

Comparative Transcriptional and Genomic Analysis of *Plasmodium falciparum* Field Isolates

Margaret J. Mackinnon^{1*}, Jinguang Li², Sachel Mok², Moses M. Kortok¹, Kevin Marsh¹, Peter R. Preiser², Zbynek Bozdech²

1 KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya, **2** Department of Biological Sciences, Nanyang Technological University, Nanyang, Singapore

Abstract

Mechanisms for differential regulation of gene expression may underlie much of the phenotypic variation and adaptability of malaria parasites. Here we describe transcriptional variation among culture-adapted field isolates of *Plasmodium falciparum*, the species responsible for most malarial disease. It was found that genes coding for parasite protein export into the red cell cytosol and onto its surface, and genes coding for sexual stage proteins involved in parasite transmission are up-regulated in field isolates compared with long-term laboratory isolates. Much of this variability was associated with the loss of small or large chromosomal segments, or other forms of gene copy number variation that are prevalent in the *P. falciparum* genome (copy number variants, CNVs). Expression levels of genes inside these segments were correlated to that of genes outside and adjacent to the segment boundaries, and this association declined with distance from the CNV boundary. This observation could not be explained by copy number variation in these adjacent genes. This suggests a local-acting regulatory role for CNVs in transcription of neighboring genes and helps explain the chromosomal clustering that we observed here. Transcriptional co-regulation of physical clusters of adaptive genes may provide a way for the parasite to readily adapt to its highly heterogeneous and strongly selective environment.

Citation: Mackinnon MJ, Li J, Mok S, Kortok MM, Marsh K, et al. (2009) Comparative Transcriptional and Genomic Analysis of *Plasmodium falciparum* Field Isolates. *PLoS Pathog* 5(10): e1000644. doi:10.1371/journal.ppat.1000644

Editor: Kirk Deitsch, Weill Medical College of Cornell University, United States of America

Received: December 19, 2008; **Accepted:** October 5, 2009; **Published:** October 30, 2009

Copyright: © 2009 Mackinnon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The costs of this work were borne by the Royal Society of London, the Wellcome Trust (grant number 077092), the University of Cambridge, the University of Oxford, the Singapore Biomedical Research Council (grant number 05/1/22/19/398) and the Singapore National Medical Research Council (grant number NMRC/1168/2008). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Mmackinnon@kilifi.kemri-wellcome.org

Introduction

Malaria is a global health problem that imposes major strain on the development of many tropical countries [1]. The challenges in combating malaria are largely due to the complexity of the parasite responsible for most of the disease, *Plasmodium falciparum*. In particular, the parasite's inherent ability to adapt to its environment, coupled with its extraordinary biological diversity can erode the efficacy of control interventions: the widespread development of drug resistance is a salient example of this [2]. Other parasite traits that are under natural selection, such as virulence (parasite-mediated damage to its host) and transmissibility (the parasite's success in spreading to new hosts), may evolve in response to control programs because they are key determinants of parasite fitness [3,4]. Understanding the molecular mechanisms of these traits, their genetic regulation and their evolution in natural populations is fundamental to designing successful, sustained control programs and new therapies.

Whole transcriptome analyses using microarrays has opened up the possibility of simultaneously exploring all the parasite's genes for their combined and individual function, and their regulation. In *Plasmodium*, through a series of approaches such as gene knockouts [5,6], comparative genomics and bioinformatics [7–12], and transcriptome-phenotype or -genotype association studies using microarrays [13–17], a picture is building up of the types of genes that are essential to the parasite's function and survival *in vivo*

and hence important in disease processes. A further theme that has emerged from these studies is the remarkable degree of coordinated regulation in the transcription of *P. falciparum* genes, both within and between life stages [18–22]. This is evident from the strong co-regulation through time within the asexual cycle of genes with related function [19,20,22,23], but has also been observed for genes that share similar location in the genome [9,20,22,24,25]. The mechanisms for coordinated control of transcription in *Plasmodium* are not well understood: the current focus is on chromatin modification, locus repositioning within nuclear sub-compartments, identification of DNA-binding regulatory proteins, and the role of RNA decay in regulating transcript abundance (reviewed in [26,27]). The contribution of transcriptional regulation to whole-organism phenotypes such as virulence and transmissibility in natural parasite populations has not yet been investigated.

Several technical hurdles need to be overcome before large-scale analyses of natural parasite populations can be done. The first is to define the amount of biological and technical variability in microarray data from field parasites so that adequately powered experiments can be designed. The second is to obtain samples throughout the 48-hour cycle since most genes in *P. falciparum* have stage-specific expression [19,20] and patient samples only yield parasites at an early stage in their cycle. This can be overcome by maturing the parasites *in vitro* and then taking samples for RNA in time series as they mature through the 48 hours. This, however,

Author Summary

Malaria parasites are particularly good at adapting to the variable environments they encounter during their life cycle. This property helps explain their widespread prevalence and persistence despite massive malaria control campaigns. The genes responsible for this adaptability are largely unknown. In this study we analyzed gene expression profiles of *P. falciparum* parasites recently taken from the field and compared them to those from laboratory-adapted parasites. Many of the genes that were up-regulated in field isolates coded for proteins which are exported, or involved in export, from the parasite onto the surface of the host cell where they interact with the immune system. Differences at the expression level were partly explainable by variation in gene copy number caused by deletions or amplification of small chromosomal segments. Usually, higher gene copy was associated with elevated expression levels, but for some genes, more copies appeared to repress expression. Remarkably, the regulatory effects of deleted or amplified segments appeared to extend to genes outside the segment boundaries. This suggests an epigenetically mediated co-regulation of tightly linked gene clusters. Such 'soft-wired', coordinated adaptation is potentially an important mechanism by which the parasite readily adapts to its current host. The molecules responsible may therefore be important targets for drugs or vaccines.

yields a third challenge in that volumes of blood from patients are usually too small to provide enough RNA for multiple time point analyses. Finally, there is some uncertainty about the adequacy of microarrays to capture the relevant variation among field parasites [28] because most arrays are designed on the basis of a single reference genome (from clone 3D7) that has been maintained in the laboratory for a long time [29].

Therefore, with a view to scaling up the application of microarray technology to studies of natural populations of *P. falciparum*, we performed whole-transcriptome analyses on a handful of field and laboratory parasites starting with the small amounts of RNA material typically obtained from a patient sample. We sampled parasite cultures at regular intervals throughout the 48-hour cycle and incorporated time point effects into the analysis. By performing technical and biological replicates we were able to quantify the level of transcriptional biological variation among field isolates relative to background technical noise, thereby testing for their true biological differences from laboratory isolates. We found that much of this variation could be explained by gene copy number variation at the sequence level due to deletion and amplification events. We further found evidence of co-regulation of genes outside as well as inside these copy number variant (CNV) regions suggesting that they act as important determinants of transcriptional regulation.

Results

A panel of 5 recently culture-adapted field isolates of *P. falciparum*, and 3 laboratory-adapted isolates were subjected to whole-genome transcriptional analysis using a 70mer DNA microarray that contained 10,417 probes representing 5577 genes [30]. Material was taken from synchronized cultures every 8 hours for 48 hours, amplified by PCR using the SMART technique (see below), and assayed on microarrays in triplicate. The focus of our analysis was on differential expression between strains rather than differential expression through time as this has been well addressed

in previous analyses of the transcriptome [19,20,31,32]. To determine whether strains differed in stage of maturity at the time of harvesting RNA we used the data from the 48-hour transcriptome [19] to estimate the 'age' of each strain at the first sampling time point [15]: this ranged between -3hr and +4hr with a mean of 0.7hr (Figure S1). This variability in time of harvesting, despite efforts to standardize stage of maturity in our cultures, highlights the need in transcriptional studies of *Plasmodium* to take samples from multiple time points to avoid confounding of true biological differences with those due to stage of maturity [13,14,33–36].

Evaluation of amplification method

To overcome the problem of limited amounts of parasite RNA obtainable from patients, we adapted and validated an RNA amplification technique [37] for use in *P. falciparum* that yields enough mRNA to perform multiple time point transcriptome analysis from a typical patient (<3ml of whole blood). Using strain 3D7, we tested the correspondence between expression profiles obtained using amplified vs. unamplified cDNA (Figures S2A, B). There was good correspondence between time profiles of individual genes measured using the two methods (Figure S2C) and in whole-genome expression profiles within each time-point (Figure S2B). Amplification of cDNA by SMART-PCR gave higher sensitivity to detect genes with low expression values (Figure S2), as found previously [37]. It also led to underestimation of expression in highly expressed genes: the regression slope of genes with above average expression values was $b = 0.78$ which is significantly ($P < 0.001$) below the value of 1 expected if the fidelity of amplification was 100%. Thus amplification improved sensitivity but introduced some downward quantitative bias into absolute measures of expression. The magnitude of this bias was such that genes that were 8-fold lower than the average gene (i.e., the common reference material) before amplification were 4.5-fold after amplification while those that were 8-fold higher than average were 5-fold higher after amplification. However, amplification made little difference to the ranking of expression values: the Spearman rank correlation coefficient across all probes and time points in amplified vs. unamplified material was $r = 0.73$ which is close to that expected from within-method technical variability alone ($r = 0.83$ for replicate hybridizations within each method). We therefore proceeded to use this technique for analysis of RNA material from culture-adapted field isolates.

Differences between field and laboratory strains

Of the 5224 genes represented on the array after excluding *vars*, *rifins*, *stevors*, tRNAs and rRNAs we obtained full data (all time points for all strains in at least one hybridization and one probe) on 4896 (94%) of these (Dataset S1). Across all probes and arrays, 9% of the data were excluded due to low levels of signal or poor spot quality.

Hierarchical clustering of the whole genome expression profiles showed clear differentiation between laboratory and field strains (Figure 1A). Principal coordinates analysis [38] confirmed these groupings (Figure 1B). After allowing for multiple testing, applying a cut-off of > 1.5 -fold difference based on a volcano plot (Figure S4), and taking only those genes for which the majority of their probes at the $P < 0.01$ level, the proportion of genes that differed between laboratory and field isolates was 5.3% (259 genes, Figure 1C). For genes with multiple probes and at least one probe significant, 55% of the remaining probes were also significant: when compared with that expected by chance (6.8% of all probes were significant), this indicates high consistency in expression levels across probes representing the same gene. Of those genes

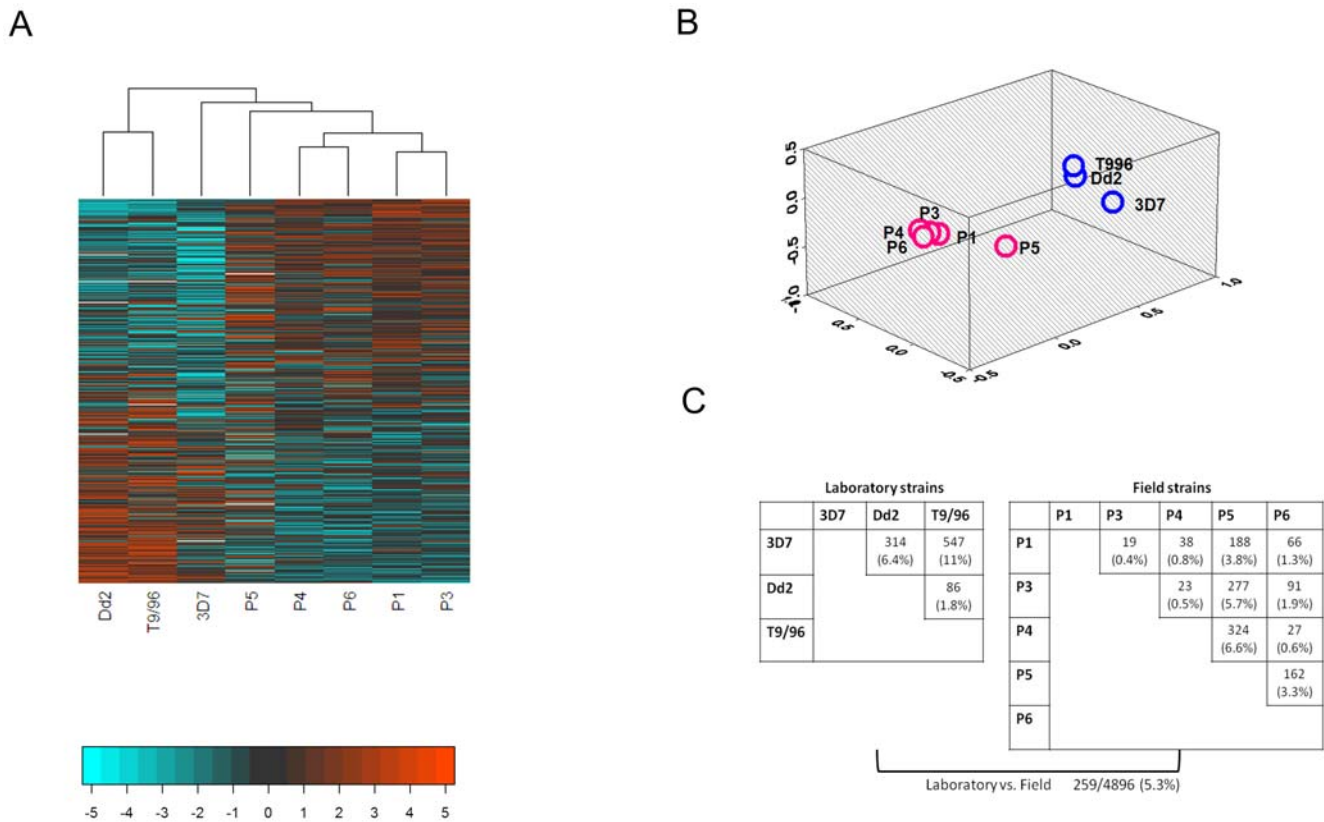


Figure 1. Overview of Pairwise Comparison of Transcriptional Differences between *P. falciparum* Laboratory and Field Strains. **A.** Heat map representing each strain's (column) expression values for all genes (rows). Rows are ordered from top to bottom according to the maximum difference between field and laboratory isolates. Values represented by colors (red, positive; blue, negative) are summary measures over time points (averages across the life cycle, as output from the statistical model) of the strain's log₂ expression value relative to the reference material. Values are mean-normalized across rows to highlight between-strain differences. Hierarchical clustering of the column data yielded a dendrogram that broadly separated the laboratory and field strains into two clusters with 3D7 falling somewhere between these groups. **B.** When represented by a 3-dimensional principal coordinates plot in which distances between points indicate the degree in similarity between strains across their whole genome expression profile, laboratory and field isolates fall into two distinct clusters. **C.** Tests of significance revealed that 5.3% of genes were significantly differentially expressed between laboratory and field strains (>1.5-fold difference, P<0.01 after allowing for multiple testing and including genes for which the majority of probes per gene were significant). 3D7 and P5 were more different from their counterparts than other strains. doi:10.1371/journal.ppat.1000644.g001

declared significant, 78% were up-regulated and 22% were down-regulated in field strains compared with laboratory strains. Among field isolates alone, 592 genes (7.6%) showed significant ratios of between-strain to within-strain variability for at least one probe at the P<0.001 level: 83 (32%) of the genes that differed significantly between field and laboratory isolates also differed among field isolates.

Figure 2A displays the time profiles of the 20 most up-regulated and 20 most down-regulated genes when comparing field vs. laboratory strains. Strain differences were highly consistent across time points and for different probes representing the same gene (Figure 2A). Notable members of the top 20 in the up-regulated group were three sexual stage-related genes (gamete antigen 27/25 (PF13_0011), a gametocyte-implicated protein (PFI1720w), and a male gamete development gene (PFL0795c)); one gene found to be suppressed in pregnancy-associated malaria (PF10_0350) [16]; a gene coding for proteins similar to the merozoite surface-expressed proteins (PF13_0196, MSP7-like protein) which sits adjacent to the MSP7 precursor that was significantly down-regulated in field isolates; a gene coding for transcription factor with AP2 DNA-binding domains (PFL1085w) which have been implicated in driving the highly coordinated regulation of transcription during the 48-hour cycle [39,40]; three (PFI1735c, PFI1740c, PFI1755c) of

four homologous genes (named *rex1*, *rex2*, *rex3*, *rex4*) found in a cluster on the end of chromosome 9 that code for ring-stage proteins that are exported into the host cell [41] and found in the vesicle-like structures called Maurer's Clefts [42] which are implicated in the transport of parasite proteins to the red cell surface [43–47]; three contiguous genes on the end of chromosome 7 (PF07_0005, PF07_0006, PF07_0007), one of these coding for the sporozoite surface antigen, STARP, also expressed in ring stages [48], and another coding for a lysophospholipase; and four genes that code for PHIST proteins that are members of the exportome (PFI14_0748, PFI1770w, PFI1780w, MAL7P1.7). The remaining four of the top 20 were genes of unknown function on the end of chromosome 14 (PF14_0644, PF14_0708 and PF14_0735 whose neighbor (PF14_0736) was also strongly up-regulated but not part of the top 20) and on chromosome 6 (PFE0940c).

Notable members of the 20 most down-regulated genes in field isolates were five genes found previously to be down-regulated in pregnancy-associated malaria, three of which code for PHIST proteins and are members of the exportome (MAL7P1.225, PFB0105c and PFD1140w), one which is a protease (PFI14_0616) and another of unknown function (PFB0115w) [15,16,49]; another exportome gene from the *hyp5* family (PFI0085c); another gene coding for a protein kinase (PFL1370w); two homologous genes

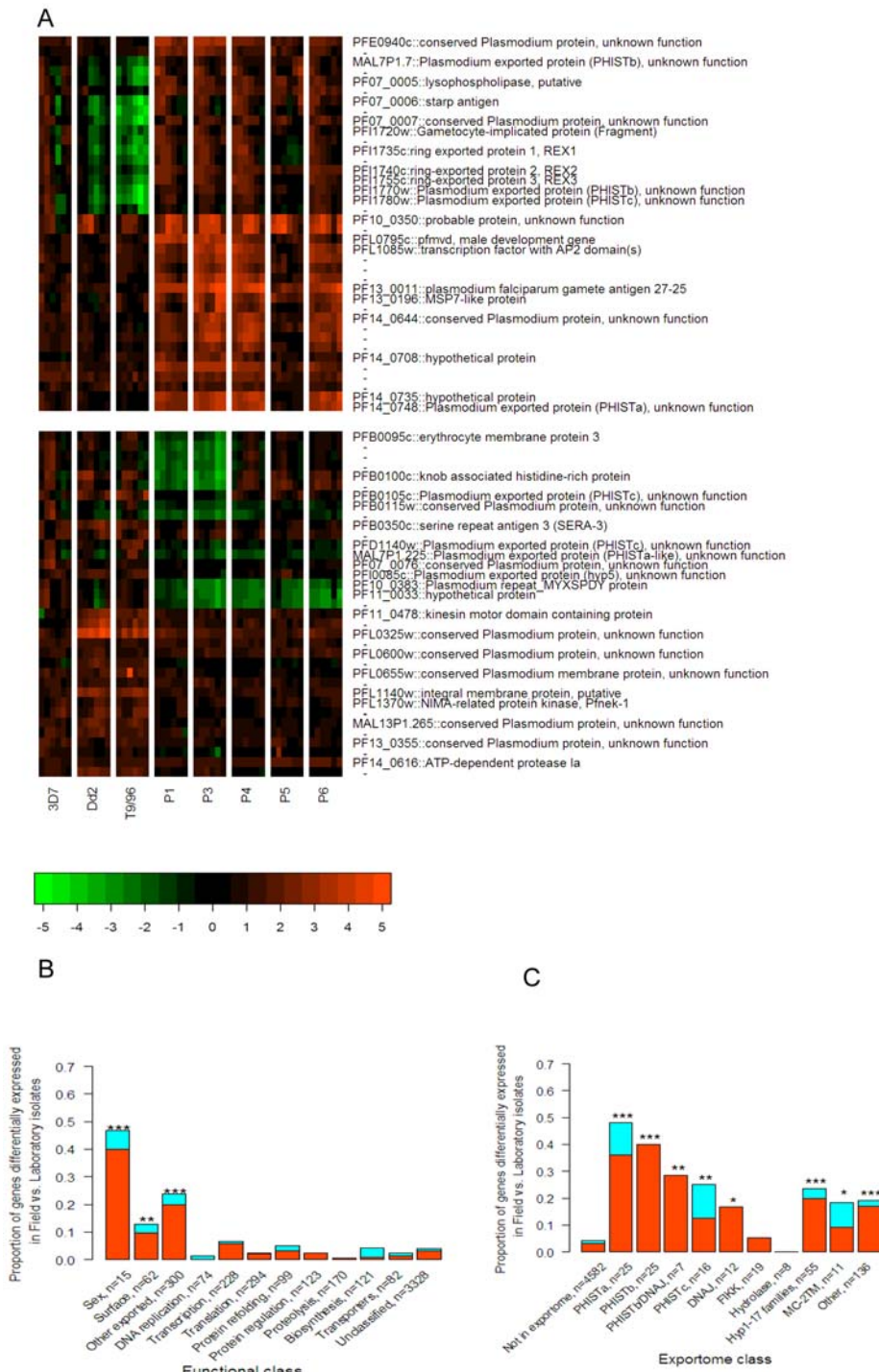


Figure 2. Genes that Differed Most Between Field and Laboratory Isolates and Their Functional Classification. **A.** Individual time point expression profiles for the 20 up-regulated (top panel) and 20 down-regulated genes (bottom panel) with lowest P-values in field vs. laboratory isolates comparisons. Rows are sorted by chromosomal location. All probes per gene, whether significant or not, are shown to illustrate the good correspondence across probes within gene. The data shown are actual expression values relative to the 3D7 reference material on the log₂ scale (see color bar), i.e., they have not been row-mean normalized. **B.** Genes were classified according to function based on their PlasmoDB annotations and GO terms, and bioinformatics studies (Table S1). The proportion of each class that was significantly expressed (>1.5-fold difference and P<0.01) between field and laboratory isolates was significantly higher in genes coding for sexual stage parasites, surface proteins and genes containing an export motif (exportome) than for other genes. Differentially regulated genes tended to be more often up-regulated (red) than down-regulated (light blue). **C.** Genes were classified according to their class within the exportome [57]. With the exceptions of the Maurer's Cleft genes (MC-2TM) and the FIKK kinases, all of the sub-classes had significantly higher numbers of up-regulated in field vs. laboratory isolates than non-exportome genes (left bar). Numbers per class are shown in the axis labels. Significance levels for differences in proportions of each class vs. all other classes using the hypergeometric test are indicated by *, P<0.05; **, P<0.01; ***, P<0.001. doi:10.1371/journal.ppat.1000644.g002

(PF10_0383 and PF11_0033) that contain a 7-amino acid repeat unit (MYXSPDY); a serine-repeat antigen (SERA-3, PFB0350c) [5,50]; an integral membrane protein (PFL1140w); a kinesin motor encoding gene (PF11_0478); two further genes (PFB0095c, PFB0100c) in addition to those above found on the left end of chromosome 2 which has been shown previously to be prone to a large deletion containing these genes [51–53], including two genes associated with knob formation [54–56]; and five genes of unknown function (PF07_0076, MAL8P1.210, PFL0325w, PFL0600w, PFL0655w). In addition, two mitochondrial DNA-encoded mitochondrial genes coding for cytochrome oxidase I and III were found to be up regulated in laboratory strains which indicates that the long term culture adaptation may alter the utilization of mitochondria for parasite growth (data not shown).

Thus among the genes that differed most between field and laboratory isolates, there was a preponderance of those coding for processes likely to be important for parasite fitness *in vivo* such as export of proteins to the red cell surface, merozoite surface proteins, production of sexual stage forms for transmission to new hosts, and adaptation to the pregnant vs. non-pregnant host environment. It was also noticed that many of these highly regulated genes were located in regions of the genome in which deletion and amplification events have been reported to occur. Both these aspects are explored more rigorously below.

Functional classification of genes that differ between field and laboratory isolates

We found that genes coding for proteins involved in sexual stage parasites, those known to be expressed on the surface of the

parasite, and those thought to be exported to the red cell surface membrane (members of the exportome, Table S1) were significantly over-represented among the genes significantly different between field and laboratory strains (Figure 2B). Most of these had higher expression levels in field isolates than in laboratory isolates. Other classes of genes involved in basic metabolic processes such as transporters, genes involved in replication, transcription, translation and protein fate were not well represented among genes that differed between field and laboratory isolates. When the exportome genes were classified into their sub-families [57], it was found that all of these classes were significantly over-represented in field isolates with two exceptions - those belonging to the hydrolase and FIKK protein kinase groups (Figure 2C). Since the number of genes per class was small and the overall proportion significant was low, this result cannot be strongly interpreted as lack of strain variability in expression of these genes. Nonetheless, they are notable for their contrasting pattern to other members of the exportome, and suggest that these genes are involved in biological processes essential to *in vivo* survival.

Chromosomal clustering of differentially regulated genes

Significant chromosomal location clustering of genes that differed between field and laboratory strains was found in 30 regions of the genome; 19 of these were in the subtelomeric regions (Figure 3A, Figure S6). Clusters were also found for genes that varied significantly within field strains, 13 of 33 which overlapped with those for between-group comparisons (Figure S7). Clustering of genes that share similar timing of expression [20,22], similar function [20,24,25], similar response to drugs [36] and

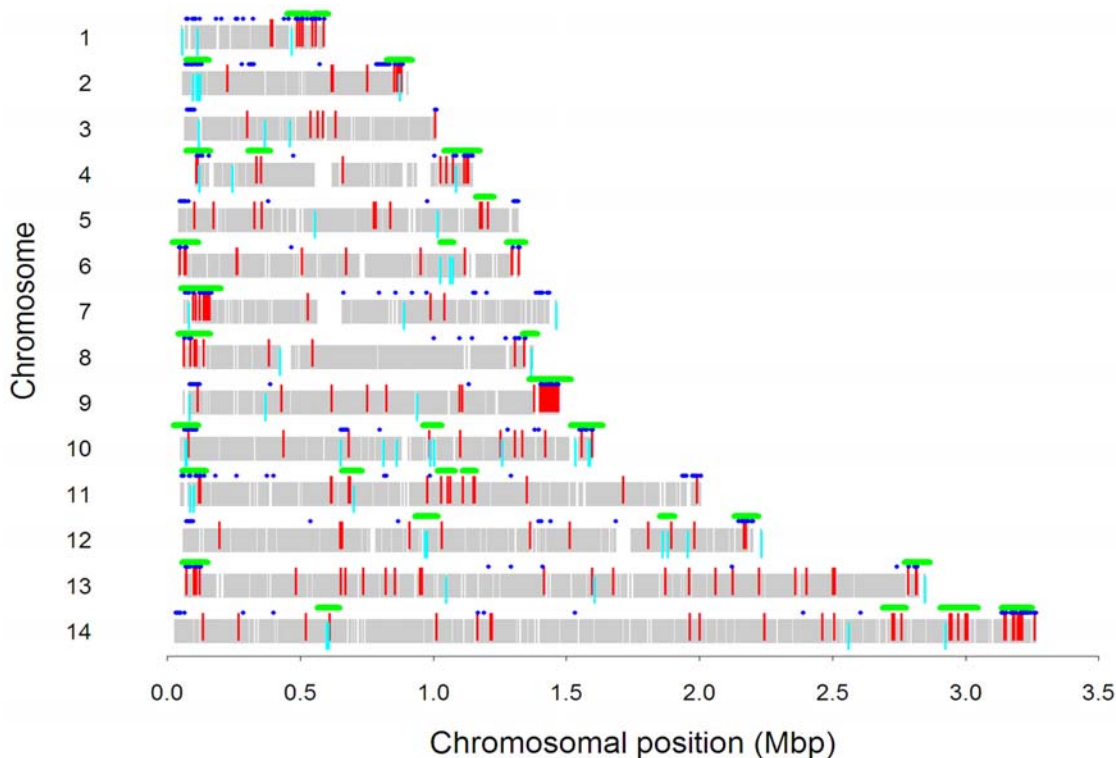


Figure 3. Chromosomal Location of Genes Differentially Expressed between Field and Laboratory Strains. Chromosomal locations of genes that had significantly ($P < 0.01$, > 1.5 -fold difference) higher (red) or lower (light blue) expression levels in field vs. laboratory strains in at least half of their probes. Regions in which there were significantly more ($P < 0.05$) differentially regulated genes than expected by chance (i.e., clustering) are indicated by a green bar above the chromosome. Genes belonging to the exportome are indicated by blue dots above the chromosome. Vertical grey bars indicate the locations of genes included in these analyses. doi:10.1371/journal.ppat.1000644.g003

general chromosomal location (subtelomeric vs. centromeric) [32,36] has been reported previously. Here we show physical clustering within short chromosomal regions of genes that exhibit strain variability between strains in their expression. This implies co-regulation of genes occupying the same genomic region. Possible reasons for such clustering are through chromatin-related epigenetic control of transcription in the region [20], or the existence of deletion or amplification events in genomic DNA segments. Many reports of the latter, termed copy number variants (CNVs), have been reported previously in laboratory and fresh field isolates of *P. falciparum* based on either microarray hybridizations of genomic DNA [19,50,53,58] or genetic studies of single genes or chromosomal regions [51,52,59–63]. To investigate the role of CNVs in chromosomal clustering of gene expression patterns, we mapped onto the genome all gene deletions and gene amplification events previously reported in the literature, and performed comparative genomic hybridization (CGH) of genomic DNA from our field isolates to detect further regions showing variability between strains (Dataset S2). From these data we generated a comprehensive map of genes that display copy number variation (Table S2, Figure S5). We included in this list, as a separate category, the nine ‘segmental duplications’ (SDs) consisting of four genes in tandem described by Lavazec et al. (2006) and Mok et al. (2008) which also display copy number variation [25]. A total of 448 genes in 264 separate CNVs were defined (see Materials and Methods), 154 of which have not been reported previously. CNVs displayed significant clustering in 22 regions of the genome, 10 of which overlapped with expression clusters (Figure S5).

Associations between gene expression and gene copy number

We overlaid on the CNV map all the genes in our study that showed strain variation in expression levels, both between field and laboratory isolates groups and within field strains (Figure S10) and observed that many of these coincided. Therefore we jointly analyzed expression and genomic DNA (CGH) data from across field strains to determine the degree of correspondence between genomic content and expression levels. We found that across-strain associations between expression levels and genomic copy number tended to be more often strongly positive or strongly negative than expected by chance (Figure 4A, B). Positive correlations between expression and DNA copy number are generally expected because of the direct effects of gene dosage. In most cases here, positive associations appeared to be due to deletions (Figure 4C). By contrast, negative associations between expression and CGH signal for some genes indicate that lower copy number leads to higher expression. This is possibly due to their release from gene repression. We cannot say whether these apparently repressed genes had been amplified in 3D7 relative to field isolates or whether, instead, field isolates contain sequence variation in the probes for genes and therefore, from CGH data, appear to be absent. The former explanation seems more probable given that the probes for this array were designed around non-variable regions, and because only those genes for which the majority of probes showed significance, i.e., consistency across probes in the same gene, were included in the analysis. Genes showing negative associations between expression and gene copy number were more prevalent in exported genes than those not showing gene copy number variation ($P < 0.001$ by Fisher’s Exact test) but less so than for genes showing positive associations ($P < 0.01$) These data suggest that during *in vitro* adaptation, both deletions and amplifications arise that cause, respectively, loss or suppression of gene expression in genes required for *in vivo*

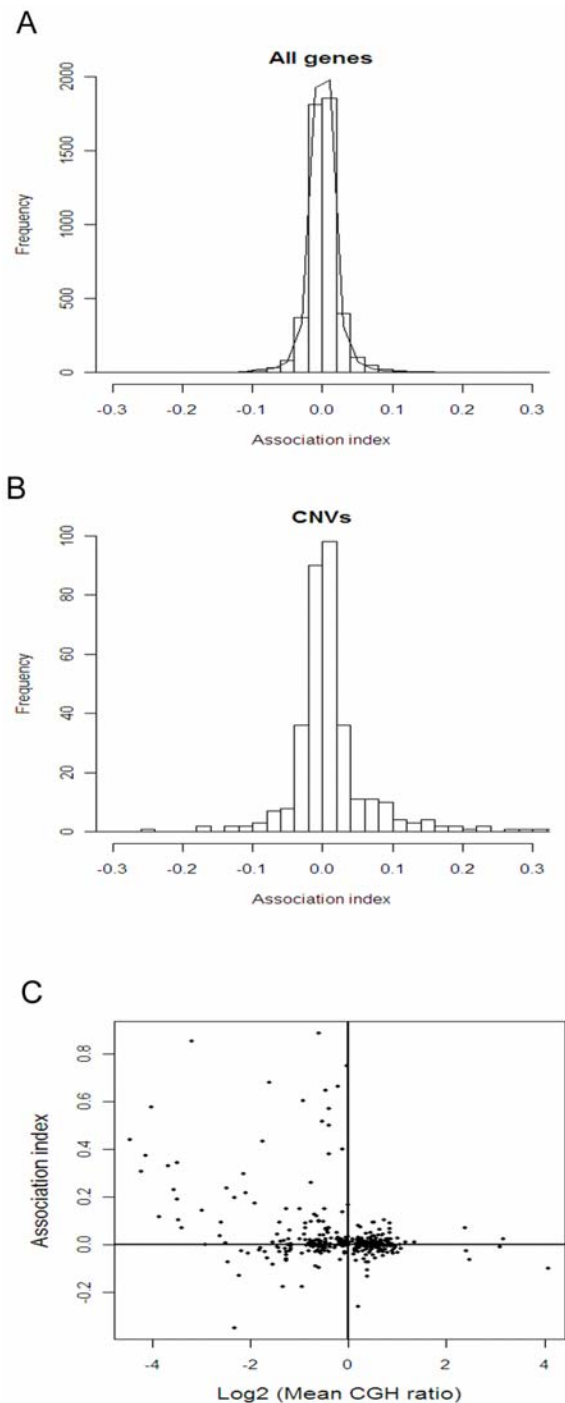


Figure 4. Associations Between Expression Levels and Genomic DNA Content among Field Isolates. Regression coefficients of normalized hybridization signals from cDNA on normalized hybridization signals from gDNA in the 5 field isolates were calculated for each gene. **A.** The frequency distribution of these coefficients (association index) across all genes was more dispersed than expected by chance (black line, obtained by randomly shuffling the data across isolates within each gene, $P < 0.001$ by Kolmogorov-Smirnov test). **B.** Genes that showed variation at the genome level (CNVs) had a higher frequency of strong positive and strong negative associations than genes that did not ($P < 0.001$ by Kolmogorov-Smirnov test). **C.** Genes with reduced genomic content relative to 3D7 had stronger associations, both positive and negative, with expression levels than genes with equivalent genomic content to 3D7. Only data from genes showing significant variation at the genomic level are shown in **C.**
doi:10.1371/journal.ppat.1000644.g004

processes. Possible mechanisms underlying amplification-mediated repression may be negative feedback effects of the gene's protein product on its own transcription, or chromatin-mediated suppressive effects that are induced by the CNV's structure, as discussed below.

Co-regulation between genes inside CNVs and adjacent genes

To investigate further CNV-related transcriptional regulation in relation to chromosomal clustering, we calculated the proportion of differentially expressed genes (between field and laboratory isolates and among field isolates) as a function of distance from CNV boundaries. Genes contained inside and close to CNVs located within subtelomeres, but not centromeric regions were significantly more often differentially expressed than those far from the CNV boundaries (Figure 5A). The same pattern was observed at the genomic level except that, in addition to subtelomeric CNVs, centromeric CNVs also showed (by definition) differential genomic content (Figure 5A).

We then calculated the correlations in expression levels between genes inside CNVs with those for genes outside CNVs. Associations in expression between pairs of genes both of which were located inside the same CNV were on average more positive than expected at random (Figure 5B, $P < 0.01$ by t-test). Associations in expression levels between pairs of genes inside the CNV and outside the CNV boundary were more strongly negative or positive than by chance alone (Figure 5B), a pattern that grew weaker as distance from the CNV boundary increased ($P < 0.001$ for the regression slope of absolute values of the observed association on distance from the CNV boundary, compared with $P = 0.14$ for randomized values; $P < 0.05$ for the interaction term testing for differences between these slopes). For these tests, only differentially expressed test genes and index genes were included in the analysis: this was to rule out bias in the association index caused by the higher frequency of differentially expressed genes closer to CNVs (Figure 5A). However, when all non-differentially expressed test genes were included, these significance values were only mildly reduced. To determine whether the co-regulation around the CNVs might be the direct result of undetected gene copy number variation in regions near to the CNVs, we performed the same association-distance analysis on data from genomic DNA. While there was a strong association in gene copy number among pairs of genes contained inside the defined CNVs ($P < 0.001$, when compared by t-test with the randomized values), no such association was found for genes outside the CNV (Figure 5B, $P > 0.5$ for regression of associations on distance for both observed and randomized values). Thus we conclude that co-regulation of genes surrounding CNVs with those inside CNVs acts through a mechanism that is not a direct effect of genomic content of the surrounding genes.

For SDs, the pattern of co-regulation was similar to that for CNVs though more striking despite being based on many fewer observations (Figure S9B, $P < 0.01$ for the interaction term using data from only differentially expressed test genes and all SDs whether the index gene was significant or not). This pattern was not due to the influence of adjacent CNVs (Figure 5B), of which there were a large number containing differentially expressed genes, because we excluded test genes falling inside CNVs from the analysis. Nor was the pattern due to genomic variation because the slope of association of genomic content by distance from SDs was not significantly different from zero ($P = 0.5$). Unlike for CNVs, however, the proportion of genes differentially regulated inside the SDs, and the strength of co-regulation among them, was relatively

low (Figure S9). Nevertheless, it appears that SDs are somehow involved in co-regulation of genes that surround them.

Large deletions

Two of the regions that showed strongest clustering and co-regulation of genes were the large deletions on the left arm of chromosome 2 and right arm of chromosome 9. Both of these deletions have been previously well documented, mainly in cultured laboratory isolates [51,64–66] but also in field isolates after short-term culture [67], as in this study. Their typical breakpoints have been well mapped in earlier studies [51,65,68,69]. However, in our study, three observations did not correspond to those from previous reports.

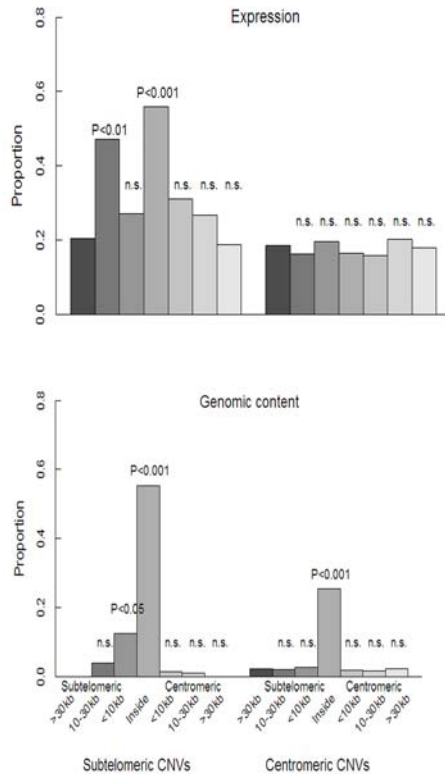
First, the region of co-regulation appeared to extend beyond the breakpoint in both deletions (Figure S8). The breakpoint in the chromosome 2 deletion is typically found in PFB0100c, the *kharp* gene: even though extensions to include its neighbor on the centromeric side, PFB0105c, have also been observed in some laboratory strains [53,62], our expression data and CGH data on the sub-lines that also contained the deletion (Figure S8) were consistent with the breakpoint falling inside PFB0100c. However, positive correlations between expression of genes inside the deletion and outside it extended for another 32kb containing 7 genes. Similarly, in chromosome 9, our expression data concur with previous data showing that the breakpoint is in *bporf* (PFI1710w), but the positive correlations in expression in this study extended for 12kb to include 3 more genes (Figure S8).

Second, expression of genes in the far telomeric ends of both deleted regions did not seem to be abrogated (PFB0056c, PFB0070c, PFI1785w, PFI1790w, PFI1795c, PFI1800w). This is surprising because previous studies of these and other deletions have shown that chromosomal breakages in the sub-telomeres are healed by the direct joining of telomeric sequences to the breakpoint [51,52,60,70] in which case all genes on the telomeric side of the breakpoint would have zero expression levels. Third, these large deletions were found in three of our five field isolates despite the essentiality of many of the genes in these regions for *in vivo* processes such as gametocyte production [67,71,72], cytoadherence [67,69,70,73,74], the formation of knobs [51,54,64,75,76], maintenance of cell deformability [77] and host cell remodeling for protein trafficking [41]. The chromosome 2 deletion has not previously been found in field isolates even after short-term culture adaptation [50,52,58,67] although the chromosome 9 deletion has [67]. Our field isolates used for expression analyses were cultured for up to 18 weeks and for CGH data for a further 8–24 weeks (Table 1). It is probable that the mutations arose in culture and were selectively favored through having higher replication rates: the relatively high mutation rates of these deletions in culture makes this a likely possibility [78]. Alternatively, these deletions may have existed at very minor frequencies within their infected hosts in which case they would not be detectable using conventional methods immediately after isolation from the host. The fact that independent sub-lines of P1 diverged for the chromosome 2 deletion, and that the chromosome 9 deletion appeared upon further rounds of culture in all 5 field isolates (Figure S8), illustrates the particular vulnerability of parasite genomes to deletions in these regions and to their frequency changes within cultures. If the same applies to other CNVs, this highlights the need to define the biological role of CNVs in isolates taken directly from patients.

Malaria-in-pregnancy genes

Of the 31 genes previously reported to be associated with pregnancy in malaria (PAM, Table S3), 9 are located within a

A



B

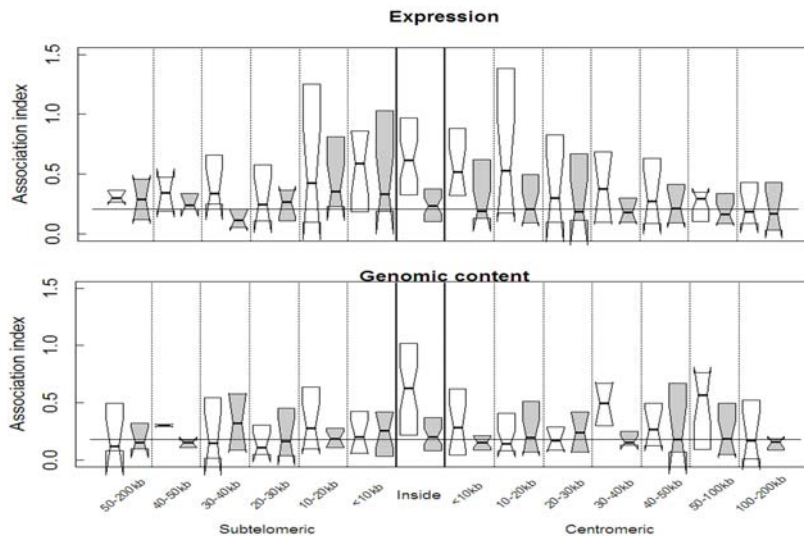


Figure 5. Co-regulation of Genes Surrounding CNVs. **A.** Proportion of genes significantly differentially regulated between field and laboratory isolates or among field isolates (upper), and showing significant between-strain variation at the genomic level (lower), grouped by distance of the gene from the CNV boundary on either the subtelomeric side or centromeric side. P-values (by Fisher's Exact test) indicate where the proportion of significant genes differ from that >30kb on the subtelomeric side of the CNV. Analyses have been split by whether the CNV occurs in the subtelomeric (left) or internal (right) regions of the chromosome. **B.** Across-strain (field and laboratory strains) measures of association (absolute values of regression coefficients, white bars) between expression levels (upper, white bars) of a single differentially expressed gene within a CNV (index gene) and expression levels in differentially expressed genes surrounding the CNV (test gene) grouped by distance from the CNV boundary. The same analysis was performed for genomic content data (lower). For comparison to that expected by chance, associations after randomly permuting the data across strains within genes are shown in grey boxes. The horizontal line inside each box shows the median, the box boundaries show the interquartile range, the whisker length is one interquartile range, the box width is proportional to the square root of the number of observations per group, and the notches show the approximate 95% confidence intervals, i.e., non-overlapping notches strongly support a hypothesis of non-equivalence [135]. The horizontal line across all the boxes indicates the median value for the association index of permuted values on all genes after excluding data for genes inside the CNV, i.e., the expected value of the association index from chance alone for test genes outside CNVs. Regression analysis of expression data from individual genes showed a significant decline in the magnitude of across-strain associations with distance from the CNV boundary ($P<0.001$) which was stronger than for randomized data ($P<0.05$), and was not observed for genomic content data ($P>0.5$). (See main text for details).

doi:10.1371/journal.ppat.1000644.g005

Table 1. *P. falciparum* Strains and Experimental Design Used in This Study.

Purpose	Strain & sub-line	Description	Method of cDNA/gDNA preparation	No. of cycles in culture	No. of replicate hybridizations
Testing amplification method					
	^a 3D7	Laboratory	Unamplified	Many	2
Transcriptome analysis					
	3D7	Laboratory	Amplified	Many	2
	^b Dd2	Laboratory	Amplified	Many	1
	^c T9/96	Laboratory	Amplified	Many	1
	^d P1	Field, severe	Amplified	62	3
	^d P3	Field, severe	Amplified	63	3
	^d P4	Field, mild	Amplified	48	3
	^d P5	Field, mild	Amplified	66	2
	^d P6	Field, mild	Amplified	59	3
Gene copy number analysis					
	^d P1.A	As above	Unamplified	^e +21	3
	^d P3.A		Unamplified	+24	3
	^d P4.A		Unamplified	+31	3
	^d P5.A		Unamplified	+21	3
	^d P6.A		Unamplified	+23	3
	^d P1.B		Unamplified	+8	1
	^d P4.B		Unamplified	+14	1
	^d P6.B		Unamplified	+8	1

^aDerived from a patient in the Netherlands living near the airport [131]. Believed to be of African origin.

^bObtained by cloning of strain W2-mef, a clone of Indochina III/CDC that had been under long-term mefloquine pressure [132].

^cObtained from Thailand [133,134].

^dObtained from Kilifi, Kenya, from patients admitted to the hospital with severe or mild malaria.

^eThe "+" indicates the number of cycles of culture in addition to those in the top part of the table.

doi:10.1371/journal.ppat.1000644.t001

CNV, and 19 are located outside but within 50kb of a CNV boundary. Of the 31, 13 showed differential expression between strains (11 of these between field and laboratory isolates and 2 among field strains). Those PAM genes that were up-regulated in pregnant women vs. non-pregnant hosts in other studies tended to be up-regulated genes in field isolates compared with laboratory isolates (Table S3), whereas those down-regulated in pregnant women were down-regulated in field vs. laboratory isolates. Notably, the two genes most consistently found to be up-regulated in previous studies (PFB0115w, PFI1785w) fall just outside the boundaries of the large deletions on chromosomes 2 and 9 (Figures S8, S10) and within the regions in which positive co-regulation with the CNV genes remained elevated (Figure S8). Thus these pregnancy-associated genes may be examples of *cis*-acting CNV-mediated gene regulation that allows the parasite to rapidly adapt to new, immunity-free host niches.

Drug resistance genes

A further example of a link between CNVs and differential gene expression is the internal region of chromosome 12 where 3 of the genes inside or very close to a CNV (denoted CNV12H, Table S2) displayed significant between-strain variation in expression (Figure S10, Table S2). One of these genes (PFL1155w) codes for GTP-cyclohydrolase I (*gch1*) which is the first enzyme in the folate synthesis pathway. Higher copy number of genes in the vicinity of PFL1155w was previously reported in several laboratory strains compared with field strains by Kidgell et al. (2006) [50]. Two alternative explanations were offered for this. Either this was generated by the different composition of folate synthesis

precursors in the *in vitro* culture media compared to serum, in which case the parasite would need to up-regulate its synthesis to compensate for not being able to scavenge it from the host. Alternatively, amplification in laboratory strains arose as a result of selection for antifolate resistance *in vitro*, while field isolates, in their study, had not. Recent evidence from field isolates sampled from areas with high and low antifolate drug pressure supports the second hypothesis [17]. However, our data are more consistent with the first hypothesis for several reasons. First, our isolates were obtained from an area where antifolate drugs had been in use for many years and have generated a very high level of resistance [79]. Second, we found high expression of *gch1* in 3D7, which is antifolate-sensitive, as for other laboratory isolates and consistent with the higher copy number reported by Kidgell et al. (2006) [50]. Third, examination of expression of the 8 other genes in the folate synthesis pathway and its co-pathway – the shikimate pathway - revealed that two of them also were significantly down-regulated in field isolates, viz., dihydrofolate synthase (PF13_0140, *dhfs*, $P < 0.05$) and bifunctional dihydrofolate reductase (PFD0830w, *dhfr*, $P < 0.001$). This suggests that there was general up-regulation of the folate synthesis pathway in laboratory isolates compared to field isolates. This support for the first hypothesis does not preclude the second also being true: direct comparisons of expression levels of this gene and copy number in resistant vs. sensitive isolates are required to determine the role of *gch1* copy number in antifolate resistance. The data do nevertheless highlight the role of gene copy number variation in parasite adaptation, of which drug resistance is just one example [17,80–82], as well as the need to explore its genetic basis in natural field isolates.

Discussion

In the process of validating microarray technology for the analysis of transcriptional variation in natural populations of malaria parasites, we found supporting evidence for many of the results from previous whole genome studies [13–16,19,20,24,31,33–36], despite many differences in experimental design, materials and methodology. Examples of differentially expressed genes discovered in multiple independent studies include those involved in a range of biological processes important to parasite fitness *in vivo* such as export of proteins to the red cell surface, transmission, and colonization of the pregnant woman, transcriptional regulation factors, nutrient biosynthesis and metabolic pathways involved in drug resistance. This highlights the robustness of the microarray approach to the quantitative analysis of transcriptional variation in natural populations of malaria parasites.

Prior to whole genome screens, genes within several large or small deletions had been shown to be important for parasite virulence and transmissibility [51,52,59–61,63–65,67,69,71–74,83]. With whole genome screens, many more smaller deletions and amplifications have been discovered in field and laboratory isolates [19,31,50,58,62]. Here we show that much of the biological variation in transcription between field and laboratory isolates, and among field isolates, is attributable to genes located within and around CNVs [14,24]. We showed that gene copy number in these CNVs associates strongly with their level of expression, either strongly promoting it or suppressing it. Intriguingly, the expression levels of genes outside these CNVs are associated with expression levels of genes contained inside the CNVs. By analyzing the genomic DNA around the CNVs we were able to rule out the possibility that this co-regulation was due to undetected copy number variation in the genes outside the CNVs. We therefore suggest that co-regulation of genes adjacent to CNVs occurs through an indirect effect of the CNV.

We envisage a number of ways in which CNVs might alter the expression, or be associated with expression of genes in their vicinity. These fall under the ‘dynamic’ vs. ‘structural’ explanations for the epigenetically-mediated phenomenon of position effect variegation (PEV) that is found in many organisms (reviewed in [84,85]). PEV refers to the maintenance and inheritance of silenced states of genes within some somatic lineages of an organism, but not others, despite there being no differences at the genetic level. Regulation of *var* genes in *P. falciparum*, in which there is mutually exclusive repression of all but one of the many genes over repeated cycles of asexual replication resembles this phenomenon [6]. The ‘dynamic’ model invokes spreading of heterochromatin from the telomeres into the adjacent regions. In *P. falciparum*, the existence of a graded band of heterochromatin in the nuclear periphery where sub-telomeric genes are silenced [6,86], and *cis* spreading of it into adjacent regions [87,88] has been demonstrated. This band is considerably wider than in other organisms where this has been observed, such as yeast (~50kb vs. ~5kb) [6,86], thus providing scope for simultaneous silencing of many sub-telomeric genes in *P. falciparum*. In addition to the maintenance of a heterochromatin environment, it has been shown that histone modifications alter the degree of the heterochromatin’s repressive effect on genes in this region in *P. falciparum* [89–92]. Such heterochromatin-modifying effects, such as nucleosome assembly during replication, histone modification, DNA methylation and RNA interference are common in other organisms and are thought to maintain the reversibility of silencing by heterochromatin (reviewed in [93]). Applying the dynamic model to our results, we postulate that CNVs may act as insulators to chromatin spreading, or as modifiers to its effectiveness, as is

common in other organisms (reviewed in [94]). Increased copy number in tandem repeat non-coding sequences have been shown to increase their repressive effect on neighboring genes in a diverse range of organisms, even at long range, by changing the relative position of genes and the heterochromatin regions (reviewed in [95]). It is possible that copy number variation in coding genes acts in a similar way in *P. falciparum*.

The structural model of PEV proposes that somatic pairing of homologous chromosomes (in the case of diploids) and of heterologous chromosomes causes the recruitment of heterochromatin and other DNA-binding proteins to the temporarily joined regions. If chromosomal rearrangements occur, or copy number variation in coding or non-coding regions, cause changes in the pairing alignment, this heterochromatin formation is altered, thus resulting in PEV (reviewed in [84]). Such effects can be manifest at single loci as well as having long-range effects on multiple loci. In *P. falciparum*, formation of clusters or “bouquets” of heterologous chromosomes has been shown to occur during mitosis and meiosis [96–98] and to be involved in recombination amongst *var* genes, presumably in order to generate new diversity [96]. (SDs may play such a role for generating diversity in the *pfmc-2TM* genes which contain a hypervariable region that is likely to be exposed to the immune system [9]). Applying the structural model to our data, we speculate that formation of junctions between heterologous chromosomes, if they rely on multi-copy genes contained in or surrounding CNVs, may mediate the recruitment of heterochromatin to these regions and thereby modify the expression of genes in the vicinity. Sub-telomeric deletion events have been shown to disrupt bouquet formation [97]. There is also evidence suggesting the existence of transcriptionally permissive sub-compartments within the silencing zone on the nuclear periphery into which sub-telomeres might be transferred in order to reverse the silencing [6,87,99]. Thus a third model to explain our observations on co-regulation are that CNVs or SDs in some way interact with the boundaries of these sub-compartments, thereby altering the positioning of surrounding genes relative to the silencing/permissive zones. Much remains to be explored on the mechanisms and consequences of heterochromatin recruitment and modification in *Plasmodium*. Doing so could lead to powerful new drugs that disrupt heterochromatin-mediated regulation of important homeostatic and adaptive mechanisms in the parasite [100].

What might be the reasons for maintenance of variation in CNV genes in natural populations? The findings of this study support the view that these unstable regions in the genome are important to parasite adaptation in its *in vivo* environment and so are maintained in nature. In the case of the large deletions on chromosomes 2 and 9, it is likely that *in vitro* culture conditions relieve the parasite from the presumably costly but necessary *in vivo* exercise of exporting proteins, constructing knobs, displaying cytoadhesive molecules and variant antigens, and producing gametocytes (which cost the parasite future asexual replication). Their ready appearance under such conditions therefore suggests that they can have a growth advantage under some conditions. One of those conditions might be during the acute stage of infection of a naive host when the short-term race against competitor parasite lineages for host cells and escape from innate immune mechanisms is of paramount importance to the parasite’s fitness. Knobs are generally assumed to be essential for parasite survival because they mediate cytoadherence which, evidence in both human and other malaria species suggests, allow the parasite to avoid clearance by the spleen [101–107]. However, knobless parasites are able to cytoadhere to some degree [70,108], and there may be other mechanisms of cytoadherence not mediated by

knobs that help the parasite to sequester [109]. Therefore, rather than viewing knobs as being universally favored, we postulate that they have both benefits and costs which vary during the course of an infection. During the acute stage of the infection the benefit may be escape from immune attention and intrinsically higher replication rate, while at the chronic stage the benefits are likely to be the ability to express new ligands and antigenic types to keep ahead of emerging variant-specific antibodies. Thus we view the postulated maintenance of large chromosomal deletions at low frequencies within hosts as arising from a play-off between a short-term advantage to winning the within-host competition for host resources versus the long-term advantage of winning the competition to transmit between hosts. This fits with two well-established evolutionary theories of why highly virulent pathogens are seen in nature despite their fitness costs through host mortality [110–114], both of which have some empirical support in malaria (reviewed in [4]). Shorter CNVs, whether deletions or amplifications, may play a similar role in adaptation to niche environments as appears to be the case for some drug resistance genes ([14,24,50] and this study).

CNVs may also play a role in regulation of non-CNV adaptive genes that surround them. It is striking that the two genes most consistently associated with placental malaria are found on the boundaries of the chromosome 2 and 9 deletions (Figure S8) and half of those reported to be associated with placental malaria are located in other CNVs or within 50kb of them. Another example may be the deletion-mediated switching on of rosetting [78], a parasite virulence-related adhesive property [115–118] which, given its widespread prevalence in nature, and the fact that it varies during an infection (at least in the rodent malaria, *P. chabaudi*, [119]) and responds to immune selection [120], suggests it may play an adaptive role. There may be many such other phenotypic, adaptive consequences of alterations in gene copy number. The existence of shared molecular mechanisms associated with these deletions, such as small regulatory elements upstream of the breakpoints [51,52,60], similar gene inversion mechanisms to create them [60,63], and healing of breakages by pasting of telomeric sequences that alter the positioning of the telomeres within the heterochromatin environment [97] supports the view that these mechanisms are maintained by the parasite to allow it to adapt to its variable host environments, both within an infection in the same host and across infections in hosts of different immune/pregnancy/genetic or other susceptibility states [75].

CNVs may also play a role in regulation of genes very distant to them or even on other chromosomes. A recent study of genome-wide transcription of a genetic cross between a drug resistant and drug sensitive line identified 12 hotspots in the genome containing genes that appear to regulate many other genes on different chromosomes to the CNV [24]. One of these regulators occurred in a region containing a CNV that is related to drug resistance [24]. Most of the 12 regulatory hotspots were found inside or close to CNVs. Thus it seems that CNVs might have genome-wide *trans-acting* effects. Other forms of master regulators of transcription in *P. falciparum* have been proposed to work through binding of specific DNA motifs of genes expressed at different stages of the parasite's asexual cycle, such as those in the ApiAP2 family [39,40]. Three of this family of 26 genes fall within CNVs and four of them showed strain variability in expression in the study here. Thus combined, our data and those from previous studies present a picture of highly coordinated transcription of genes involved in functional processes related to parasite adaptation to its *in vivo* environment. Understanding the source of this co-regulation will provide clues as to how the parasite has maintained its extraordinary evolutionary success and may, perhaps, provide new targets for intervention.

Materials and Methods

Ethics statement

Collection of parasite samples from patients was approved by the KEMRI Ethical Review Committee.

Parasite material and preparation of RNA and DNA

Five malaria parasite isolates (denoted P1, P3, P4, P5 and P6) were obtained from malaria-diagnosed inpatients and outpatients at the Kilifi District Hospital, Kenya, and adapted to *in vitro* conditions through continuous culture for 48–66 cycles (Table 1). During culture adaptation and for the experiments reported here, parasites were grown at 2–3% hematocrit with 0.5% Albumax II as a replacement for human serum and 0.1mM hypoxanthine in the standard malaria culture media [121]. At the point of collection from the patient the minimum number of genotypes per isolate were 3, 1, 3, 2 and 1 for P1, P3, P4, P5 and P6 respectively, as determined by genotyping at the MSP-2 locus [122] but appeared to consist of a single genotype by the time the experiments here were performed. Cultures of these lines were expanded to a volume of 3–4ml packed cell volume (PCV), synchronized by sorbitol treatment in two consecutive cycles, and early in the following cycle were diluted to 1% parasitemia with fresh red blood cells and split into six separate flasks containing 75ml and 0.25ml PCV. At six 8-hour intervals starting at approximately 4 hours post-invasion, cells were washed once in warm PBS, snap-frozen on dry ice, then stored at -80°C . Similar preparations were made from the laboratory strains 3D7, Dd2 and T9/96.

RNA was prepared from frozen samples after lysing them in a water bath at 62°C for 3 minutes, and then adding 2.5ml of Trizol (Invitrogen) while keeping on ice. After 10 minutes of mixing by pipetting, the RNA was extracted using two chloroform treatments, followed by overnight precipitation with isopropanol/sodium acetate precipitation. An average of 16 μg (range 2–64) of RNA was obtained from each 0.25ml PCV culture sample and used at an average concentration of 0.8 $\mu\text{g}/\mu\text{l}$ in the subsequent amplification step. A pool of material from 3D7 for use as the common reference in the microarray hybridizations was constructed by pooling equal quantities of RNA collected at the 6 time-points.

Genomic DNA for comparative genomic hybridization (CGH) of the field isolates to the array was obtained by growing up each line used in the transcriptome analyses for a further 8–24 culture cycles (Table 1). Two cultures of each line were maintained independently by two different people to create “sub-lines”. Competitive hybridizations against clone 3D7 of one set of independent set sub-lines from all 5 strains, and 3 strains from the other set, were performed using standard methods [19] in triplicate for the first set of sub-lines and in one replicate for the second set. Strain identity was confirmed using the hybridization patterns of the hypervariable *rifin*, *var* and *stevor* genes.

Amplification of RNA using template switching and microarray hybridizations

For analyses of the transcriptome, we modified the previously reported method Switch Mechanism at the 5' end of Reverse Transcription (SMART) [37] to obtain sufficient target cDNA for microarray assays from as little as ~ 100 ng of total RNA. Briefly 100ng of total RNA was subjected to reverse transcription that was primed by two oligonucleotides (pdN9-tag, 5'-AAG CAG TGG TAT CAA CGC AGA GTN NNN NNN NN-3' and dT25-tag, 3'-AAG CAG TGG TAT CAA CGC AGA GTA C (T)₂₅ -3') using Powerscript (Clontech) for reverse transcription. To

compensate for the AT rich bias of the *Plasmodium* genes, the dNTPs mix was enriched for A-T at a ratio of 2:1. After the first-strand synthesis, the generated cDNA was used in a second PCR with the tag oligonucleotide (3'-AAG CAG TGG TAT CAA CGC AGA GTA C-5') and a hybrid DNA:RNA (d(AAG CAG TGG TAT CAA CGC AGA GTA CGC)r(GGG)). In this step, aminoallyl dUTPs were incorporated into the dNTP mix (dATP:dTTP:aadUTP:dCTP:dGTP = 2:1:1:1:1) for the subsequent coupling of fluorescent dyes. cDNA was hybridized to a 70mer oligonucleotide microarray [30] as previously described [19]. Hybridizations were performed in triplicate or duplicate for all isolates except Dd2 and T9/96 for which single replicates were performed. Microarrays were scanned with a GenePix 4000B scanner and the images analyzed using GenePix Pro 3.0 software (Axon Instruments, Union City, CA).

Pre-analysis of expression and CGH data

The initial raw microarray data were first filtered by eliminating poor quality spots (manual inspection) and low signal spots (median signal in both Cy3 and Cy5 channel > median background signal +3 standard deviations of the background signal as defined by the GenePix software. For expression data, in order to standardize the variance and mean in expression ratios across the range of background intensities, the “normexp” [123,124] and the “robustspline” methods from the Limma package [124] were applied to each array separately (within-array normalization) (Figure S3). The data were then variance-normalized between arrays using the “scale” method in Limma because it was found that the variability for 3D7 was much lower than for other strains [125] (Figure S3). The reason for this lower variance was because of the positive correlation, and hence restricted variance, between the expression ratios of the sample and pooled reference material as the latter was created by pooling the former. For CGH data, it was found necessary to normalize differently to that used for expression data because of the zero hybridization levels for some genes compared with 3D7. Within-array normalization was performed by subtraction of background intensities from foreground intensities and between-array normalization was performed using the “quantile” method to ensure equal empirical distributions across arrays.

Data from the hypervariable gene families (*rifin*, *var* and *stevor*, and fragments and pseudogenes of these, collectively termed variant surface antigens, VSA) were excluded from the analysis prior to normalization because many were not recognized by the probes which were based on 3D7 VSA sequences. Data from genes coding for ribosomal RNAs and transfer RNAs were also excluded because of potential error caused by differences in the dynamic range of the signal.

All statistical analyses and graphs were generated using the R programming language [126]. The raw filtered data have been deposited in the Array Express database (www.ebi.ac.uk/microarray-as/ae/).

Validating the RNA amplification method. To describe the faithfulness of SMART amplification in reproducing expression differences, we compared expression profiles of all genes using cDNA prepared directly from large cultures by the usual single-step first-strand synthesis method (“Unamplified”) to that from amplified material using SMART. To examine the effects of amplification on time profiles of expression of the same gene, Pearson correlation coefficients between the two methods were calculated across the six time-points for each probe. To examine the effects of amplification on gene expression levels relative to all the other genes, Pearson correlations and Spearman rank correlations between the two methods (Unamplified vs.

Amplified) were calculated across all probes within each of the 6 time points. The observed across-time point and across-gene correlations were compared to distributions of correlations obtained after 1000 random permutations of the data across time points and genes respectively. Regression analysis of data from amplified on unamplified material from all probes and time points was used to determine whether quantitative differences between genes were faithfully reproduced by amplification. For these analyses, data on the *rifins*, *vars* and *stevors* were retained as we wanted to determine whether there was amplification bias in genes with very low relative expression levels.

Estimating the time zero for each strain. Since each strain was synchronized and harvested separately there were potential differences in the stage of maturity (phase) of each strain at the starting time point. Therefore we estimated the phase of each strain by finding the minimum mean Euclidean distance between data on 6 time points in our strains and data from the 48-hour transcriptome data set [15,19].

Testing for differences in expression among strains. To statistically test for differences among strains, a linear model with fixed effects for time point and strain was fitted to data on each probe separately using the Limma package in R [124]. A block effect was included in the model to allow for correlations between data from repeat hybridizations of the same cDNA material for each strain-time point combination [127]. For each probe, contrasts were formed between each pair of strains, and between the grouped laboratory vs. grouped field strains. These differences were tested for statistically significant deviations from zero using modified t-tests and F-tests in which, instead of using the error variance for each probe, the pooled error variance across all probes, estimated using a Bayesian approach [128], is substituted. P-values on these modified t-tests were adjusted for multiple testing using the Benjamini and Hochberg [129] (BH) method which gives the required false discovery rate, i.e., the rate of false positives among the declared positives. For the comparison of field vs. laboratory isolates, those probes with $P < 0.01$ and a fold-change of greater than 1.5 (i.e., 50% up- or down-regulation) were declared significant. For declaring significant variation among field isolates, a cut-off value of $P < 0.001$ by F-test was used. For secondary analyses that required data on a gene-wise rather than probe-wise basis, only genes with greater than 50% of their probes were taken to be significant.

The strain coefficient from the linear model above was used as a summary measure for each strain: this coefficient represents the average expression level across time-points. Hierarchical clustering of these summary data was applied to whole-genome expression profiles to obtain a strain dendrogram that represented relative differences among strains. Differences among strains were also described using a principal coordinates plot [38] which represents dissimilarities among strains as distances in a strain ‘map’. To create this map, the matrix of Pearson correlations between each strain’s whole-genome profile were computed, and the Euclidean distances from this matrix were used as the similarity measure in a multidimensional scaling analysis using the “cmdms” command in R [126]. Here we used a 3-dimensional map because it explained considerably more of the variation than a 2-dimensional map.

Testing for differences in genomic content among strains by CGH and definitions of copy number variants (CNVs). As for expression data, a model was fitted to each gene with a fixed effect for each sub-line (8 lines representing 5 isolates) with allowance for correlations among replicate hybridizations. Genes with F-tests significant at the $P < 0.001$ level after correcting for multiple comparisons and at least two lines different from 3D7 by

at least 1.5-fold in the majority of probes per gene were declared as a copy number variant (CNV). Significant genes that were adjacent to other significant genes in this study or previously published CNVs were assigned to the same CNV (Table S2).

Testing for chromosomal clustering of differentially expressed genes. To examine for chromosomal clustering of genes that were differentially expressed between field and laboratory strains, for each chromosome a sliding window average expression difference was calculated using a window width of 50kb and an increment of 5kb. To statistically test for clustering, the probability of getting equal or greater than the observed number of significant genes within the window, given the number of genes that fell within that window and the proportion of all genes in the genome that showed significance (excluding those in this window), was calculated from the hypergeometric distribution, i.e. sampling from a finite pool of differently labeled objects without replacement. Clustering was declared to be present if this probability fell below $P < 0.01$. These clustering tests were performed on both field vs. laboratory isolate comparisons and comparisons among field isolates.

Testing for associations between differential expression and gene function. Genes were classified according to functional classes using the gene ontology (GO) annotations obtained from version 5.5 of PlasmoDB. The genes were further classified based on lists compiled from specialist studies on genes with related functions, viz., the proteins involved in membrane transport (the “permeome” [11]), protein kinases (the “kinome” [12]), and proteins exported out of the parasite into the host cell either into the parasitophorous vacuole, the erythrocyte cytosol or the red cell surface (the “exportome”, [7,8,57,130]). The criteria used for classification are given in Table S1. When genes fell into more than one class, they were assigned to the class at the highest level in Table S1. Over-representation of significantly differentially expressed genes in these classes relative to all other classes was assessed using the hypergeometric test.

Analysis of co-regulation. To determine whether there were associations between the expression of genes inside CNVs and SDs with other genes inside them and genes outside their boundaries, the across-strain (field and laboratory isolates) correlations and regression coefficients were computed for pairs of data from an index gene inside the CNV or SD (chosen to be that with the lowest P-value for the test of a difference between field and laboratory isolates in the case of the CNV, and the *hyp4* gene in the case of the SDs) and all other genes 200kb either side of the CNV boundary (test genes). The absolute value of the regression coefficient was taken as a measure of the strength of association and was analyzed by linear regression as a function of distance from the CNV or SD boundary. The same analysis was performed on CGH data (using data from all eight sub-lines of the field isolates) to determine whether any associations seen in expression data were attributable to variation at the genomic level.

Supporting Information

Figure S1 Estimation of the Timing of Each Strain Relative to the 48-hour Transcriptome. For each strain and each gene, Euclidean distances between the 6-timepoint expression profiles from this study and the quality-controlled data set from the 48-hour transcriptome [19] were calculated for each of the 48 hour time points (x-axis). The mean Euclidean distance across all genes is on the y-axis. Where this distance reaches a minimum gives the best fit of the data to the reference set, i.e., the best estimate for the stage of maturity at which the samples were harvested at the first time point. Due to variation in the reference data set, there were

systematic differences (i.e., across all strains) between time points in the distance measure. Therefore the curves were smoothed using a quadratic loess fitting method with an effective number of parameters of 10. The minima were then found from the smoothed curves. The best estimates ranged from 45h (–3h) to 4h with an average of 0.67h across strains. Therefore the first time point is labeled as zero hours in all other figures showing time profiles of expression data.

Found at: doi:10.1371/journal.ppat.1000644.s001 (2.17 MB TIF)

Figure S2 Evaluation of the SMART RNA Amplification Technique for Use in Whole-Transcriptome Analyses of *P. falciparum*. Genome-wide gene expression levels were analyzed by DNA microarrays using parasite material taken from a synchronized culture of the laboratory-adapted strain, 3D7, every 8 hrs during the parasite’s 48hr asexual cycle. cDNA for hybridizing to the array was prepared from total RNA either using the normal first strand reverse transcription method (“Unamplified”) or by SMART-PCR (“Amplified”) (see Materials and Methods). A. Heat maps of gene expression levels across time points (x-axis) using Unamplified and Amplified material demonstrates faithful reproduction of gene expression time profiles after amplification. Only probes exhibiting at least 2-fold change during the 48hr cycle are included in the heat map. Rows (probes) are ordered according to expression levels at the first time point (0hrs) in the Unamplified material. Expression values are mean-normalized across rows. Relative expression values on the \log_2 scale are shown in the color bar below the plots. B. Scatterplot of expression values over all probes and time points using Unamplified (x-axis) and Amplified (y-axis) material. Density shading has been used to replace individual points in high density regions: 5000 individual points are shown in the least dense areas. Slopes of regression lines (dashed lines) fitted separately to data with x-values above and below zero were both significantly ($P < 0.001$) lower than the expected value of 1 (solid line) indicating that RNA amplification leads to underestimation of relative expression differences. The boxed area indicates the few genes that were markedly preferentially amplified. Closer analysis revealed that these contain repetitive sequences that are also shared across genes: therefore their data were excluded from the main analyses. C. The frequency distribution of observed Pearson correlations between each probe’s expression time-profile using Unamplified vs. Amplified material (bars) shows high correspondence between time profiles for the two methods. 57% of the probes had correlations greater than 0.8. Permutation analysis (solid line) found that this would occur by chance 3.6% of the time. D. Distributions of observed correlations between Unamplified and Amplified material across all probes at each time point ranged from 0.76 to 0.92 (solid dots for Pearson correlations, open circles for Spearman rank correlations, increasing in y-value from first to last time point). These were much higher than the range of –0.03 to 0.03 expected by chance (solid line, Pearson correlations obtained by permutation analysis). *Note for Figure S2.* A small subset of probes appeared to be preferentially amplified in strain 3D7 (Figure 1B). Of the 17 data points highlighted as over amplified, 15 of these fell within 2 time points, and they represented 11 genes. Since these over-amplification events were observed in both hybridizations, and these events were restricted to 2 time points and were consistent across multiple probes per gene, it seems likely that they arose as an artifact of the PCR conditions during amplification. Of the 11 genes showing amplification, four were paralogous genes coding for hypothetical proteins (MAL8P1.335, PFC1125w, PFC0002c, and PFD1250w). The two probes for these genes all had very similar sequences, and so it is likely that these genes were cross-hybridizing. These genes also contain repetitive sequences that are similar which may also have contributed to over-amplification. As these genes do not

appear in the sequence database of strains other than 3D7, and were not expressed in strains other than 3D7 (our results), it is concluded that these were not real genes but are instead artifacts in the 3D7 genome annotation. They were therefore ignored in further analyses. Another gene that was over amplified, PF10_0374 that codes for a large gametocyte-expressed protein Pfl1-1 and has 10 probes, had multiple probes that displayed over amplification. This gene contains a repetitive sequence that evolves rapidly to produce strain variation [136–138] and which is also found in other antigens (antigen 332 PF11_0507 and RESA). Thus the expression differences based on this gene should be considered unreliable, as supported by discrepant results for this gene in 3D7 from other studies [50,58]. Probes for the remaining 6 genes displaying over-amplification (MAL8P1.38, PFL1210w, PF11_0330, PF11_0462, PF14_0107, and PF14_0235) showed no homology with other sequences in the database and so cannot be explained. However, only one of these (PF11_0462) showed significant differences between strains (higher in field than laboratory isolates, even when 3D7 was ignored) and so these genes were retained for further analysis. Nevertheless, the existence of amplification bias in a few genes highlights the need for highly specific oligonucleotide probes that avoid repetitive or low complexity regions [19], and for independent verification of genes that show expression differences using unamplified material [58].

Found at: doi:10.1371/journal.ppat.1000644.s002 (3.53 MB TIF)

Figure S3 Quality Control and Normalizations. A. Expression ratio (y-axis) vs. mean intensity (x-axis) plots from 1 array on one field isolate before and after the “normexp” normalization and “robustspline” within-array normalization procedures were applied [123,124]. Data points shown in black are those that were filtered out due to poor quality spots or screened out because they coded for VSAs or non-messenger RNA species. Data points shown in red are those probes included in the final analysis. B. Box-plots of data from each of 108 arrays before and after between-array normalization using the “scale” method in Limma [124]. Arrays 41 and 70 were excluded from analyses due to poor quality hybridizations. Slides 85–96 are for strain 3D7 (6 time-points, 2 replicates each).

Found at: doi:10.1371/journal.ppat.1000644.s003 (5.38 MB TIF)

Figure S4 Volcano Plot Showing Relationship Between Expression Differences and P-values in Comparison of Field vs. Laboratory Isolates. Individual points are for probes and there are multiple probes per gene. The significance criteria of 1.5-fold expression and $P < 0.01$ are indicated by the vertical and horizontal solid dashed lines, respectively, not allowing for multiple testing. After allowing for multiple testing, 6.8% (647/9493) of probes were significant. After including only genes for which the majority of their probes were significant, 259/4896 (5.3%) of genes were declared significant.

Found at: doi:10.1371/journal.ppat.1000644.s004 (0.57 MB TIF)

Figure S5 Chromosomal Location of CNVs and SDs and Functional Classification of Genes Contained Within Them. A. Chromosomal locations of genes that had significant variation among field isolates ($P < 0.01$, > 1.5 -fold difference in at least two field isolates and the majority of their probes significant by these criteria) in genomic DNA content as measured by CGH. Pink bars indicate those CNVs detected in this study and brown bars indicate CNVs reported in previous studies (Table S2). Orange bars indicate the locations of SDs. Regions in which there were significantly more ($P < 0.05$) CNVs than expected by chance (i.e., clustering) are indicated by a dark green bar above the chromosome. Clusters found in expression data (Figure 3) are shown in light green for comparison. Vertical grey bars indicate

the locations of genes analyzed in this study. B. Genes were classified according to function based on their PlasmoDB annotations and GO terms, and bioinformatics studies (Table S1). The proportion of each class that were located in CNVs are shown for new CNVs found in this study (pink) and CNVs reported from previous studies (brown). C. As for B but classified according to their class within the exportome [57]. Numbers per class are shown in the x-axis labels. Significance levels for differences in proportions of each class vs. all other classes using the hypergeometric test are indicated by *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Found at: doi:10.1371/journal.ppat.1000644.s005 (1.90 MB TIF)

Figure S6 Sliding Window Plots of Average Expression Differences (\log_2 scale) Along Each Chromosome. Differences in expression between field vs. laboratory strains averaged over a sliding window of a 0.5Mbp region are shown as solid lines. The horizontal line marks the expected value under the null hypothesis of there being no difference between the groups. Expression differences for individual genes that were significant ($P < 0.01$, > 1.5 -fold change) are shown as dots. Regions where there were significantly more genes than expected by chance ($P < 0.01$) after allowing for gene densities are indicated by horizontal green bars. The vertical bars in the top of each panel represent the locations of all the genes in the genome.

Found at: doi:10.1371/journal.ppat.1000644.s006 (1.90 MB TIF)

Figure S7 Chromosomal Location of Genes Differentially Expressed within Field Strains. Chromosomal locations of genes that showed significant variation ($P < 0.001$ by F-test after correcting for multiple comparisons and which had the majority of their probes significant) in expression levels among field strains (black bars). Regions in which there were significantly more ($P < 0.05$) differentially regulated genes than expected by chance (i.e., clustering) are indicated by a yellow-green bar above the chromosome. Genes belonging to the exportome are indicated by blue dots above the chromosome. Vertical grey bars indicate the locations of genes included in these analyses.

Found at: doi:10.1371/journal.ppat.1000644.s007 (1.70 MB TIF)

Figure S8 Gene Expression Levels and Genomic DNA Content for Each Parasite Isolate for Genes in the Region of the Chromosome 2 and 9 Deletions. Expression levels (left axis, upper panel) are expressed as differences from the mean over all time points for strain 3D7. Genomic content levels (lower panel) are expressed relative to 3D7 and are shown only for field isolates. The vertical dotted line indicates the site of the breakpoint as reported in previous studies. For Chromosome 2, based on expression levels, the deletion appears to have occurred in strains P1 and P3 beginning at PFB0100c (*kharp*) or PFB0105c as reported previously (marked by a vertical dashed line) and extends through to PFB0075c. This was confirmed by CGH in sub-lines P1.A and P3.A. In Chromosome 9, from the expression data, the laboratory strains Dd2 and T9/96 appear to have a deletion between genes PFI1710w (*bporf*) and PF1780w, consistent with previous reports of the breakpoint [41]. Strain P5 appears to be carrying a mixture of deleted and non-deleted parasites. These breakpoints were confirmed in CGH analysis of the field strains, all of which carried the deletion in at least one sub-line after further culturing. In the expression data (upper panel), the black dashed line shows the across-strain correlations (right axis) between expression profiles for an index gene inside the deleted region (PFB0105c or PFI1780w) and other genes in the region. The correlations remain positive beyond the deletion boundaries, indicating co-regulation of genes neighboring the deletion. Two strongly down-regulated genes adjacent to the deletion boundaries, PFB0115w and

PFI1785w, have been shown in previous studies to be up-regulated in parasites obtained from pregnant women (Table S3).

Found at: doi:10.1371/journal.ppat.1000644.s008 (2.45 MB TIF)

Figure S9 Co-regulation of Genes Surrounding SDs. A. Proportion of genes significantly differentially regulated between field and laboratory isolates or among field isolates (upper) and showing significant between-strain variation at the genomic level (lower) grouped by distance of the gene from the SD boundary on either its subtelomeric side or centromeric side. P-values (by Fisher's Exact test) indicate where the proportion of significant genes was different from that >30kb centromeric to the SD boundary (right-hand bar in each group of bars). Results are shown with (left) and without (right) data from genes contained within CNVs that are not also SDs. B. Across-strain (field and laboratory strains) measures of association (regression coefficients, white bars) between expression levels (upper) of a single gene within a SD (the index gene, *hyp4*, or for SD11LA, gene PF11_0015, not necessarily differentially expressed) and expression levels of genes surrounding the SD (test gene, not necessarily differentially expressed) grouped by distance from the SD boundary. The same analysis was performed for genomic content data (lower). To compare these results with those expected by chance, associations after randomly permuting the data across strains within genes were computed (grey bars). The horizontal line inside each box shows the median, the box boundaries show the interquartile range, the whisker length is one interquartile range, the box width is proportional to the square root of the number of observations per group, and the notches show the approximate 95% confidence intervals, i.e., non-overlapping notches strongly support a hypothesis of non-equivalence [135]. Regression analysis of the data from individual genes, using only gene that were differentially expressed in order to rule out bias due to higher numbers of significant genes nearer to SDs, showed a significant decline in the strength of association with distance from the SD boundary ($P < 0.001$). This decline was stronger than for data that had been randomized ($P < 0.01$) for expression data. A decline in association strength with distance from the SD in genomic content data was not observed ($P > 0.5$). (See main text for details).

Found at: doi:10.1371/journal.ppat.1000644.s009 (4.01 MB TIF)

Figure S10 Regional Genomic Maps of CNVs, SDs, Differentially Expressed Genes, Expression Clusters, CNV Clusters and Pregnancy-Associated Genes from This and Other Studies. Grey bars indicate the locations of those genes analysed in this study. Superimposed black bars mark those genes differentially regulated within field isolates. In the next row up, red and light blue bars

indicate genes up- and down-regulated genes in field vs. laboratory isolates, respectively. Pink bars in the third row mark CNVs found in this or other studies. In the fourth row orange bars represent SDs. Steel blue bars in the fifth row up show genes reported to be associated with placental malaria. Blue dots below the first row indicate exportome genes. Green horizontal lines marks regions where significant clustering ($P < 0.05$) of genes differentially expressed between field and laboratory isolates were found, while yellow-green lines indicate significant clustering of CNVs.

Found at: doi:10.1371/journal.ppat.1000644.s010 (0.14 MB PDF)

Table S1 Terms used for classifying genes according to function. String searches were applied to the fields of Product Description, GO Annotated Function, GO Annotated Process and GO Annotated Component downloaded from the list of updated annotations in PlasmoDB version 5.5.

Found at: doi:10.1371/journal.ppat.1000644.s011 (0.05 MB DOC)

Table S2 CNVs and SDs identified in this and other studies.

Found at: doi:10.1371/journal.ppat.1000644.s012 (0.82 MB PDF)

Table S3 Differential expression results from this study for genes associated with pregnancy malaria in previous studies.

Found at: doi:10.1371/journal.ppat.1000644.s013 (0.07 MB DOC)

Dataset S1 Raw data from microarray hybridization of cDNA from 3 laboratory and 5 field isolates from *P. falciparum* taken at 6 time points during the intra-erythrocytic cycle.

Found at: doi:10.1371/journal.ppat.1000644.s014 (4.59 MB ZIP)

Dataset S2 Raw data on microarray hybridizations of genomic DNA for 5 field isolates of *P. falciparum*.

Found at: doi:10.1371/journal.ppat.1000644.s015 (0.98 MB ZIP)

Acknowledgments

Vandana Thathy provided very useful discussion of the work and Gordon Smyth advised on statistical analysis. Teresa Tiffert, Virgilio Lew, Karl Hoffmann and David Dunne are thanked for making laboratory facilities available and for their advice and support. This paper is published with the permission of the Director of the Kenyan Medical Research Institute.

Author Contributions

Conceived and designed the experiments: MJM ZB. Performed the experiments: MJM JL SM MMK ZB. Analyzed the data: MJM ZB. Contributed reagents/materials/analysis tools: MJM ZB. Wrote the paper: MJM KM PRP ZB.

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 434: 214–217.
2. White NJ (1998) Drug resistance in malaria. *Br Med Bull* 54: 703–715.
3. Gandon S, Mackinnon MJ, Nee S, Read AF (2001) Imperfect vaccines and the evolution of parasite virulence. *Nature* 414: 751–755.
4. Mackinnon MJ, Read AF (2004) Virulence in malaria: an evolutionary viewpoint. *Phil Trans R Soc Lond B* 359: 965–986.
5. Maier AG, Rug M, O'Neill MT, Brown M, Chakravorty SJ, et al. (2008) Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell* 134: 46–61.
6. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, et al. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 121: 13–24.
7. Marti M, Good RT, Rug M, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* 306: 1930–1933.
8. Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, et al. (2004) A host-targeting signal in virulence proteins reveals a secretome in malaria infection. *Science* 306: 1934–1937.
9. Lavazec C, Sanyal S, Templeton TJ (2006) Hypervariability within the Rifin, Stevor and Pfinc-2TM superfamilies in *Plasmodium falciparum*. *Nucleic Acids Res* 34: 6696–6707.
10. Ansari FA, Kumar N, Bala SM, Gnanamani M, Ramachandran S (2008) MAAP: malarial adhesins and adhesin-like proteins predictor. *Proteins* 70: 659–666.
11. Martin RE, Henry RI, Abbey JL, Clements JD, Kirk K (2005) The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. *Genome Biol* 6: R26.
12. Ward P, Equinet L, Packer J, Doerig C (2004) Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics* 5: 79.
13. Daily JP, Scanfeld D, Pochet N, Le Roch K, Plouffe D, et al. (2007) Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature* 450: 1091–1095.
14. Jiang H, Patel JJ, Yi M, Mu J, Ding J, et al. (2008) Genome-wide compensatory changes accompany drug-selected mutations in the *Plasmodium falciparum crt* gene. *PLoS ONE* 3: e2484. doi:10.1371/journal.pone.0002484.
15. Francis SE, Malkov VA, Oleinikov AV, Rossnagle E, Wendler JP, et al. (2007) Six genes are preferentially transcribed by the circulating and sequestered

- forms of *Plasmodium falciparum* parasites that infect pregnant women. *Infect Immun* 75: 4838–4850.
16. Ndam NT, Bischoff E, Proux C, Lavstsen T, Salanti A, et al. (2008) *Plasmodium falciparum* transcriptome analysis reveals pregnancy malaria associated gene expression. *PLoS ONE* 3: e1855. doi:10.1371/journal.pone.0001855.
 17. Nair S, Miller B, Barends M, Jaidec A, Patel J, et al. (2008) Adaptive copy number evolution in malaria parasites. *PLoS Genet* 4: e1000243. doi:10.1371/journal.pgen.1000243.
 18. Ben Mamoun C, Gluzman IY, Hott C, MacMillan SK, Amarakone AS, et al. (2001) Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite *Plasmodium falciparum* revealed by microarray analysis. *Mol Microbiol* 39: 26–36.
 19. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu JC, et al. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1: e5. doi:10.1371/journal.pbio.0000005.
 20. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301: 1503–1508.
 21. Hayward RE, DeRisi JL, Aljadhi S, Kaslow DC, Brown PO, et al. (2000) Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum*. *Mol Microbiol* 35: 6–14.
 22. Scholz M, Fraunholz MJ (2008) A computational model of gene expression reveals early transcriptional events at the subtelomeric regions of the malaria parasite, *Plasmodium falciparum*. *Genome Biol* 9: R88.
 23. Spielmann T, Beck HP (2000) Analysis of stage-specific transcription in *Plasmodium falciparum* reveals a set of genes exclusively transcribed in ring stage parasites. *Mol Biochem Parasitol* 111: 453–458.
 24. Gonzales JM, Patel JJ, Pongmee N, Jiang L, Tan A, et al. (2008) Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol* 6: e238. doi:10.1371/journal.pbio.0060238.
 25. Mok BW, Ribacke U, Sherwood E, Wahlgren M (2008) A highly conserved segmental duplication in the subtelomeres of *Plasmodium falciparum* chromosomes varies in copy number. *Malar J* 7: 46.
 26. Scherf A, Lopez-Rubio JJ, Riviere L (2008) Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* 62: 445–470.
 27. Deitsch K, Duraisingh M, Dzikowski R, Gunasekera A, Khan S, et al. (2007) Mechanisms of gene regulation in *Plasmodium*. *Am J Trop Med Hyg* 77: 201–208.
 28. Bozdech Z, Zhu JC, Joachimiki MP, Cohen FE, Pulliam BL, et al. (2003) Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* 4: R9.1–R9.14.
 29. Gardner MJ, Hall N, Fung E, White O, Berriman M, et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511.
 30. 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.
 31. Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 34: 1166–1173.
 32. Le Roch KG, Zhou Y, Batalov S, Winzeler EA (2002) Monitoring the chromosome 2 intraerythrocytic transcriptome of *Plasmodium falciparum* using oligonucleotide arrays. *Am J Trop Med Hyg* 67: 233–243.
 33. Daily JP, Le Roch KG, Sarr O, Fang X, Zhou Y, et al. (2004) *In vivo* transcriptional profiling of *Plasmodium falciparum*. *Malaria J* 3: 30.
 34. Daily JP, Le Roch KG, Sarr O, Ndiaye G, Lukens A, et al. (2005) *In vivo* transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. *J Infect Dis* 191: 1196–1203.
 35. Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, 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 U S A* 106: 7559–7564.
 36. Natalang O, Bischoff E, Deplaine G, Proux C, Dillies MA, et al. (2008) Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* 9: 388.
 37. Petalidis L, Bhattacharyya S, Morris GA, Collins VP, Freeman TC, et al. (2003) Global amplification of mRNA by template-switching PCR: linearity and application microarray analysis. *Nucleic Acids Res* 31: e142.
 38. Gower J (1966) Some distance properties of latent roots and vector methods used in multivariate analysis. *Biometrika* 53: 325–338.
 39. Balaji S, Babu MM, Iyer LM, Aravind L (2005) Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* 33: 3994–4006.
 40. De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, et al. (2008) Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc Natl Acad Sci U S A* 123: 8393–8398.
 41. Spielmann T, Hawthorne PL, Dixon MW, Hannemann M, Klotz K, et al. (2006) A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in *Plasmodium falciparum* codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. *Mol Biol Cell* 17: 3613–3624.
 42. Hawthorne PL, Trenholme KR, Skinner-Adams TS, Spielmann T, Fischer K, et al. (2004) A novel *Plasmodium falciparum* ring stage protein, REX, is located in Maurer's clefts. *Mol Biochem Parasitol* 136: 181–189.
 43. Wickham ME, Rug M, Ralph SA, Klonis N, McFadden GI, et al. (2001) Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. *EMBO J* 20: 5636–5649.
 44. Kneupfer E, Rug M, Klonis N, Tilley L, Cowman AF (2005) Trafficking determinants for PfEMP3 export and assembly under the *Plasmodium falciparum*-infected red blood cell membrane. *Mol Microbiol* 58: 1039–1053.
 45. Kneupfer E, Rug M, Klonis N, Tilley L, Cowman AF (2005) Trafficking of the major virulence factor to the surface of transfected *P. falciparum*-infected erythrocytes. *Blood* 105: 4078–4087.
 46. Kriek N, Tilley L, Horrocks P, Pinches R, Elford BC, et al. (2003) Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes. *Mol Microbiol* 50: 1215–1227.
 47. Bhattacharjee S, van Ooij C, Balu B, Adams JH, Haldar K (2008) Maurer's clefts of *Plasmodium falciparum* are secretory organelles that concentrate virulence protein reporters for delivery to the host erythrocyte. *Blood* 111: 2418–2426.
 48. Fidock DA, Bottius E, Brahimi K, Moelans II, Aikawa M, et al. (1994) Cloning and characterization of a novel *Plasmodium falciparum* sporozoite surface antigen, STARP. *Mol Biochem Parasitol* 64: 219–232.
 49. Fried M, Hixson KK, Anderson L, Ogata Y, Mutabingwa TK, et al. (2007) The distinct proteome of placental malaria parasites. *Mol Biochem Parasitol* 155: 57–65.
 50. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, et al. (2006) A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog* 2: e57. doi:10.1371/journal.ppat.0020057.
 51. Polge LG, Ravetch JV (1988) Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* 55: 869–874.
 52. Scherf A, Mattei D (1992) Cloning and characterization of chromosome breakpoints of *Plasmodium falciparum*: breakage and new telomere formation occurs frequently and randomly in subtelomeric genes. *Nucleic Acids Res* 20: 1491–1496.
 53. Carret CK, Horrocks P, Konfortov B, Winzeler EA, Qureshi M, et al. (2005) Microarray-based comparative genomic analyses of the human malaria parasite *Plasmodium falciparum* using Affymetrix arrays. *Mol Biochem Parasitol* 144: 177–186.
 54. Crabb BS, Cooke BM, Reeder JC, Waller RF, Caruana SR, et al. (1997) Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* 89: 287–296.
 55. Pasloske BL, Baruch DI, van Schravendijk MR, Handunnett SM, Aikawa M, et al. (1993) Cloning and characterization of a *Plasmodium falciparum* gene encoding a novel high-molecular weight host membrane-associated protein, PfEMP3. *Mol Biochem Parasitol* 59: 59–72.
 56. Pasloske BL, Howard RJ (1994) Malaria, the red cell, and the endothelium. *Annu Rev Med* 45: 283–295.
 57. Sargeant TJ, Marti M, Caler E, Carlton JMR, Simpson K, et al. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol* 7: R12.
 58. Ribacke U, Mok BW, Wirta V, Normark J, Lundeberg J, et al. (2007) Genome wide gene amplifications and deletions in *Plasmodium falciparum*. *Mol Biochem Parasitol* 155: 33–44.
 59. Petersen C, Nelson R, Magowan C, Wollish W, Jensen J, et al. (1989) The mature erythrocyte surface antigen of *Plasmodium falciparum* is not required for knobs or cytoadherence. *Mol Biochem Parasitol* 36: 61–65.
 60. Polge LG, de Bruin D, Ravetch JV (1990) A and T homopolymeric stretches mediate a DNA inversion in *Plasmodium falciparum* which results in loss of gene expression. *Mol Cell Biol* 10: 3243–3246.
 61. Welles TE, Walliker D, Smith CL, Rosario VE, Maloy WL, et al. (1987) A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of *Plasmodium falciparum*. *Cell* 49: 633–642.
 62. Lavazec C, Sanyal S, Templeton TJ (2007) Expression switching in the stevor and Pfmc-2TM superfamilies in *Plasmodium falciparum*. *Mol Microbiol* 64: 1621–1634.
 63. Cappai R, van Schravendijk MR, Anders RF, Peterson MG, Thomas LM, et al. (1989) Expression of the RESA gene in *Plasmodium falciparum* isolate FCR3 is prevented by a subtelomeric deletion. *Mol Cell Biol* 9: 3584–3587.
 64. Biggs BA, Kemp DJ, Brown GV (1989) Subtelomeric chromosome deletions in field isolates of *Plasmodium falciparum* and their relationship to loss of cytoadherence *in vitro*. *Proc Natl Acad Sci U S A* 86: 2428–2432.
 65. Shirley MW, Biggs BA, Forsyth KP, Brown HJ, Thompson JK, et al. (1990) Chromosome 9 from independent clones and isolates of *Plasmodium falciparum* undergoes subtelomeric deletions with similar breakpoints *in vitro*. *Mol Biochem Parasitol* 40: 137–145.
 66. Polge LG, Ravetch JV (1986) A chromosomal rearrangement in a *Plasmodium falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* 322: 474–477.
 67. Day KP, Karamalis F, Thompson J, Barnes D, Brown H, et al. (1993) Virulence and transmissibility of *Plasmodium falciparum* map to chromosome 9. *Proc Natl Acad Sci USA* 90: 8292–8296.
 68. Barnes DA, Thompson J, Triglia T, Day K, Kemp DJ (1994) Mapping the genetic locus implicated in cytoadherence of *Plasmodium falciparum* to melanoma cells. *Mol Biochem Parasitol* 66: 21–29.
 69. Bourke PF, Holt DC, Sutherland CJ, Kemp DJ (1996) Disruption of a novel open reading frame of *Plasmodium falciparum* chromosome 9 by subtelomeric and

- internal deletions can lead to loss or maintenance of cytoadherence. *Mol Biochem Parasitol* 82: 25–36.
70. Horrocks P, Pinches R, Chakravorty SJ, Papakrivovs J, Christodoulou Z, et al. (2005) PfEMP1 expression is reduced on the surface of knobless *Plasmodium falciparum* infected erythrocytes. *Journal of Cell Science* 118: 2507–2518.
 71. Gardiner DL, Dixon MW, Spielmann T, Skinner-Adams TS, Hawthorne PL, et al. (2005) Implication of a *Plasmodium falciparum* gene in the switch between asexual reproduction and gametocytogenesis. *Mol Biochem Parasitol* 140: 153–160.
 72. Alano P, Read D, Bruce MC, Aikawa M, Kaido T, et al. (1995) Cos cell expression cloning of *Pf377*, a *Plasmodium falciparum* gametocyte antigen associated with osmiophilic bodies. *Mol Biochem Parasitol* 74: 143–156.
 73. Trenholme KR, Gardiner DL, Holt DC, Thomas EA, Cowman AF, et al. (2000) *clag9*: a cytoadherence gene in *Plasmodium falciparum* essential for binding of parasitized erythrocytes to CD36. *Proc Natl Acad Sci USA* 97: 4029–4033.
 74. Chaiyaroj SC, Coppel RL, Magowan C, Brown GV (1994) A *Plasmodium falciparum* isolate with a chromosome 9 deletion expresses a trypsin-resistant cytoadherence molecule. *Mol Biochem Parasitol* 67: 21–30.
 75. Ravetch JV (1989) Chromosomal polymorphisms and gene expression in *Plasmodium falciparum*. *Exp Parasitol* 68: 121–125.
 76. Kilejian A, Olson J (1979) Proteins and glycoproteins from human erythrocytes infected with *Plasmodium falciparum*. *Bull World Health Organ* 57 Suppl 1: 101–107.
 77. Glenister FK, Coppel RL, Cowman AF, Mohandas N, Cooke BM (2002) Contribution of parasite proteins to altered mechanical properties of malaria-infected red blood cells. *Blood* 99: 1060–1063.
 78. Horrocks P, Kyes S, Pinches R, Christodoulou Z, Newbold C (2004) Transcription of subtelomerically located var gene variant in *Plasmodium falciparum* appears to require the truncation of an adjacent var gene. *Mol Biochem Parasitol* 134: 193–199.
 79. Certain LK, Briceno M, Kiara SM, Nzila AM, Watkins WM, et al. (2008) Characteristics of *Plasmodium falciparum* dhfr haplotypes that confer pyrimethamine resistance, Kilifi, Kenya, 1987–2006. *J Infect Dis* 197: 1743–1751.
 80. Price RN, Uhlemann AC, Brockman A, McGready R, Ashley E, et al. (2004) Mefloquine resistance in *Plasmodium falciparum* and increased pfmdr1 gene copy number. *Lancet* 364: 438–447.
 81. Sidhu AB, Uhlemann AC, Valderramos SG, Valderramos JC, Krishna S, et al. (2006) Decreasing pfmdr1 copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J Infect Dis* 194: 528–535.
 82. Nair S, Nash D, Sudimack D, Jaidee A, Barends M, et al. (2007) Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol Biol Evol* 24: 562–573.
 83. Scherf A, Carter R, Petersen C, Alano P, Nelson R, et al. (1992) Gene inactivation of Pfl1-1 of *Plasmodium falciparum* by chromosome breakage and healing: identification of a gametocyte-specific protein with a potential role in gametogenesis. *EMBO J* 11: 2293–2301.
 84. Henikoff S (1994) A reconsideration of the mechanism of position effect. *Genetics* 138: 1–5.
 85. Henikoff S (1990) Position-effect variegation after 60 years. *Trends Genet* 6: 422–426.
 86. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, et al. (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 121: 25–36.
 87. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, et al. (2006) A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 439: 1004–1008.
 88. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, et al. (2009) Sir2 paralogs cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biol* 7: e1000084. doi:10.1371/journal.pbio.1000084.
 89. Chookajorn T, Dzikowski R, Frank M, Li F, Jiwan AZ, et al. (2007) Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci U S A* 104: 899–902.
 90. Lopez-Rubio JJ, Riviere L, Scherf A (2007) Shared epigenetic mechanisms control virulence factors in protozoan parasites. *Curr Opin Microbiol* 10: 560–568.
 91. 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.
 92. Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, et al. (2009) *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res*.
 93. Vermaak D, Ahmad K, Henikoff S (2003) Maintenance of chromatin states: an open-and-shut case. *Curr Opin Cell Biol* 15: 266–274.
 94. Talbert P, Henikoff S (2006) Spreading of silent chromatin: inaction at a distance. *Nat Rev Genet* 7: 793–803.
 95. Henikoff S (1998) Conspiracy of silence among repeated transgenes. *Bioessays* 20: 532–535.
 96. Freitas-Junior LH, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, et al. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 407: 1018–1022.
 97. Figueredo LM, Freitas-Junior LH, Bottius E, Olivo-Marin JC, Scherf A (2002) A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J* 21: 815–824.
 98. Marty AJ, Thompson JK, Duffy MF, Voss TS, Cowman AF, et al. (2006) Evidence that *Plasmodium falciparum* chromosome end clusters are cross-linked by protein and are the sites of both virulence gene silencing and activation. *Mol Microbiol*.
 99. Ralph SA, Bischoff E, Mattei D, Sismeiro O, Dillies MA, et al. (2005) Transcriptome analysis of antigenic variation in *Plasmodium falciparum*—var silencing is not dependent on antisense RNA. *Genome Biol* 6: R93.
 100. Henikoff S, Vermaak D (2000) Bugs on drugs go GAGAA. *Cell* 103: 695–698.
 101. Hommel M, David PH, Oligino LD (1983) Surface alterations of erythrocytes in *Plasmodium falciparum* malaria: antigenic variation, antigenic diversity and the role of the spleen. *J Exp Med* 157: 1137–1148.
 102. David PH, Hommel M, Miller LH, Udeinya JJ, Oligino LD (1983) Parasite sequestration in *Plasmodium falciparum* malaria - spleen and antibody modulation of cytoadherence of infected erythrocytes. *Proc Natl Acad Sci USA* 80: 5075–5079.
 103. Barnwell JW, Howard RJ, Miller LH (1983) Influence of the spleen on the expression of surface antigens on parasitized erythrocytes. In: Evered D, Whelan J, eds. Ciba Foundation Symposium on Malaria and the red cell. London: Pitman. pp 117–132.
 104. Barnwell JW, Howard RJ, Coon HG, Miller LH (1983) Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned *Plasmodium knowlesi* malaria. *Infect Immun* 40: 985–994.
 105. Gilks CF, Walliker D, Newbold CI (1990) Relationships between sequestration, antigenic variation and chronic parasitism in *Plasmodium chabaudi chabaudi* - a rodent malaria model. *Parasite Immunol* 12: 45–64.
 106. Langreth SG, Peterson E (1985) Pathogenicity, stability and immunogenicity of a knobless clone of *Plasmodium falciparum* in Colombian owl monkeys. *Infect Immun* 47: 760–766.
 107. Handunneti SM, Mendis KN, David PH (1987) Antigenic variation of cloned *Plasmodium fragile* in its natural host *Macaca sinica*. *J Exp Med* 165: 1269–1283.
 108. Pasloske BL, Baruch DI, Ma C, Taraschi TF, Gormley JA, et al. (1994) PfEMP3 and HRP1: co-expressed genes localized to chromosome 2 of *Plasmodium falciparum*. *Gene* 144: 131–136.
 109. Pouvelle B, Buffet PA, Lepoat C, Scherf A, Gysin J (2000) Cytoadhesion of *Plasmodium falciparum* ring-stage-infected erythrocytes. *Nat Med* 6: 1264–1268.
 110. Levin BR, Svanborg Eden C (1990) Selection and evolution of virulence in bacteria: an ecumenical excursion and modest suggestion. *Parasitology* 100: S103–S115.
 111. Levin BR, Bull JJ (1994) Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol* 2: 76–81.
 112. Frank M, Dzikowski R, Costantini D, Amulic B, Berdugo E, et al. (2006) Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite *Plasmodium falciparum*. *J Biol Chem* 281: 9942–9952.
 113. Levin SA, Pimentel D (1981) Selection of intermediate rates of increase in parasite-host systems. *Am Nat* 117: 308–315.
 114. Anderson RM, May RM (1982) Co-evolution of hosts and parasites. *Parasitology* 85: 411–426.
 115. Carlson J, Helmby H, Hill AVS, Brewster D, Greenwood BM, et al. (1990) Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet* 336: 1457–1460.
 116. Rowe JA, Obeiro J, Newbold CI, Marsh K (1995) *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. *Infect Immun* 63: 2323–2326.
 117. Treutiger C, Hedlund I, Helmby H, Carlson J, Jepson A, et al. (1992) Rosette formation in *Plasmodium falciparum* isolates and anti-rosette activity of sera from Gambians with cerebral or uncomplicated malaria. *Am J Trop Med Hyg* 46: 503–510.
 118. Ringwald P, Peyron F, Lepers JP, Rabarison P, Rakptmalala C, et al. (1993) Parasite virulence factors during falciparum-malaria - rosetting, cytoadherence, and modulation of cytoadherence by cytokines. *Infect Immun* 61: 5198–5204.
 119. Mackinnon MJ, Walker PR, Rowe JA (2002) *Plasmodium chabaudi*: rosetting in a rodent malaria model. *Exp Parasitol* 101: 121–128.
 120. Mackinnon MJ, Gandon S, Read AF (2008) Virulence evolution in response to vaccination: the case of malaria. *Vaccine* 26 Suppl 3: C42–C52.
 121. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. *Science* 193: 673–675.
 122. Cattamanchi A, Kyabayinze D, Hubbard A, Rosenthal PJ, Dorsey G (2003) Distinguishing recrudescence from reinfection in a longitudinal antimalarial drug efficacy study: comparison of results based on genotyping of MSP-1, MSP-2 and GLURP. *Am J Trop Med Hyg* 68: 133–139.
 123. Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, et al. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 23: 2700–2707.
 124. Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer. pp 397–420.
 125. Smyth GK, Speed TP (2003) Normalization of cDNA microarray data. *Methods* 31: 265–273.

126. R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
127. Smyth GK, Michaud J, Scott H (2005) The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067–2075.
128. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Medical Biology* 3: 3.
129. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 57: 289–300.
130. Cooke BM, Lingelbach K, Bannister LH, Tilley L (2004) Protein trafficking in *Plasmodium falciparum*-infected red blood cells. *Trends Parasitol* 20: 581–589.
131. Delemarre BJ, van der Kaay HJ (1979) [Tropical malaria contracted the natural way in the Netherlands]. *Ned Tijdschr Geneesk* 123: 1981–1982.
132. Guinet F, Dvorak JA, Fujioka H, Keister DB, Muratova O, et al. (1996) A developmental defect in *Plasmodium falciparum* male gametogenesis. *J Cell Biol* 135: 269–278.
133. Thaithong S, Beale GH (1981) Resistance of ten Thai isolates of *Plasmodium falciparum* to chloroquine and pyrimethamine by in vitro tests. *Trans R Soc Trop Med Hyg* 75: 271–273.
134. Rosario V (1981) Cloning of naturally occurring mixed infections of malaria parasites. *Science* 212: 1037–1038.
135. Chambers JM, Cleveland WS, Kleiner B, Tukey PA (1983) *Graphical Methods for Data Analysis*. Wadsworth & Brooks/Cole.
136. Koenen M, Scherf A, Mercereau O, Langsley G, Sibilli L, et al. (1984) Human antisera detect a *Plasmodium falciparum* genomic clone encoding a nonapeptide repeat. *Nature* 311: 382–385.
137. Scherf A, Hilbich C, Sieg K, Mattei D, Mercereau-Puijalon O, et al. (1988) The 11-1 gene of *Plasmodium falciparum* codes for distinct fast evolving repeats. *EMBO J* 7: 1129–1137.
138. Kahane B, Sibilli L, Scherf A, Jaureguiberry G, Langsley G, et al. (1987) The polymorphic 11.1 locus of *Plasmodium falciparum*. *Mol Biochem Parasitol* 26: 77–85.