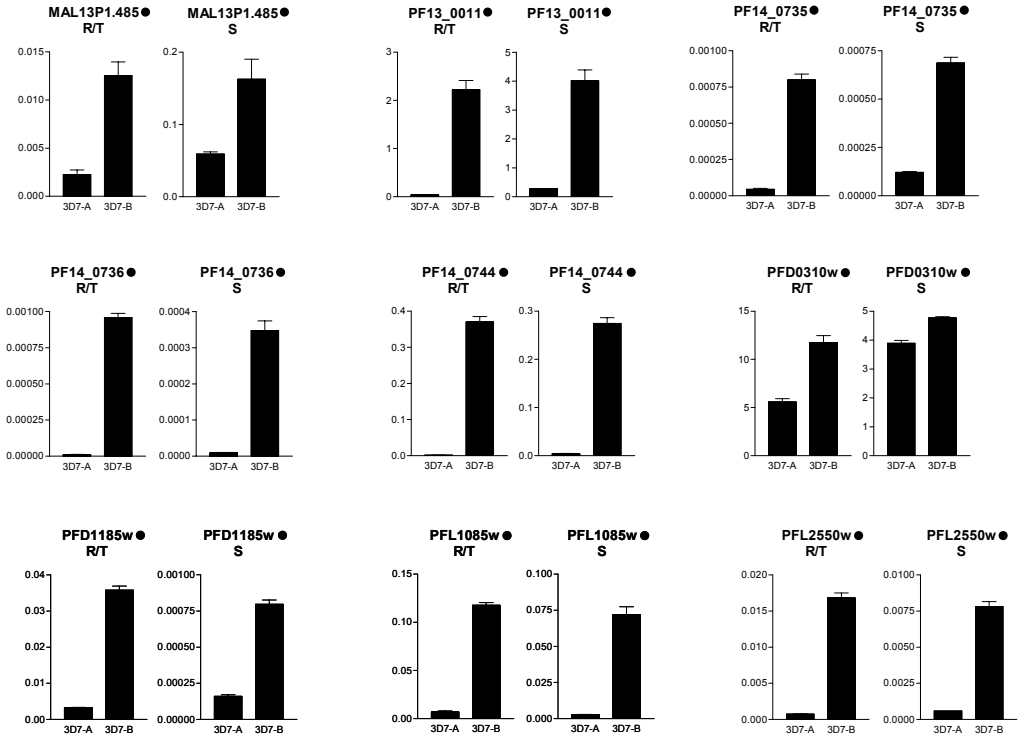
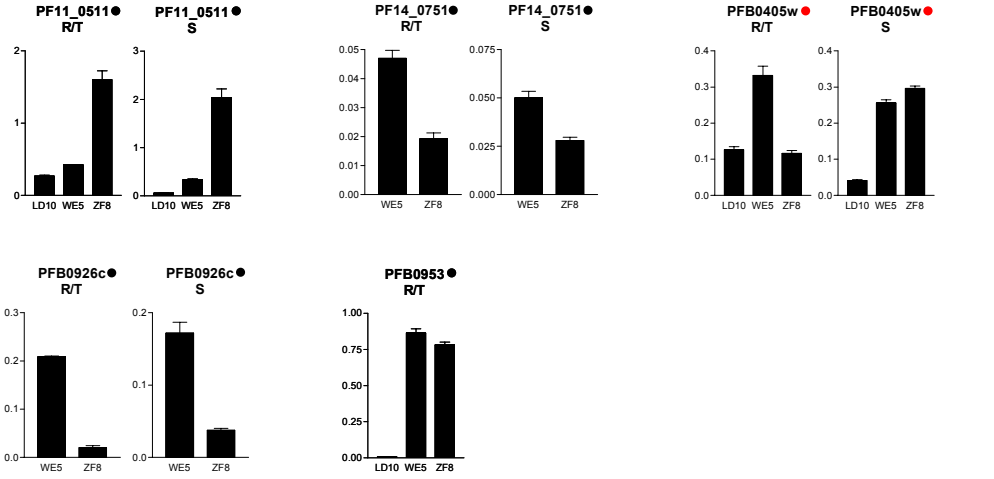


Figure S7. Time-course expression plots for six validated early gametocyte markers (Carter et al. 1989; Eksi et al. 2005; Furuya et al. 2005; Silvestrini et al. 2005; Olivieri et al. 2009). Expression levels relative to the reference pool [$\log_2$(Cy5/Cy3)] and Cy5 intensities ($\log_2$Cy5) are plotted against statistically-estimated parasite age. Results are shown for the 3D7 and parentals comparisons. In the parentals comparison, 3D7-A is not included in the "ratio" plots (expression relative to the reference pool), because 3D7-derived parasite lines were hybridized against a different reference pool than non-3D7 parasite lines (see Methods).

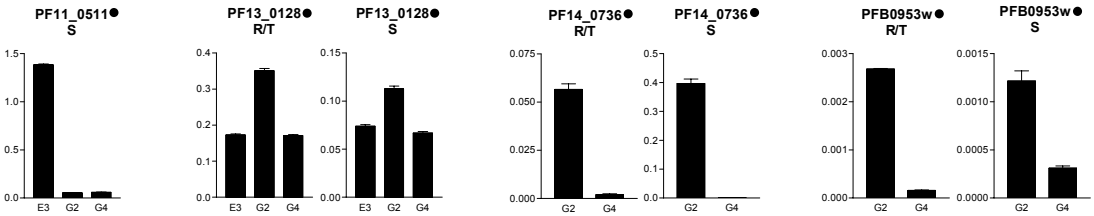
3D7


MAL13P1.485● R/T
MAL13P1.485● S
PF13_0011● R/T
PF13_0011● S
PF14_0735● R/T
PF14_0735● S
PF14_0736● R/T
PF14_0736● S
PF14_0744● R/T
PF14_0744● S
PFD0310w● R/T
PFD0310w● S
PFD1185w● R/T
PFD1185w● S
PFL1085w● R/T
PFL1085w● S
PFL2550w● R/T
PFL2550w● S
3D7-A
3D7-B

7G8


PF11_0511● R/T
PF11_0511● S
PF14_0751● R/T
PF14_0751● S
PFB0405w● R/T
PFB0405w● S
PFB0926c● R/T
PFB0926c● S
PFB0953● R/T
LD10
WE5
ZF8

D10


PF11_0511● S
PF13_0128● R/T
PF13_0128● S
PF14_0736● R/T
PF14_0736● S
PFB0953w● R/T
PFB0953w● S
E3
G2
G4

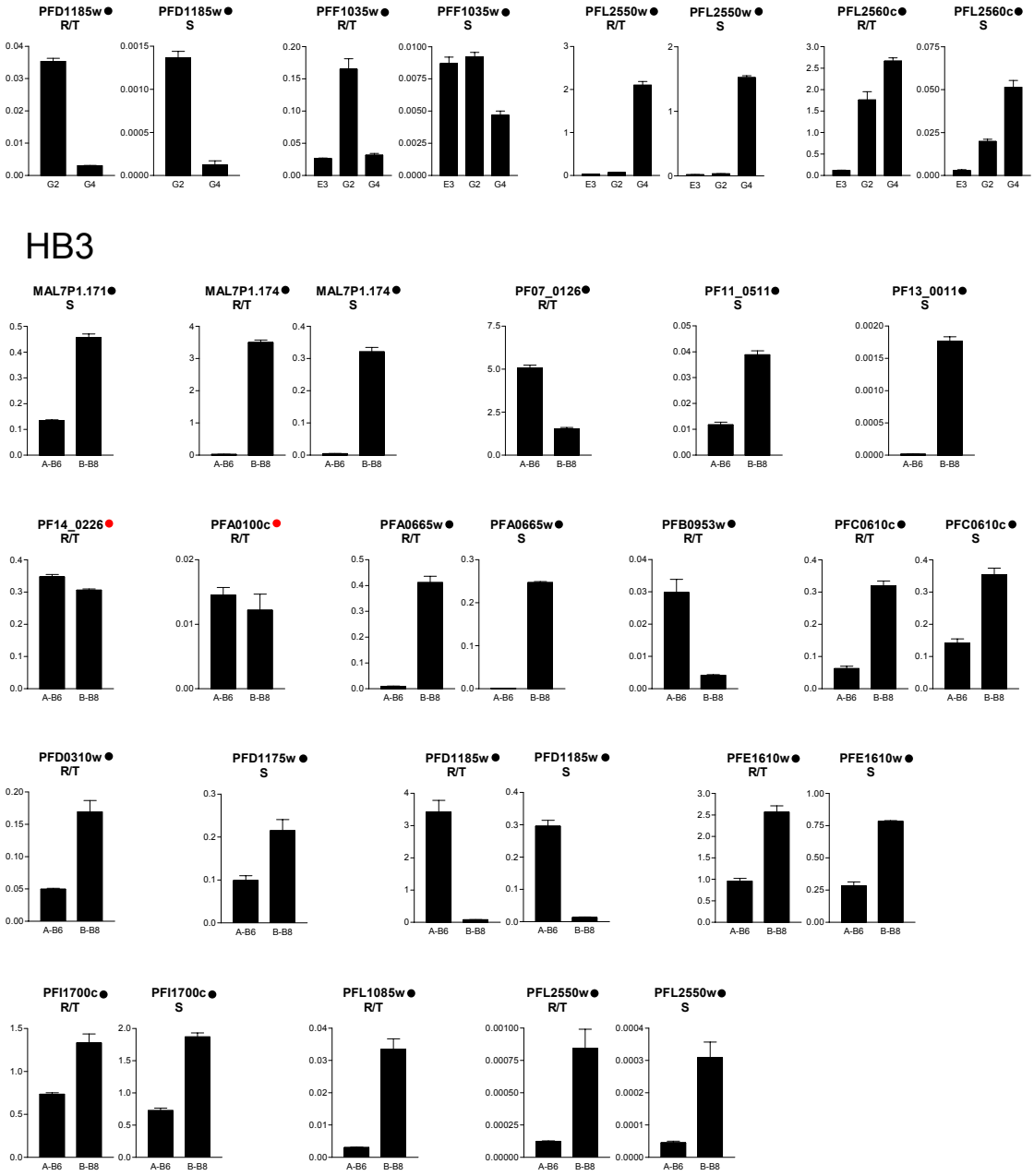


Figure S8. qPCR validation of variant expression in independent biological samples. Expression levels for selected genes were determined by qPCR in cDNA obtained from parasite preparations independent from those used for microarray analysis. Expression was analyzed at the stage(s) of the life cycle at which differences were observed in the microarray analysis. RNA preparation and reverse transcription were performed as previously described (Cortés et al. 2007). All qPCRs were performed using PowerSYBR Green Master Mix (Applied Biosystems), and expression values calculated using the relative standard curve method. The primers used are detailed in Supplemental Table S3. Results are expressed in arbitrary units, relative to a gDNA standard curve, and were normalized against *seryl tRNA synthetase* (PF07_0073). The samples used for qPCR analysis were from parasites at the ring (R), trophozoite (T), or schizont (S) stages, as indicated, but parasites were not as synchronized as in the cultures used for microarray analysis.

Genes that showed transcriptional patterns consistent with the patterns observed in the microarray analysis are indicated by a black dot here and in Fig. 2, whereas three genes

that showed different patterns are indicated by a red dot. Discrepancies may be explained by the lower synchrony of parasites in the preparations used for qPCR analysis, or may indicate switches in the expression of these genes. However, qPCR analysis performed on the same cDNA samples used for microarray analysis confirmed in all cases microarray results (not shown).

For this qPCR analysis, we selected genes distributed along the different clusters of variant genes (Fig. 2) that showed clear differences in expression between the parasite lines for which we prepared new biological samples (3D7-A, 3D7-B, LD10, WE5, ZF8, E3, G2, G4, AB6 and BB8). We tried to avoid genes with abrupt expression changes along the asexual cycle, to prevent artifacts related to differences in the stage of the life cycle, because stage effects are more difficult to control in qPCR experiments (statistical parasite age estimation is not possible) and samples were not synchronized to a defined age-window. We also included in the analysis genes for which we already had qPCR primers in the lab from previous studies.

Expression of several other variant genes had been previously assessed by qPCR or semi-quantitative PCR on independent biological preparations [PFC0110w (*clag3.2*), PFC0120w (*clag3.1*), PFB0935w (*clag2*), MAL13P1.60 (*eba-140*), MAL13P1.59 (*phista* family), and PF14_0749 (*acbp14*)] (Cortés et al. 2007; Crowley et al. 2011). In all cases, the results were fully consistent between the previous analysis and our microarray results. These genes are also marked with black dots in Fig. 2.

This analysis on independent biological preparations included 20 genes that were found to be variantly expressed in only one comparison. The microarray result was confirmed for 18 of these genes, indicating that the reliability of our data is reasonably high even in this group. This is also supported by the observation that even after excluding *var*, *rif*, *stevor* and *pfmc-2tm* from the analysis, 42% of the genes found to be variant in only one comparison were part of multigene families with at least one additional gene showing variant expression. Gene families with variant genes represent only 8.5% of the genome. For genes variant in more than one comparison, the percentage belonging to gene families with at least one more variant gene was 59%.

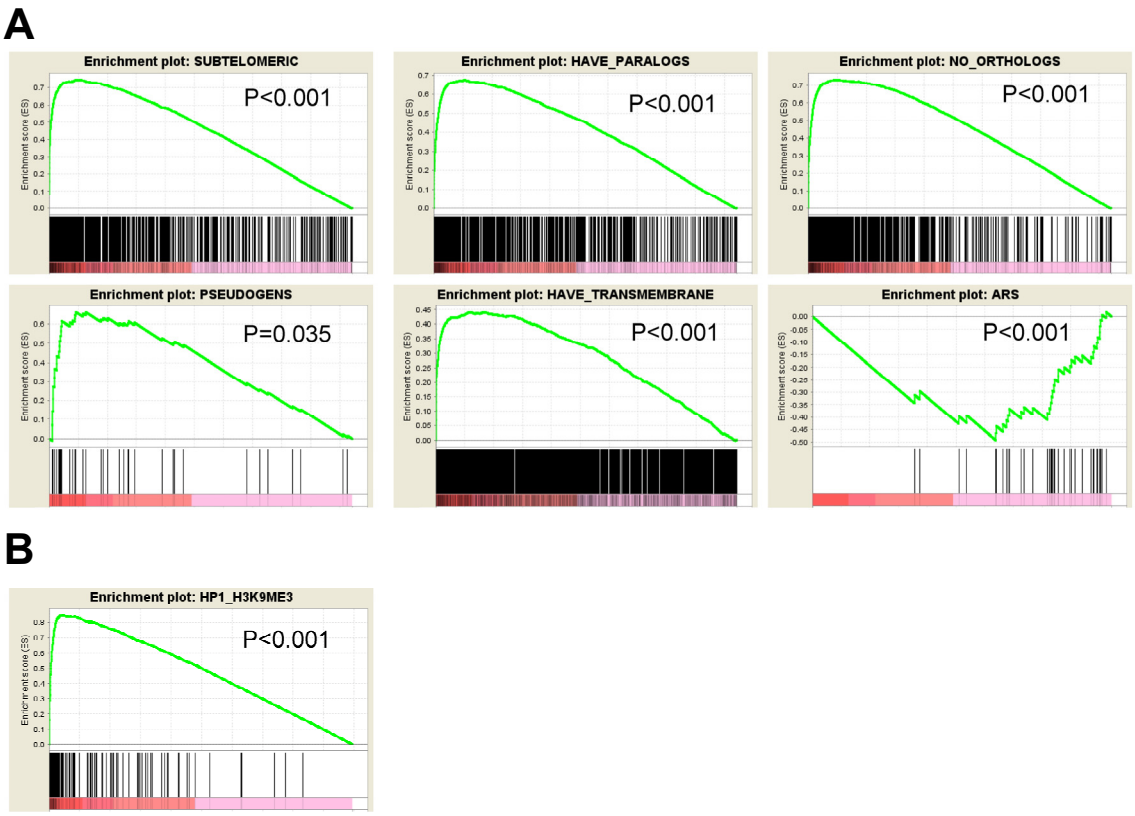


Figure S9. Gene set enrichment analysis (GSEA) of variantly-expressed genes. Genes were ranked in descending order (from left to right) according to their highest aMAFC among 3D7, 7G8 and HB3 comparisons. The parentals comparison was excluded because it would dominate aMAFC for the majority of genes (Supplemental Fig. S13B), and the D10 comparison was excluded because some genes showed an anomalous behavior (Supplemental Results), but analysis using the five sets of parasite lines yielded almost identical results to the analysis presented here (not shown). The distribution of several gene sets along the ranked genes is shown. **(A)** The gene sets correspond to subtelomeric genes (less than 150 Kb from the chromosome end), genes with paralogs, genes that lack identified orthologs in other species, pseudogenes, and genes encoding proteins with transmembrane domains. These five gene sets showed a significantly skewed distribution towards high levels of variant expression (P<0.05). For comparison, the sixth panel is a gene set representative of essential genes not showing variant expression (aminoacyl tRNA synthetases, ARS). This gene set showed a significantly skewed distribution towards low levels of variant expression. When the analysis was performed excluding the dominant variant gene families *var*, *rif*, *stevor* and *pfmc-2tm*, the five gene sets still showed a highly significant association with high levels of transcriptional variability. All gene sets were obtained from plasmoDB, and the analysis was performed using GSEA preranked (Subramanian et al. 2005). We also tested the distribution of other gene sets, including gene families (Supplemental Table S4), GO terms, Interpro domains, metabolic pathways (Ginsburg 2006), and OPI clusters (Zhou et al. 2008). This analysis confirmed the conclusions obtained from the analysis based on assigning variant genes to known gene families (main text), but did not reveal any additional information. **(B)** Distribution of genes that are HP1 or H3K9me3 positive, according to published chip on ChIP data (Flueck et al. 2009; Lopez-Rubio et al. 2009; Salcedo-Amaya et al. 2009). Genes from the families *var*, *rif*, *stevor* and *pfmc-2tm* were excluded from the analysis. Even after excluding these large gene families, there was a significant association between clonally variant gene expression and heterochromatin marks (P<0.001).

	3D7.	D10.	HB3.	7G8.
var	71	0	0	0
rif	86	0	0	0
stevor	90	0	0	0
pfmc-2tm	69	0	0	0
phista	65	42	48	58
phistb	92	80	88	92
phistc	100	94	100	100
Phistb_dnaj	100	100	100	100
ab_hyd	100	100	100	100
dnaj_I	100	100	100	100
dnaj_III	89	89	89	89
Other with dnaj	94	94	94	94
emp3	100	100	100	100
gbp130/gbph	100	67	100	100
fikk	88	88	79	88
hyp1	100	100	100	100
hyp2	100	100	100	100
hyp4	11	11	11	11
hyp5	11	11	11	11
hyp6	100	50	100	50
hyp7	100	100	100	100
hyp8	100	50	100	100
hyp9	100	100	100	100
hyp10	100	100	100	100
hyp11	100	80	100	100
hyp12	100	67	100	100
hyp13	100	100	100	100
hyp15	100	75	75	75
hyp16	100	100	100	100
hyp17	100	100	100	100
clag	100	60	60	80
acs	92	92	85	92
acbp	100	100	100	100
msrps	100	100	100	100
msp3/6	100	100	100	100
ars	100	100	100	100
surfin	100	90	100	100
etramps	100	100	100	100
apiAP2-tf	100	100	100	100
eba	100	75	100	100
PfRh	86	71	86	86
hrp	50	50	50	50
crmp	100	100	100	100
PFD0075-like	67	67	67	67
Lysophospholipase	100	87	100	100
articulin/alveolin.	100	100	100	100
cpw-wpc	100	100	100	100
macpf	100	100	100	100
6-cys	100	100	100	100
ron	100	100	100	100
sera	100	100	100	100
falcipain	100	100	75	75
plasmepsins	100	100	100	100
subtilisins	100	100	100	100
rhomboid	80	80	80	80
TRAP-like (or TSR domain).	83	83	83	83
tRNA nuclear	73	73	73	73
tRNAs apic.	74	74	74	74
rRNAs.	15	15	15	15
snRNAs	0	0	0	0
hsp100	100	100	100	100
hsp90	100	100	100	100
hsp70	100	100	100	100
hsp60	100	100	100	100
CCT/TCP chaperonin	100	100	100	100
prefoldin	100	100	100	100
small HSP	100	100	100	100
Hsp90 co-chaperones	100	100	100	100
Cyclophilins	90	90	90	90
ABC superfamily	100	100	100	100
Histones	100	100	100	100

Figure S10. Probe coverage for different gene families. Values are the percentage of genes of each family that were included in each comparison, after excluding probes on the basis of blast and CGH values (see Supplemental Methods). Orange shading indicates gene families with less than 50% of genes retained, whereas yellow indicates gene families with 50-75% of genes retained. Gene families are detailed in Supplemental Table S4, which also shows the specific genes that were included or excluded in each comparison. Supplemental Table S5 shows the probes that were retained for the analysis of each genetic background.

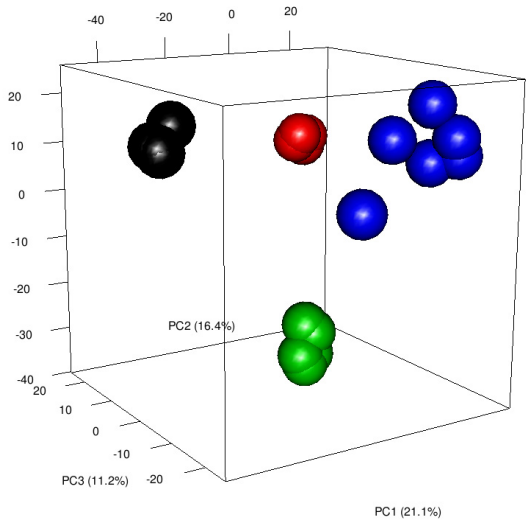
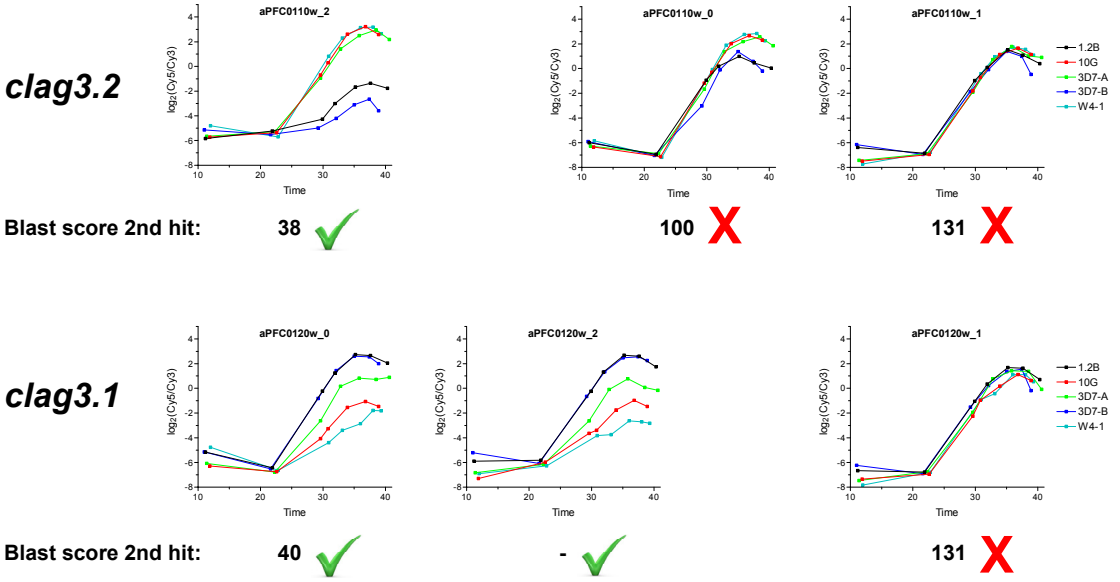


Figure S11. Three-dimensional principal components analysis (PCA) of CGH values for 20 parasite lines used in this study (genomic DNA from each parasite line was hybridized against genomic DNA from 3D7-A). Each sphere corresponds to a different parasite line, color indicates expected genetic background (black 3D7, red 7G8, green D10, blue HB3). Parasite lines of the same genetic background clustered together.

A



B

	7G8 probes		D10 probes		HB3 probes	
	All	Variants only 7G8	All	Variants only D10	All	Variants only HB3
Average blast score	94.6	91.3	92.1	88.77	136.3	132.5
% blast score >100	58.2	54.8	56.1	54.55	97.8	93.4
Average CGH low blast genes	0.11	0.17	0.12	0.31	0.11	0.11

Figure S12. Validation of the criteria used for exclusion of oligonucleotides from the analysis (see Supplemental Methods). **(A)** As a stringent test for these criteria, we compared the expression patterns obtained with different probes for the genes PFC0110w (*clag3.2*) and PFC0120w (*clag3.1*), which are 95% identical between them but highly polymorphic among genetically different parasite lines. This panel shows expression patterns in 3D7-derived parasite lines. A green tick indicates that the probe was retained, whereas a red cross indicates that the probe was excluded. Three probes that were excluded because they are potentially cross-reactive (second hit blast score > 70) show a different pattern from non cross-reactive probes, such that they are unable to capture the variant expression of these genes, demonstrating that they had been correctly excluded by our criteria. Similar analysis of the probes excluded in non-3D7 parasite lines revealed that the exclusion criteria resulted in correct inclusion or exclusion for the majority of probes (not shown). **(B)** To confirm that genetic differences in non-3D7 parasite lines did not result in artifacts in our analysis using a 3D7-based microarray, we specifically analyzed the sequence of genes found to be variantly expressed in only one non-3D7 comparison (the set of variant genes that could be suspected to be more affected by sequence variation-related artifacts). The table shows the average blast score and the percent of probes with a blast score >100, for all probes included in the analysis of a particular genetic background and for probes targeting genes variantly expressed only in that particular comparison. Both in the case of the nearly complete HB3 genome and the partial D10 and 7G8 genomes, values were similar between the two sets of probes. We next analyzed CGH values for probes of

genes showing variant expression in only one comparison and low blast score (row "Average CGH low blast genes"). They had CGH values [$\log_2$(Cy5/Cy3)] near 0 in the hybridization against 3D7 gDNA, indicating that the low blast score was not attributable to impaired hybridization due to polymorphism. Analysis of individual probes confirmed that the low blast score of these probes was predominantly explained by absence of sequence data at these positions, as they mapped to large sequence gaps or long strings of "N" in the incomplete genomes, according to analysis of neighboring sequences (not shown).

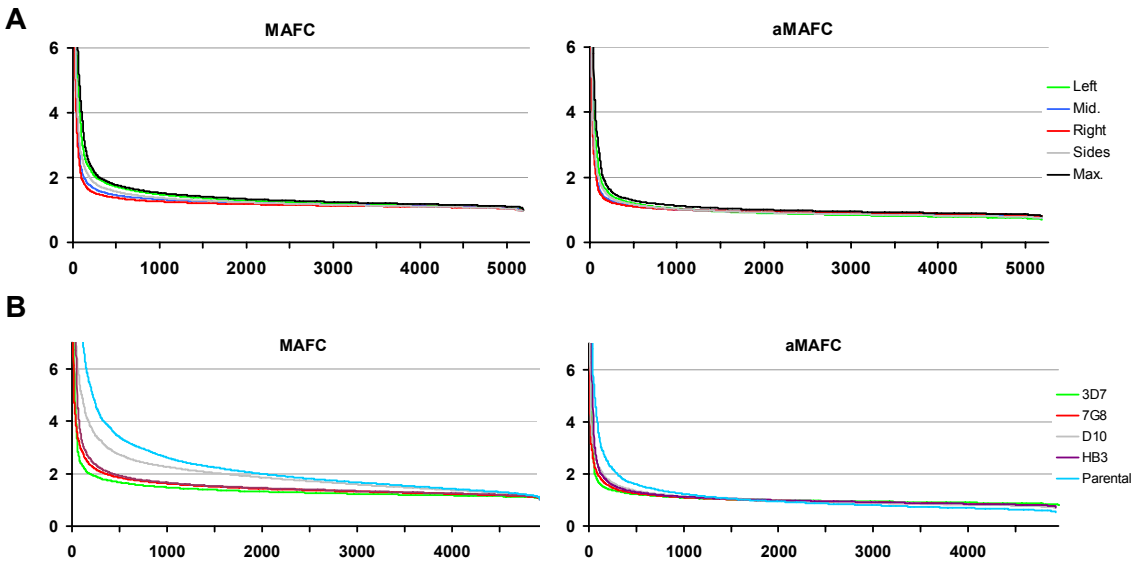


Figure S13. Comparison of the distribution of the maximum average fold change across half the time compared (MAFC) and the adjusted maximum average fold change across half the time compared (aMAFC) (see Methods). Genes (x axis) were ranked by their MAFC or aMAFC in descending order. **(A)** Distribution of MAFC and aMAFC for the different half-TC (time compared) intervals in 3D7 parasite lines. Left is 0 to 0.5x TC, right is 0.5x to1x TC, Mid is 0.25x to 0.75x TC, and Sides is 0 to 0.25x TC + 0.75x to 1x TC. **(B)** Distribution of MAFC and aMAFC for the five comparisons of parasite lines.

SUPPLEMENTAL REFERENCES

Blair PL, Witney A, Haynes JD, Moch JK, Carucci DJ, Adams JH. 2002. Transcripts of developmentally regulated Plasmodium falciparum genes quantified by real-time RT-PCR. *Nucleic Acids Res* **30:** 2224-2231.

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium falciparum*. *PLoS Biol* **1:** E5.

Cabral FJ , Wunderlich G. 2009. Transcriptional memory and switching in the *Plasmodium falciparum rif* gene family. *Mol Biochem Parasitol* **168:** 186-190.

Carter R, Graves PM, Creasey A, Byrne K, Read D, Alano P, Fenton B. 1989. *Plasmodium falciparum*: an abundant stage-specific protein expressed during early gametocyte development. *Exp Parasitol* **69:** 140-149.

Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. 2007. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog* **3:** e107.

Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. 2011. 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.

Dzikowski R, Frank M, Deitsch K. 2006. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathog* **2:** e22.

Eksi S, Haile Y, Furuya T, Ma L, Su X, Williamson KC. 2005. Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol* **143:** 90-99.

Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z et al. 2009. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog* **5:** e1000569.

Furuya T, Mu J, Hayton K, Liu A, Duan J, Nkrumah L, Joy DA, Fidock DA, Fujioka H, Vaidya AB et al. 2005. Disruption of a *Plasmodium falciparum* gene linked to male sexual development causes early arrest in gametocytogenesis. *Proc Natl Acad Sci U S A* **102:** 16813-16818.

Ginsburg H. 2006. Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends Parasitol* **22:** 238-240.

Hu G, Llinas M, Li J, Preiser PR, Bozdech Z. 2007. Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy. *BMC Bioinformatics* **8:** 350.

Joannin N, Abhiman S, Sonnhammer EL, Wahlgren M. 2008. Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family. *BMC Genomics* **9:** 19.

Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, Pinches R, Day F, Kyes SA, Conway DJ, Holmes CC et al. 2009. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. *Proc Natl Acad Sci USA* **106:** 7559-7564.

Lopez-Rubio JJ, Mancio-Silva L, Scherf A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5:** 179-190.

Olivieri A, Camarda G, Bertuccini L, van de Vegte-Bolmer M, Luty AJ, Sauerwein R, Alano P. 2009. The *Plasmodium falciparum* protein *Pfg27* is dispensable for gametocyte and gamete production, but contributes to cell integrity during gametocytogenesis. *Mol Microbiol* **73:** 180-193.

Petter M, Haeggstrom M, Khattab A, Fernandez V, Klinkert MQ, Wahlgren M. 2007. Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol* **156:** 51-61.

Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, van de Vegte-Bolmer M, Selzer RR, Iniguez AL et al. 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **106:** 9655-9660.

Silvestrini F, Bozdech Z, Lanfrancotti A, Di Giulio E, Bultrini E, Picci L, Derisi JL, Pizzi E, Alano P. 2005. Genome-wide identification of genes upregulated at the onset of gametocytogenesis in *Plasmodium falciparum*. *Mol Biochem Parasitol* **143:** 100-110.

Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, Sanchez M, Younis Younis S, Sauerwein R, Alano P. 2010. Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* **9:** 1437-1448.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102:** 15545-15550.

Urbán P, Estelrich J, Cortés A, Fernàndez-Busquets X. 2011. A nanovector with complete discrimination for targeted delivery to *Plasmodium falciparum*-infected versus non-infected red blood cells in vitro. *J Control Release*.

Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SF, Milner DA, Jr., Daily JP, Sarr O, Ndiaye D, Ndir O et al. 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet* **39:** 113-119.

Wang CW, Magistrado PA, Nielsen MA, Theander TG, Lavstsen T. 2009. Preferential transcription of conserved *rif* genes in two phenotypically distinct *Plasmodium falciparum* parasite lines. *Int J Parasitol* **39:** 655-664.

Young JA, Fivelman QL, Blair PL, de la Vega P, Le Roch KG, Zhou Y, Carucci DJ, Baker DA, Winzeler EA. 2005. The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol Biochem Parasitol* **143:** 67-79.

Zhou Y, Ramachandran V, Kumar KA, Westenberger S, Refour P, Zhou B, Li F, Young JA, Chen K, Plouffe D et al. 2008. Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. *PLoS One* **3:** e1570.