Plasmodium falciparum mature schizont transcriptome variation among clinical isolates and laboratory-adapted clones

Sarah J Tarr[1,*], Ofelia Díaz-Ingelmo[1], Lindsay B Stewart[1], Suzanne E Hocking[1], Lee Murray[1], Craig W Duffy[1], Thomas D Otto[2+], Lia Chappell[2], Julian C Rayner[2], Gordon A Awandare[3] and David J Conway[1,*]

[1] London School of Hygiene and Tropical Medicine, London, UK

[2] Wellcome Trust Sanger Institute, Hinxton, UK.

[3] West African Centre for Cell Biology of Infectious Pathogens, Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Ghana.

+Current Address: Centre of Immunobiology, Institute of Infection, Immunity & Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

[*] Corresponding authors
david.conway@lshtm.ac.uk
sarah.tarr@lshtm.ac.uk

Abstract

Malaria parasite genes can exhibit variation in sequence or transcriptional activity. A wealth of information is available on sequence polymorphism in clinical malaria isolates, but there are few quantitative whole-transcriptome studies of natural variation at specific developmental stages. It is challenging to obtain adequately precise and well-replicated transcriptome measurements in order to account for technical and biological variation among preparations, which can otherwise introduce noise or bias in analyses. We address the issue by obtaining of RNA-seq profiles of multiple independently cultured replicates of mature schizont-stage malaria parasites from a panel of clinical isolates and laboratory-adapted lines. With a goal of robustly identifying variably expressed genes, we show that increasing the biological sample replication improves the true positive discovery rate, and that six independent replicates of each isolate is significantly superior to lower numbers. Focusing on genes that are more highly expressed on average improves the discovery rate when fewer biological replicates are available. We identify genes encoding transcription factors and proteins implicated in gametocytogenesis that differ in expression between cultured clinical and laboratory adapted lines. We confirm the variable expression of known merozoite invasion ligands, and identify previously uncharacterised genes as highly differentially expressed among isolates. RT-qPCR assays confirm the variation, and extend quantitation of expression of these genes to a wider panel of *ex vivo* clinical isolate samples. This highlights new candidates for investigation as potential markers of alternative developmental pathways or targets of immunity.

Author summary

We analyse the transcriptomes of mature *Plasmodium falciparum* schizonts using RNA-sequencing. We use large numbers of replicates per sample to minimise the impact of inter-replicate biological variation on observed patterns of differential expression. We identify genes that are differentially expressed between isolates. We extensively validate our findings for novel putative targets of immunity. For a panel of *ex vivo* clinical isolates, we show that expression levels of these candidate genes in fall within the ranges observed for the cultured isolates. These genes constitute novel targets for characterisation for merozoite-stage intervention.

Introduction

Symptoms of malaria occur as the malaria parasite undergoes cycles of invasion and replication inside erythrocytes. Towards the end of each intra-erythrocytic cycle of cell division, 'schizont'-stage malaria parasites upregulate the expression of genes encoding proteins that contribute to daughter cell ('merozoite') egress and invasion of a new host cell. Despite the highly controlled transcriptional program throughout the parasite life-cycle [1], transcriptional variation exists between parasite clones [2, 3], and can also be induced by external cues [4-7]. In *P. falciparum*, variable expression of

members of the *var* gene family, which encode hypervariable *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) antigens, leads to extensive antigenic variation [8]. However, the extent to which gene expression variation contributes to the diversity of other antigens is not known.

Merozoite-stage antigens are also targets of acquired immunity [9] and antibodies against many merozoite antigens are correlated with protection from malaria [10]. Analysis of genome sequences from clinical *P. falciparum* isolates has shown that directional selection by acquired immune responses [11-14] has led to the evolution of multiple alleles of many merozoite antigens. Nonetheless, merozoite antigens do not exhibit sequence diversity to the extent of gene families such as var, rifin [15] or stevor [16], and therefore are appealing targets for vaccination. In addition to allelic diversity, many merozoite antigens exhibit variable expression patterns [11, 17-19]. While whole-genome sequencing studies of clinical and laboratory-adapted parasite isolates have provided a wealth of information contributing to our understanding of diversity among immune antigens and adaptive traits such as transmission, the extent of antigenic diversity cause by schizont-stage transcriptional plasticity has not been studied at the whole-transcriptome level. Antigenic diversity caused by variations in gene expression must be characterised so that we can understand features of parasite virulence such as adaptation and immune evasion.

Accurate quantitation of differential gene expression in whole-transcriptome analyses is affected by the level of biological replication within sample groups [20, 21]. In the absence of biological replication, it is impossible to account for variation that exists due to the stochasticity of gene expression [21]. The impact of sample replication on the detection of differential gene expression has been the focus of much research [22], and bioinformatics tools have been developed to determine replicate numbers appropriate to experimental designs [23, 24]. In transcriptomic studies of malaria parasites, life-cycle stage, overall synchronicity and gametocyte conversion are technically challenging to replicate [2, 25]. Continuous epigenetic switching of large numbers of genes also creates transcriptional diversity within genetically clonal parasite populations [3]. These features of parasite biology can confound the transcript profile of a given sample and contribute to apparent gene-wise expression variation among samples. Approaches to mitigate the effect of life-cycle length and staging have involved sampling isolate transcriptomes throughout the parasite life-cycle [2, 26]. However, no large-scale studies into malaria parasite transcriptomes have been undertaken that include independent biological replication to allow adjustment for gene-wise variation within samples.

Much of our understanding of malaria parasite biology is derived from parasite lines that have been adapted to *in vitro* culture for decades [27-30]. Genome sequencing studies have identified polymorphisms in transcriptional factors and cell signalling proteins [31, 32], as well as genetic

deletions and amplifications [33, 34] that are associated with adaptation of parasites to *in vitro* culture. While laboratory-adapted isolates are antigenically and transcriptionally diverse [3, 17, 26, 35, 36], it is not clear to what extent they reflect the diversity among *ex vivo* parasites. Therefore, *ex vivo* clinical isolates analysed in the first cycle of growth following isolation from an infected person are the most relevant parasites for the study of *in vivo* phenotypes such as growth rate and antigenic variation. *Ex vivo* isolates matured to the schizont stage show variations in expression of immunologically relevant genes encoding proteins with roles to merozoite invasion [11, 18, 19]. However, *ex vivo* isolates are not replicable, and are often not available in sufficient quantities for whole transcriptomic analysis. Furthermore, despite efforts to adapt *ex vivo* isolates to *in vitro* culture [37], *in vitro*-adapted cultured clinical lines may lose *in vivo* characteristics throughout the process of culture adaptation. Therefore, research to capture the extent of variation of the malaria parasite antigenic repertoire must compare multiple independent isolates, including laboratory-adapted isolates, and cultured and *ex vivo* clinical isolates, with a caveat of limited replicability for the latter.

Here we present gene expression profiles of schizont-stage malaria parasites from multiple cultured clinical isolates and laboratory-adapted lines. We conduct RNA-seq analysis with large numbers of replicates per sample, in order to minimise the impact of biological variation, and to generate highly resolved expression profiles of schizont-stage parasites. We show that increased replication within samples improves the true-positive discovery rate for identifying differentially expressed genes. We identify schizont-stage genes that are differentially expressed between laboratory-adapted and cultured clinical isolates, as well as genes variably expressed among cultured clinical isolates. We identify a set of putative merozoite antigens that show variable expression among cultured clinical isolates. We show that the expression levels of these genes in a panel of *ex vivo* clinical isolates are within the ranges of expression observed for laboratory-adapted and cultured clinical isolates. These data highlight the value of *in vitro* cultured parasite lines as proxy for *ex vivo* expression profiles.

Results

Generation of replicated, highly correlated schizont-stage transcriptomes from laboratory-adapted and cultured clinical isolates

Multiple replicates of schizont-stage parasites from four laboratory-adapted *P. falciparum* lines (3D7, Dd2, D10 and HB3) and six recently culture-adapted *P. falciparum* clinical isolates from Navrongo, Ghana (INV271, INV278, INV280, INV286, INV293 and INV296) were isolated by either MACS or discontinuous density centrifugation (Percoll) methods. Since our cultured clinical isolates could not be consistently or reproducibly synchronised, we incubated parasite preparations with the egress inhibitor, E64, to block parasites at the schizont stage, in order to improve the yield of schizont-stage parasites and improve the resolution of parasite transcriptomes at this stage. Transcriptomes of the

laboratory-adapted and cultured clinical isolates were generated by RNA sequencing (RNA-seq). RNA-seq analysis of four paired replicates of E64-treated and untreated *P. falciparum* 3D7 cultures showed that E64 treatment *per se* did not affect the transcriptomes; only a single gene was differentially expressed by $\log_2$ fold change > 2 (S1 File).

Gene expression differences between parasite isolates have previously been shown to be affected by genetic changes such as copy-number variations and polymorphisms that impact on the quantitation of transcripts when counted with respect to a reference strain genome such as 3D7 [2]. To minimize the impact of inter-isolate sequence differences on gene expression measures, we generated a highly curated gene annotation file for RNA-seq read mapping that excluded regions of putative or known polymorphism (S2 file). As such, only reads mapping to non-polymorphic portions of polymorphic genes were used in quantitative expression analysis. We did not remove genes for which deletions were known from any of our analyses, as this would have resulted in an entire loss of data for genes that are absent only in subsets of isolates.

Mapped RNA-seq reads for each gene for each replicate were converted to fragments per kilobase of transcript per million mapped reads (FPKM) expression level values. To assess the contribution of non-schizont-stage transcripts to each replicate, FPKMs for each replicate were correlated with FPKM values for seven time points of a published 48 hour *P. falciparum* lifecycle RNA-seq time-course ([38]; see S1 Table). Replicates were excluded that did not have a maximum Spearman's correlation with parasites at 40 or 48 hours post-invasion. After this filtering step, each laboratory-adapted isolate was represented by at least six independently prepared replicates, and each cultured clinical isolate was represented by at least three independently prepared replicates (summarised in Table 1).

Table 1. Numbers of replicates of *P. falciparum* schizont samples before and after curation for staging

Strain	Starting number of replicates	Purification method	Final number of replicates
3D7	10	Percoll	10
3D7	9	MACS	9
Dd2	7	MACS	6
D10	10	Percoll	10
HB3	7	Percoll	7
INV271	6	MACS	3
INV278	5	MACS	5
INV280	6	MACS	6
INV286	6	MACS	6
INV293	6	MACS	6
INV296	5	MACS	3

Within each sample, pairwise correlations of FPKMs for all replicates correlated with Spearman's $\rho > 0.7$ (S2 Table).

Sample replication improves detection of differentially expressed genes

Transcriptomic comparisons of sample groups containing in excess of 40 replicates per group have shown that the sensitivity for detecting differential expression of genes between groups increases with the number of replicates in each group [22]. We assessed the impact of increasing replication by comparing our 3D7 (prepared with Percoll) and D10 sample groups, which each contained ten replicates. Differential expression analysis comparing the full 10 replicate groups of 3D7 and D10 identified 123 differentially expressed genes (absolute $\log_2$ fold change > 2, an adjusted P-value < 0.01). This number may increase with additional replication. However, this level of replication is rarely practical, particularly when handling rare or sensitive isolates. To identify an optimal level of replication that balances transcript discovery with achievable replicate numbers, we assessed what proportion of the 123 genes from the 10 replicate dataset were captured through comparisons that used two, four, six and eight replicates within each group. The true-positive rate of genes detected for 100 comparisons (using randomly selected replicates from each sample group) increased with the number of replicates within each group; the median true-positive rates were 0.30, 0.50, 0.67 and 0.80, when two, four, six and eight replicates were included, respectively (Fig 1). While the true-positive rate of genes detected increased with the number of replicates, the false-positive rate remained low irrespective of the number of replicates included (Fig 1).

We also determined the true-positive rate for detecting differential expression among the most highly expressed schizont-stage genes. For each of the 100 comparisons, replicate FPKMs for each sample were averaged to give 'per-sample' FPKM expression values and the genes with non-zero FPKM values were ranked by their maximum expression level in either strain. For each comparison, we determined the differentially expressed genes that fell within the top quartile of most highly expressed genes, and compared these to the differentially expressed genes identified among the top quartile of genes in the full (two x 10 replicate) analysis. Through analysis of differential expression among genes in the top quartile of expression levels, an improved true-positive rate was observed for replicate numbers tested. A median true-positive rates of 0.45, 0.63, 0.78 and 0.86 were achieved for comparisons containing two, four, six and eight replicates, respectively. Based on these data, we advocate the use of six independent biological replicates. Fewer replicates can be tolerated when analyses are focussed on the most highly expressed genes.

Comparison of gene expression in cultured clinical and laboratory-adapted isolates.

Pressures experienced by malaria parasites *in vivo*, such as immune selection, nutrient availability and febrile episodes, select for parasites with particular transcriptomic characteristics [3, 17]. *In vitro*,

these pressures are minimised, therefore transcriptomic comparisons of laboratory-adapted and clinical isolates can explain phenotypic differences that occur through the process of culture adaption [2], such as the use of distinct invasion pathways [39]. We used our combined data for laboratory-adapted and cultured clinical strains to identify schizont-stage genes that change in expression through adaptation to *in vitro* culture. In an analysis comparing the transcriptomes of laboratory and cultured clinical isolates, allowing for differences in purification method, one hundred and thirty four genes were identified as being differentially expressed between the two groups (S1 Fig and S3 Table) across all fourteen chromosomes (Fig 2a), representing 2.6 % of genes within the analysis. Within the top quartile of gene expression values (genes with a $\log_2$ FPKM value of > 6.94), twenty genes were significantly differentially expressed between the laboratory-adapted and clinical isolate groups (Fig 2b and S3 Table), representing 1.6 % of genes in the top quartile. This indicates a slight over representation of differentially expressed genes in the lower 75 % of gene expression values (Fisher's exact test odds ratio 0.518, P-value 0.006), which may reflect an inflated number of differentially expressed genes among these genes, due to false positives among low-expressed genes. After exclusion of genes for which the differential expression was likely as a result of known genetic deletions, and others belonging to sub-telomeric multi-gene families (S3 Table), twelve genes were identified as showing strong signals of differential expression between laboratory and clinical isolates (Table 2).

Table 2. $\log_2$ fold changes for genes differentially expressed between laboratory-adapted and clinical isolates

Gene ID	Product Description	$\log_2$ Fold Change	Wald statistic	Adjusted P-value
PF3D7_0104300	ubiquitin carboxyl-terminal hydrolase 1, putative	-2.18	-6.13	1.77E-08
PF3D7_0410000	erythrocyte vesicle protein 1	-2.38	-10.62	1.77E-23
PF3D7_0420300	AP2 domain transcription factor, putative	-2.02	-7.19	3.70E-11
PF3D7_0422900	methyltransferase, putative	-2.50	-6.48	2.78E-09
PF3D7_0501300	skeleton-binding protein 1	-2.50	-4.98	4.99E-06
PF3D7_0522300	18S rRNA (guanine-N(7))-methyltransferase, putative	-2.15	-8.67	1.03E-15
PF3D7_0935700	Plasmodium exported protein, unknown function	-2.50	-4.06	2.15E-04
PF3D7_1030200	claudin-like apicomplexan microneme protein, putative	-2.00	-9.21	1.50E-17
PF3D7_1036300	duffy binding-like merozoite surface protein 2	3.02	5.32	9.81E-07
PF3D7_1302100	gamete antigen 27/25	2.02	3.99	2.76E-04
PF3D7_1327300	conserved Plasmodium protein, unknown function	-2.00	-4.97	5.16E-06
PF3D7_1371600	erythrocyte binding like protein 1, pseudogene	-2.17	-3.93	3.45E-04

With two exceptions, all genes were downregulated in cultured clinical isolate group, suggesting that their transcription is more relaxed in long-term culture adapted laboratory lines. This includes PF3D7_0104300, encoding ubiquitin carboxyl-terminal hydrolase 1 also known as ubiquitin-binding protein 1, which is involved in protein turnover and in which mutations are known to be associated with artemisinin resistance [40]. A single AP2 transcription factor, encoded by PF3D7_0420300, was downregulated among the cultured clinical isolates, as well as two predicted methyltransferases (PF3D7_0422900 and PF3D7_0522300). PF3D7_1036300, which encodes MSPDBL2, and PF3D7_1302100, which encodes gamete antigen 27/25, were the only two genes among this group that were upregulated in cultured clinical isolates. MSPDBL2 is a merozoite surface protein that is variably expressed among *P. falciparum* isolates [11], and was upregulated at the transcriptional level in schizont-stage parasites that were committed to gametocytogenesis [41], although a role for MSPDBL2 in gametocytogenesis remains to be confirmed. Gamete antigen 27/25 is a marker of early gametocytogenesis [42] that is induced upon removal of the repressive heterochromatin protein 1 (HP1) by gametocyte development protein 1 (GDV1) [43]. An additional gametocyte transcript (PF3D7_1327300, [44]) was down-regulated in the cultured clinical isolate samples. There was an over-representation of GDV1-regulated genes [43] among the 214 genes differentially expressed in cultured clinical isolate comparisons (Fisher's exact test odds ratio 12.55, P-value 1.871×10^{-09}).

The trend towards repression of expression in the cultured clinical isolate group prompted us to investigate whether there was an association between the genes differentially expressed between laboratory-adapted and culture clinical isolates, and targets of HP1 regulation [41, 45]. Of the 258 HP1-regulated genes compiled by Filarsky *et al* [41], 118 genes remain within our dataset (largely due to the removal of hypervariable var, rifin and stevor gene families, that are targets of HP1 regulation). Of the 134 genes differentially expressed between the two groups, 24 (17.9 %) were targets of HP1 regulation. Among twenty differentially expressed genes in the top quartile of expression values, two were targets of HP1 regulation (10 %). These values signify an overrepresentation of HP1-regulated genes among genes differentially expressed between the laboratory-adapted and cultured clinical isolate groups when considering all genes (Fisher's exact test odds ratio 11.29, P-value 1.32×10^{-5}), or when limited to genes among the top quartile of expression values (Fisher's exact test odds ratio 7.61, P-value 0.037).

The distinct expression levels of the genes differentially expressed between the laboratory-adapted and cultured clinical isolate groups (Table 2) is apparent through analysis of the normalised read counts for each gene (Fig 2b). Differences in expression were largely consistent across all isolates within each group (Fig 2b). This was despite low counts for PF3D7_1371600 (EBL-1) and PF3D7_0935700 in the HB3 strain and D10 strain, respectively, due to previously documented deletions [46, 47], that were nonetheless upregulated in laboratory isolates. Our comparisons between

laboratory-adapted and cultured clinical isolates were robust to the effects of gene copy-number variations (CNVs) on signals of differential expression between laboratory-adapted and cultured clinical isolates. CNVs can affect gene dosage, leading to apparent changes in gene expression between isolates [2]. The effects of genetic deletions on gene expression profiles were evident in comparative analyses of laboratory isolates. All isolates under study here have sustained CNVs across their genomes [33]. In pairwise comparisons among the four laboratory lines under study, 126 genes were differentially expressed (log_2 fold change >2 and adjusted P-value < 0.01) among those genes in the top quartile of schizont-stage gene expression (log_2 FPKM > 7.42; Fig 3a). A group of 20 genes showed exceptionally high log_2 fold changes (outliers, Fig 3b), up to a maximum of 14.42 (Fig 3b). These highly differentially expressed genes also contributed to the largest mean fold changes in gene expression among the laboratory-line comparisons (Fig 3c). Further investigation confirmed that differential expression of the majority of these genes may be explained by genomic differences between strains (S5 Table). For example, 13 genes in close proximity on chromosome 9 appear strongly down-regulated in the D10 strain (Fig 3d), in keeping with previously documented deletions in this region of chromosome 9 for this strain [33].

These data reiterate the importance of considering genetic differences between strains when making transcriptomic comparisons involving long-term laboratory adapted lines, and demonstrate how the use of multiple strains can minimise the impact of CNVs on studies of culture adaptation.

Differential expression among schizont-stage genes in cultured clinical isolates

To focus on differentially expressed genes among cultured clinical isolates without the impact of strong signals of differential expression due to genomic differences among laboratory strains, we analysed the replicated RNA-seq data for the cultured clinical samples in pairwise combinations. Among the fifteen pairwise comparisons, two hundred and fourteen genes showed an absolute log_2 fold change > 2 with an adjusted P-value < 0.01. Since the sample sizes for three of the six isolates contained fewer than six replicates each, these data may be an under-estimate of the true extent of differential expression. However, this was deemed preferable to inclusion of replicates with poor correlation to either the other replicates for that sample, or to overall schizont-stage transcriptomes. Differentially expressed genes among the top quartile of gene expression values were further analysed, in order to focus on the most highly expressed schizont-stage genes and to maximise the discovery rate given than some samples contained as few as three replicates. Replicate FPKMs were averaged to give 'per-sample' FPKM expression values. Of those genes in the top quartile of expression levels (log_2 FPKM > 6.34; Fig 4a), thirty-nine genes were differentially expressed (Fig 4b and S4 Table). The highest observed log_2 fold change was 13.51 (Fig 4c).

A feature of the genes differentially expressed among cultured clinical isolates were differences in expression of merozoite invasion ligands. The genes encoding the erythrocyte binding antigens,

EBA-140, EBA-175 and EBA-181, exhibited significant differential expression between isolates (S4 Table). These genes are among a panel of invasion ligand genes that were previously investigated for transcript-level variation in *ex vivo* clinical isolates [19] (S2 Fig). Members of the merozoite-surface protein 3 family also showed variation in expression levels (S4 Table) including PF3D7_1036300, which encodes the duffy binding-like merozoite surface protein MSPDBL2, which was also differentially expressed when broadly comparing laboratory and cultured clinical isolates, in keeping with the highly variable expression levels documented for this gene [11].

While fewer genes were differentially expressed among cultured clinical samples than laboratory-adapted isolates, there was significant overlap in the genes differentially expressed in comparisons among laboratory-adapted isolates and comparisons among cultured clinical isolates. Of the genes differentially expressed in comparisons among cultured clinical samples, 64 % were also differentially expressed in comparisons among laboratory-adapted lines (Fisher's exact test odds ratio 17.37, P-value $< 2.2 \times 10^{-16}$; S4 Table). This was also the case when only considering genes in the top quartile of expression values; 43.6 % of genes differentially expressed in comparisons among cultured clinical isolates were also differentially expressed in comparisons among laboratory-adapted isolates (Fisher's exact test odds ratio 7.94, P-value 2.4×10^{-08}). This indicates that signals of variable gene expression in cultured clinical isolates are also captured through analysis of laboratory-adapted lines.

To validate the data obtained through RNA-seq, we identified a refined subset of genes for quantitation by reverse-transcription quantitative PCR (RT-qPCR). Since expression variation may be a mechanism of immune evasion, we focussed further validation on a subset of the differentially expressed genes encoding proteins that are likely to enter the parasite secretory pathway (either by virtue of transmembrane domains or signal peptides), since relevant targets of immunity are located at the parasite surface, or secreted during the process of invasion. We did not further validate genes that had received previous functional characterisation, or that were members of sub-telomeric multi-gene families (S4 Table). In all, eight genes were identified for further characterisation (Table 3), which included three genes that were also differentially expressed among laboratory-adapted isolate comparisons (PF3D7_0423900, PF3D7_1252900 and PF3D7_1476500). Putative merozoite surface proteins encoded by PF3D7_0102700 (merozoite-associated tryptophan-rich antigen, MaTrA [48]) and PF3D7_0220000 (liver-stage antigen-3, LSA-3 [49]), were also among the subset of genes identified for further characterisation.

Table 3. Log_2 fold changes for genes in the top quartile of FPKM expression values, and with a log_2 fold change of > 2 in any comparison of six cultured clinical isolates. Denominators for each comparison are bold.

				Pairwise log_2 fold changes in expression for comparisons among schizont-stage cultured clinical isolates														
Gene ID	**Product Description**	**#TM[#]**	**SignalP Scores**	**INV 271 INV 278**	**INV 271 INV 280**	**INV 271 INV 286**	**INV 271 INV 293**	**INV 271 INV 296**	**INV 278 INV 280**	**INV 278 INV 286**	**INV 278 INV 293**	**INV 278 INV 296**	**INV 280 INV 286**	**INV 280 INV 293**	**INV 280 INV 296**	**INV 286 INV 293**	**INV 286 INV 296**	**INV 293 INV 296**	**Adjusted P-value**
PF3D7_0102700	merozoite-associated tryptophan-rich antigen	0	NN Sum: 3 NN D: .6 HMM Prob: .77	2.44	1.07	0.24	2.15	1.26	-1.37	-2.20	-0.28	-1.18	-0.83	1.09	0.19	1.92	1.02	-0.90	1.24E-04
PF3D7_0220000	liver stage antigen 3	2	null	1.53	1.98	2.13	2.79	1.16	0.45	0.61	1.26	-0.37	0.16	0.81	-0.82	0.65	-0.98	-1.63	8.45E-05
PF3D7_0423900*	probable protein, unknown function	1	null	-3.19	-0.56	0.08	-0.70	0.07	2.63	3.27	2.49	3.26	0.64	-0.14	0.64	-0.78	-0.01	0.78	1.07E-17
PF3D7_0605900	long chain polyunsaturated fatty acid elongation enzyme, putative	7	null	1.37	1.61	1.32	2.47	1.22	0.25	-0.04	1.11	-0.14	-0.29	0.86	-0.39	1.15	-0.10	-1.25	5.19E-03
PF3D7_0935300	phosphatidylinositol N-acetylglucosaminyltransferase subunit P, putative	2	NN Sum: 4 NN D: .6 HMM Prob: .03	-1.57	-1.52	-1.69	-2.58	-1.45	0.05	-0.12	-1.01	0.12	-0.17	-1.05	0.08	-0.88	0.24	1.13	2.06E-06
PF3D7_1252900*	Plasmodium exported protein, unknown function	0	NN Sum: 4 NN D: .58, HMM Prob: .39	-0.57	2.90	0.38	1.45	1.00	3.47	0.96	2.02	1.58	-2.52	-1.45	-1.89	1.07	0.62	-0.45	3.52E-05
PF3D7_1461700	conserved Plasmodium protein, unknown function	0	NN Sum: 3 NN D: .6 HMM Prob: .92	0.25	1.63	0.61	2.13	1.54	1.38	0.36	1.88	1.29	-1.02	0.50	-0.09	1.52	0.93	-0.59	5.47E-06
PF3D7_1476500*	probable protein, unknown function	1	NN Sum: 3 NN D: .43 HMM Prob: .15	1.04	1.52	0.95	3.06	2.19	0.48	-0.08	2.02	1.15	-0.56	1.54	0.67	2.10	1.24	-0.87	2.50E-10

*Also differentially expressed among comparisons of transcriptomes of laboratory-adapted schizonts

Number of predicted transmembrane domains

RT-qPCR assays were designed for absolute quantitation of these eight target genes. The transcripts of these eight genes were quantified by RT-qPCR for 58 of the 71 RNA preparations previously analysed by RNA-seq. Transcript copy numbers normalised to a house-keeping gene (HKG; PF3D7_0717700, serine-tRNA ligase) were correlated with HKG-normalised FPKM expression values for the same RNA preparations. With the exception of PF3D7_0935300, all genes showed a strong positive correlation between RNA-seq and RT-qPCR-derived expression measures (Fig 5). PF3D7_0935300 showed a weaker (r = 0.46) positive correlation.
Together these demonstrate the concordance of RNA-seq-derived and RT-qPCR-derived expression measures for these genes.

Expression of key variable genes in unreplicated *ex vivo* clinical isolates

Having robustly detected a panel of genes differentially expressed among cultured clinical isolates, we determined the expression levels of these genes in schizont-stage *ex vivo* clinical isolates (detailed in S6 Table). RNA-seq libraries were generated for nine *ex vivo* clinical isolates (INV018, INV020, INV027, INV032, INV051, INV054, INV055, INV056 and INV060) from Kintampo, Ghana, that had been matured *in vitro* until the schizont stage. Whole-transcriptome staging of FPKM values for these isolates identified two isolates (INV020 and INV032) that showed maximal correlation with a *P. falciparum* 3D7 time-course at earlier that 40 hours post-invasion, and were therefore not further analysed (Fig 6a). We also identified six *ex vivo* isolates for quantification by RT-qPCR. These isolates were selected based on them having at least 50 % schizonts in the culture, of which at least 50 % had more than six nuclei. RNA-sequenced isolates INV018 and INV060 were also quantified by RT-qPCR.
RNA-seq and RT-qPCR expression values for the eight gene panel (Table 3) were normalised against respective expression values for the house-keeping gene, PF3D7_0717700. The *ex vivo* samples were distributed within the range observed for the laboratory-adapted and cultured clinical parasite samples (Fig 6b). Without replication, the impact of biological variation the expression levels of these genes in *ex vivo* clinical isolates cannot be accounted for. However, the fact that the normalised expression values are distributed within the range observed for laboratory-adapted and cultured clinical lines indicates that these genes are unlikely to exhibit expression levels more extreme than the cultured lines. These data emphasise the value of laboratory-adapted and cultured clinical lines as models for *ex vivo* parasite isolates.

Discussion

Variations in gene expression contribute to the adaptation strategies of many organisms, and influence phenotypes such as sexual differentiation [50], adaptation to different growth conditions [51] and immune evasion [8, 52]. The malaria parasite is no exception [2, 3]. Transcriptome variations occur

due to various stimuli such as heat-shock [4, 5], hypoglycaemia [6] and oxidative stress [7]. Particular transcriptomic profiles are selected for by pressures such as febrile episodes [3] and altered erythrocyte invasion conditions [17, 53]. Transcriptomic differences between *in vitro*-adapted and field parasite isolates have identified differentially expressed genes with roles linked to parasite fitness [2]. Extensive whole-genome sequencing projects have shown that parasites respond to selective pressure through genetic sequence-level variations, which is particularly relevant for genes expressed at the schizont-stage. However, the extent to which transcriptional variation might lead to alterations in phenotypes such as growth adaptation and antigenic variation has not been studied. Here, we have used highly replicated whole-transcriptome analyses to study variations in gene expression between schizont-stage malaria parasite isolates.

Many transcriptomic analyses perform optimally with increased levels of sample replication to minimise the impact of stochastic differences in transcription [22, 54]. This is particularly important for studies that focus on a defined transcriptional window such as the schizont stage, in which noise is generated due to subtle stage-dependent differences between samples. However, the nature of *ex vivo* clinical samples of malaria parasites is such that replication is technically very difficult, and may be impossible. In order to assess the degree of replication necessary to robustly identify differentially expressed genes, we have used laboratory-adapted isolates samples with ten replicates to calculate the true- and false-discovery rates obtained using fewer replicates. In keeping with studies in *S. cerevisiae* [22], we have shown in *P. falciparum* that increasing the sample replication improves the true-positive discovery rate. We conclude that where possible, six independent replicates per isolate optimally balances gene discovery and experimental feasibility. When only fewer replicates were included in the analysis, we found that focussing on genes that are more highly expressed improves the discovery rate.

The highly variable nature of epigenetically-regulated genes in *P. falciparum* will add to the apparent gene-wise variation observed among cloned parasite samples. The cultured clinical lines under study here had not been previously cloned and it is likely that the isolates will contain multiple parasite genotypes [55]. Consequently, we expect the effect of clonally-variant gene expression on overall gene-wise variation to be minimised. It has been proposed that spontaneous transcriptional variation within genetically identical malaria parasite populations is a strategy that ensures fitness of parasites facing a range of potential and changing selective pressures [3]. It is likely that the rates that genes are up- or down-regulated by these mechanisms differ on a gene-by-gene basis. Therefore, bulk analysis of cloned lines will always be confounded by transcriptional heterogeneity. The recent adaption of single-cell transcriptomics methods to *P. falciparum* parasites will allow the definition of population-wide transcriptomic profiles [56, 57]. Application of these methods to *ex vivo* clinical isolates is essential to understand the diversity of antigen expression *in vivo*.

We addressed the effect of undescribed sequence polymorphisms on transcriptome analysis, by refining and masking 'non-mapping' transcript regions prior to transcriptome analysis, rather than removing polymorphic genes altogether. As earlier observed [19], our comparisons among isolates highlighted how gene copy-number variations impact the observed transcriptional differences among lines, highlighting the importance of, where possible, generating appropriate reference genomes for inter-isolate transcriptomic comparisons.

In a broad comparison of laboratory-adapted strains and cultured clinical isolates, we found that among the genes most strongly differentially expressed were genes involved in sexual differentiation, and potential transcription factors, which may play a role in transcriptional differences between long-term adapted laboratory isolates and recently cultured clinical lines. Of particular interest were genes encoding an AP2-transcription factor [58] and two methyltransferases, the roles of which remain to be investigated in *P. falciparum*. The over-representation of gametocyte-related genes differentially expressed between the two groups may reflect a lesser state of adaptation to culture within the cultured clinical isolates, and as such these isolates may commit to gametocytogenesis at a higher rate. This was also a feature of pan-lifecycle transcriptome comparisons of laboratory-adapted and field isolates [2]. Studies of gametocytogenesis in clinical isolates will provide a wealth of information on the process of commitment to sexual differentiation and transmission.

In general, differentially expressed genes were repressed in the cultured clinical isolates, suggesting that transcription is more relaxed in laboratory lines. HP1 is responsible for extensive gene silencing in malaria parasites [43]. We showed that there was an over representation of HP1-regulated genes among the genes differentially expressed between laboratory and cultured clinical isolate, although further studies are required to determine whether HP1 strictly determines gene expression in clinical isolates.

Through pairwise comparisons of individual cultured clinical isolates, we identified many genes that were differentially expressed among clinical isolates. Many of the genes identified have previously been shown to be differentially expressed through targeted qRT-PCR assays of schizont-stage *ex vivo* clinical isolate material [11, 18, 19]. In order to focus on transcriptional variation as a mechanism of immune evasion, we identified a panel of eight highly differentially expressed genes that encode secreted proteins, since these may be targets of immunity. In particular, PF3D7_0102700 (merozoite-associated tryptophan-rich antigen), PF3D7_0220000 (liver stage antigen 3) have both been identified as merozoite proteins [48, 49]. Another gene identified in out study, PF3D7_0423900, sits near the genes encoding cysteine-rich protective antigen (CyRPA) and reticulocyte binding homolog 5 (PfRh5) on chromosome 4. Differential expression of PF3D7_0423900 has not before been described. However, there is evidence for variable expression of an adjacent gene in field isolates [2], suggesting plasticity in expression around these loci, despite conserved expression of PfRh5 [59]. We

extensively validated our RNA-seq data by RT-qPCR for these genes and extended quantitation of expression of these genes to additional *ex vivo* clinical isolate samples. Despite the unreplicated nature of these samples, our data showed that expression measures for the genes tested fell within those of the replicated measures for other samples. Importantly this implies that cultured clinical isolates are a valuable proxy for studying *ex vivo* samples.

Together, these data highlight exciting new candidates for investigation as markers of alternative developmental pathways or targets of immunity.

Methods

Ethical approval and sampling of *P. falciparum* from clinical malaria cases.

Blood samples were collected from clinical malaria cases attending Ghana government health facilities between 2012 and 2013, in Kintampo (Brong-Ahafo Region of central Ghana), and Navrongo (Kassena-Nankana East Municipality, in the Upper East Region of northern Ghana). Ethical approval to collect and analyse the clinical samples was granted by the Ethics committees of the Ghana Health Service, the Noguchi Memorial Institute for Medical Research, University of Ghana, the Kintampo Health Research Centre, the Navrongo Health Research Centre and the London School of Hygiene and Tropical Medicine.

Patients were eligible to participate in the study if they had uncomplicated clinical malaria, were aged 2–14 years, tested positive for *P. falciparum* malaria by Rapid Diagnostic Test (First Response®, Transnational Technologies) or blood smear and had not taken antimalarial drugs during the 72 h preceding sample collection. Written informed consent was obtained from parents or legal guardians of all participating children, and additional assent was received from children over 10 years old. Antimalarial treatment and care was provided according to the Ghana Health Service guidelines. Venous blood samples (up to 5 ml) were collected into heparinized Vacutainer tubes (BD Biosciences). Blood samples were centrifuged, plasma and leukocyte buffy coats were removed, and erythrocytes were cryopreserved in glycerolyte and stored frozen at −80°C or in liquid nitrogen until shipment on dry ice to the London School of Hygiene and Tropical Medicine. A summary of clinical samples included in this study is shown in S6 Table.

Parasite culture and maintenance.

Parasites from thawed, *ex vivo* clinical samples were cultured at 2 % hematocrit in RPMI 1640 medium containing 2 % human AB serum (GE Healthcare) and 0.3 % Albumax II (Thermo Fisher Scientific) under an atmosphere of 5 % O_2, 5 % CO_2, and 90 % N_2 at 37 °C. Following thaw, parasites were matured for up to 48 h in until most parasites were at the schizont stage of parasite development, at which point RNA was extracted directly from the bulk cultures.

Laboratory-adapted and long-term cultured clinical isolates were cultured at 2 – 5 % haematocrit in RPMI 1640 medium containing 0.5 % Albumax II, at 37 °C. Laboratory-adapted isolates were cultured under atmospheric air with 5 % CO_2, and cultured clinical isolates were cultured in 5 % O_2, 5 % CO_2, and 90 % N_2. Schizonts were isolated from cultured lines by either MACS or discontinuous density centrifugation methods. For MACS purification of cultured clinical isolates, late stage parasites were isolated from 100 ml cultures with at least 0.7 % schizonts. For MACS purification of laboratory lines 3D7 and Dd2, 25 ml cultures with at least 1 % schizonts were used. Parasites were isolated by magnetic purification using magnetic LD Separation columns (Miltenyi Biotech). One column was used per 25 ml culture. Columns were washed twice in 3 ml of room temperature culture medium. Parasite culture was pelleted at 500 x g for 5 minutes. The pellet was resuspended in 3 ml culture medium per 1 ml of packed cell volume. The resuspended material was bound to the MACS column, which was then washed three times with 3 ml culture medium. Schizonts were eluted twice by removing the magnet from the column and forcing 2 ml culture medium through the column into a 15 ml falcon. Finally, the schizonts were pelleted at 500 g for 5 minutes and the pellet volume was estimated. 0.5 µl was harvested for giemsa-stained smear to assess staging, and 1 µl was added back to 250 µl culture at 0.8 % hematocrit to follow the progression. Remaining parasites were resuspended in 1.5 ml of culture medium with 10 µM E64 in a 12 well plate. Parasites were incubated for 5.5 hours in 5 % CO_2 at 37 °C, before pelleting in a 1.5 ml tube and proceeding with RNA extraction.

For discontinuous density centrifugation purification, parasites were maintained as 25 ml cultures at 2.5 % hematocrit. Cultures were used when they contained at least 1 % parasites with multiple nuclei. Schizonts were purified on a discontinuous density gradient. Specifically, 70 % Percoll (GE Healthcare)/2.93 % sorbitol/PBS was overlaid with 35 % Percoll/1.47 % sorbitol/PBS, which was in turn overlaid with resuspended pRBCs. Schizonts were separated by centrifugation at 2500 g for 10 minutes at 24 °C, with a light break. Purified schizonts were washed once in complete medium and the pellet volume was estimated. Six pellet volumes of 50 % haematocrit erythrocytes were added to the pellet and resuspended. The sample was smeared, and resuspended in 6 ml of complete culture medium. Of this, 1 ml was used as a control untreated sample to track parasite egress. E64 was added to the remaining 5 ml at a final conentraction of 10 µM. Control and E64-treated samples were placed at 37 °C, 5 % CO_2 static incubator for 5.5 hours. Schizonts from the E64-treated culture were overlaid on 70 % Percoll/2.93 % sorbitol/PBS and separated by centrifugation at 2500 g for 10 minutes at 24 °C, with a light break. The schizont layer was washed once in complete culture medium and final pellets yielding 10 - 20 µl packed material were used for RNA extraction.

RNA extraction.

Pellets were resuspended 500 µl in TRIzol® reagent (Thermo Fisher Scientific)as per manufacturer's instructions, and were stored at -80 °C until RNA extraction. RNA extraction from TRIzol® reagent was as per manufacturer's instructions. Final RNA pellets were resuspended in 100 µl RNase-free

H_2O. A second RNA clean-up and on-column DNase treatment was carried out using RNeasy mini columns (Qiagen) as per manufacturer's instructions. RNA was eluted in 30-50 µl RNase-free H_2O. RNA concentration was quantified by Qubit High Sensitivity RNA Assay (Thermo Fisher Scientific) as per manufacturer's instructions. For samples containing at least 500 ng RNA, RNA integrity was checked on an Agilent Bioanalyzer using RNA 6000 Nano reagents and chips (Agilent Genomics) as per manufacturer's instructions.

RNA-seq library preparation and sequencing.

Laboratory isolate and cultured clinical isolate RNA-seq libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina) using 500 ng – 1 µg RNA as per the Illumina TruSeq Stranded mRNA protocol for MiSeq sequencers. Libraries were validated on an Agilent Bioanalyzer using DNA 1000 reagents and chips (Agilent Genomics) to quantify library sizes and confirm the absence of primer dimers. Libraries were quantified using a KAPA Universal Library Quantification kit (Roche Diagnostics Limited) on a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific) and library concentrations were adjusted for library size. 12 – 15 pM pooled libraries were sequenced on a MiSeq System (Illumina) using a MiSeq Reagent Kit v3 (Illumina) with 2 x 75 cycles.

Ex vivo isolate RNA-seq libraries were prepared using a modified protocol. PolyA+ RNA (mRNA) was selected using magnetic oligo-d(T) beads. mRNA was reverse transcribed using Superscript III® (Thermo Fisher Scientific), primed using oligo-d(T) primers. dUTP was included during second-strand cDNA synthesis. The resulting double stranded cDNA was fragmented using a Covaris AFA sonicator. Sheared double stranded cDNA was dA-tailed, end repaired, and "PCR-free" barcoded sequencing adaptors (Bioo Scientific) [60] were ligated (NEB). Libraries were cleaned up twice, using solid phase reversible immobilisation beads, and eluted in EB buffer (Qiagen). Second strand cDNA was removed using uracil-specific excision reagent enzyme mix (NEB). Libraries were quantified by quantitative PCR prior to sequencing on an Illumina MiSeq sequencer.

Data handling.

Raw sequence reads were aligned to the *P. falciparum* 3D7 v3 genome and converted to 'bam' format using samtools [61]. Reads with MAPQ scores < 60 were removed. Reads were counted using the "summarizeOverlaps" feature of the GenomicAlignments package [62] in R, against a previously published *P. falciparum* genome annotation file that had been masked for regions of polymorphism (detailed in S2 File) to remove known regions of genomic polymorphism, highly polymorphic gene families and duplicated genes.

Data curation and analysis.

Fragments Per Kilobase of transcript per Million mapped reads (FPKMs) for our own data and that of Otto et al [38] were calculated using the 'fpkm' function of DESeq2 [63] in R. FPKMs for each replicate of our own samples were correlated using a Spearman's Rank correlation with FPKMs for each of the seven time points (0, 8, 16, 24, 32, 40 and 48 hours post-invasion; S1 Table). Replicates

with a correlation of Spearman $\rho > 0.7$ at the 40 or 48 hour time-points were included for further analysis. Pairwise correlation of FPKMs among remaining replicates of each sample confirmed that all pairs of replicates had a correlation of Spearman $\rho > 0.7$. Differential expression analysis was conducted using DESeq2 in R.

qRT-PCR.

150 – 500 ng total RNA was reverse transcribed using Superscript II® (Thermo Fisher Scientific) with 250 ng random hexamer per 20 µl reaction, as per manufacturer's instructions. Quantitative PCR (qPCR) was carried out using SYBR® Select Master Mix (Thermo Fisher Scientific) with 500 nM forward and reverse primers, in a Prism 7500 Fast qPCR machine (Thermo Fisher Scientific). For each gene, threshold-cycle values were quantified against a serially diluted genomic DNA (Dd2 strain) standard curve, run on the same plate. Cycling parameters were: 50°C for 2 minutes, 95°C for 2 minutes followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. All wells were run as 10 µl volumes in technical duplicate. PCR primer sequences are shown in S7 Table. qPCR copy numbers were normalised against copies of a house-keeping gene, PF3D7_0717700 [64].

Data open access

Laboratory-adapted and cultured clinical RNA-seq data are submitted for access via Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/), entry: GSE113718.

Ex vivo RNA-seq data are submitted for access via the European Nucleotide Accession Archive (https://www.ebi.ac.uk/ena), study: ERP103955.

Acknowledgements

We thank Alistair Miles and Antoine Claessens for extraction of coordinates of windows of polymorphism.

References

1. Bozdech, Z., M. Llinás, B.L. Pulliam, E.D. Wong, J. Zhu and J.L. DeRisi, *The transcriptome of the intraerythrocytic developmental cycle of P. falciparum.* PLoS Biology, 2003. **1**(1): p. e5.
2. Mackinnon, M.J., J. Li, S. Mok, M.M. Kortok, K. Marsh, P.R. Preiser and Z. Bozdech, *Comparative transcriptional and genomic analysis of Plasmodium falciparum field isolates.* PLoS Pathog, 2009. **5**(10): p. e1000644.
3. Rovira-Graells, N., A.P. Gupta, E. Planet, V.M. Crowley, S. Mok, L. Ribas de Pouplana, P.R. Preiser, Z. Bozdech and A. Cortes, *Transcriptional variation in the malaria parasite Plasmodium falciparum.* Genome Res, 2012. **22**(5): p. 925-38.
4. Fang, J. and T.F. McCutchan, *Malaria: Thermoregulation in a parasite's life cycle.* Nature, 2002. **418**(6899): p. 742-742.
5. Oakley, M.S.M., S. Kumar, V. Anantharaman, H. Zheng, B. Mahajan, J.D. Haynes, J.K. Moch, R. Fairhurst, T.F. McCutchan and L. Aravind, *Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic Plasmodium falciparum parasites.* Infection and Immunity, 2007. **75**(4): p. 2012-2025.
6. Fang, J., H. Zhou, D. Rathore, M. Sullivan, X.-Z. Su and T.F. McCutchan, *Ambient glucose concentration and gene expression in Plasmodium falciparum.* Molecular and Biochemical Parasitology, 2004. **133**(1): p. 125-129.

7. Akide-Ndunge, O., E. Tambini, G. Giribaldi, P. McMillan, S. Müller, P. Arese and F. Turrini, *Co-ordinated stage-dependent enhancement of Plasmodium falciparum antioxidant enzymes and heat shock protein expression in parasites growing in oxidatively stressed or G6PD-deficient red blood cells.* Malaria Journal, 2009. **8**(1): p. 113.
8. Scherf, A., R. Hernandez-Rivas, P. Buffet, E. Bottius, C. Benatar, B. Pouvelle, J. Gysin and M. Lanzer, *Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in Plasmodium falciparum.* The EMBO Journal, 1998. **17**(18): p. 5418-5426.
9. Beeson, J.G., D.R. Drew, M.J. Boyle, G. Feng, F.J. Fowkes and J.S. Richards, *Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria.* FEMS Microbiol Rev, 2016. **40**(3): p. 343-72.
10. Richards, J.S., T.U. Arumugam, L. Reiling, J. Healer, A.N. Hodder, F.J. Fowkes, N. Cross, C. Langer, S. Takeo, A.D. Uboldi, J.K. Thompson, P.R. Gilson, R.L. Coppel, P.M. Siba, C.L. King, M. Torii, C.E. Chitnis, D.L. Narum, I. Mueller, B.S. Crabb, A.F. Cowman, T. Tsuboi and J.G. Beeson, *Identification and prioritization of merozoite antigens as targets of protective human immunity to Plasmodium falciparum malaria for vaccine and biomarker development.* J Immunol, 2013. **191**(2): p. 795-809.
11. Amambua-Ngwa, A., K.K. Tetteh, M. Manske, N. Gomez-Escobar, L.B. Stewart, M.E. Deerhake, I.H. Cheeseman, C.I. Newbold, A.A. Holder, E. Knuepfer, O. Janha, M. Jallow, S. Campino, B. Macinnis, D.P. Kwiatkowski and D.J. Conway, *Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites.* PLoS Genet, 2012. **8**(11): p. e1002992.
12. Ochola, L.I., K.K. Tetteh, L.B. Stewart, V. Riitho, K. Marsh and D.J. Conway, *Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in Plasmodium falciparum.* Mol Biol Evol, 2010. **27**(10): p. 2344-51.
13. Tetteh, K.K., F.H. Osier, A. Salanti, G. Kamuyu, L. Drought, M. Failly, C. Martin, K. Marsh and D.J. Conway, *Analysis of antibodies to newly described Plasmodium falciparum merozoite antigens supports MSPDBL2 as a predicted target of naturally acquired immunity.* Infect Immun, 2013. **81**(10): p. 3835-42.
14. Mobegi, V.A., C.W. Duffy, A. Amambua-Ngwa, K.M. Loua, E. Laman, D.C. Nwakanma, B. MacInnis, H. Aspeling-Jones, L. Murray, T.G. Clark, D.P. Kwiatkowski and D.J. Conway, *Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity.* Mol Biol Evol, 2014. **31**(6): p. 1490-9.
15. Kyes, S.A., J.A. Rowe, N. Kriek and C.I. Newbold, *Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9333-8.
16. Niang, M., X. Yan Yam and P.R. Preiser, *The Plasmodium falciparum STEVOR multigene family mediates antigenic variation of the infected erythrocyte.* PLoS Pathogens, 2009. **5**(2): p. e1000307.
17. Stubbs, J., K.M. Simpson, T. Triglia, D. Plouffe, C.J. Tonkin, M.T. Duraisingh, A.G. Maier, E.A. Winzeler and A.F. Cowman, *Molecular mechanism for switching of P. falciparum invasion pathways into human erythrocytes.* Science, 2005. **309**(5739): p. 1384-7.
18. Gomez-Escobar, N., A. Amambua-Ngwa, M. Walther, J. Okebe, A. Ebonyi and D.J. Conway, *Erythrocyte invasion and merozoite ligand gene expression in severe and mild Plasmodium falciparum malaria.* J Infect Dis, 2010. **201**(3): p. 444-52.
19. Bowyer, P.W., L.B. Stewart, H. Aspeling-Jones, H.E. Mensah-Brown, A.D. Ahouidi, A. Amambua-Ngwa, G.A. Awandare and D.J. Conway, *Variation in Plasmodium falciparum erythrocyte invasion phenotypes and merozoite ligand gene expression across different populations in areas of malaria endemicity.* Infect Immun, 2015. **83**(6): p. 2575-82.
20. Hansen, K.D., Z. Wu, R.A. Irizarry and J.T. Leek, *Sequencing technology does not eliminate biological variability.* Nat Biotechnol, 2011. **29**(7): p. 572-3.

21. Elowitz, M.B., A.J. Levine, E.D. Siggia and P.S. Swain, *Stochastic gene expression in a single cell.* Science, 2002. **297**(5584): p. 1183-6.
22. Schurch, N.J., P. Schofield, M. Gierlinski, C. Cole, A. Sherstnev, V. Singh, N. Wrobel, K. Gharbi, G.G. Simpson, T. Owen-Hughes, M. Blaxter and G.J. Barton, *How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?* RNA, 2016. **22**(6): p. 839-51.
23. Seo, J., H. Gordish-Dressman and E.P. Hoffman, *An interactive power analysis tool for microarray hypothesis testing and generation.* Bioinformatics, 2006. **22**(7): p. 808-14.
24. Busby, M.A., C. Stewart, C.A. Miller, K.R. Grzeda and G.T. Marth, *Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression.* Bioinformatics, 2013. **29**(5): p. 656-7.
25. Baker, D.A., *Malaria gametocytogenesis.* Mol Biochem Parasitol, 2010. **172**(2): p. 57-65.
26. Llinas, M., Z. Bozdech, E.D. Wong, A.T. Adai and J.L. DeRisi, *Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains.* Nucleic Acids Res, 2006. **34**(4): p. 1166-73.
27. Bhasin, V.K. and W. Trager, *Gametocyte-forming and non-gametocyte-forming clones of Plasmodium falciparum.* Am J Trop Med Hyg, 1984. **33**(4): p. 534-7.
28. Culvenor, J.G., C.J. Langford, P.E. Crewther, R.B. Saint, R.L. Coppel, D.J. Kemp, R.F. Anders and G.V. Brown, *Plasmodium falciparum: identification and localization of a knob protein antigen expressed by a cDNA clone.* Exp Parasitol, 1987. **63**(1): p. 58-67.
29. Guinet, F., J.A. Dvorak, H. Fujioka, D.B. Keister, O. Muratova, D.C. Kaslow, M. Aikawa, A.B. Vaidya and T.E. Wellems, *A developmental defect in Plasmodium falciparum male gametogenesis.* J Cell Biol, 1996. **135**(1): p. 269-78.
30. Walliker, D., I.A. Quakyi, T.E. Wellems, T.F. McCutchan, A. Szarfman, W.T. London, L.M. Corcoran, T.R. Burkot and R. Carter, *Genetic analysis of the human malaria parasite Plasmodium falciparum.* Science, 1987. **236**(4809): p. 1661-6.
31. Claessens, A., M. Affara, S.A. Assefa, D.P. Kwiatkowski and D.J. Conway, *Culture adaptation of malaria parasites selects for convergent loss-of-function mutants.* Sci Rep, 2017. **7**: p. 41303.
32. Kafsack, B.F., N. Rovira-Graells, T.G. Clark, C. Bancells, V.M. Crowley, S.G. Campino, A.E. Williams, L.G. Drought, D.P. Kwiatkowski, D.A. Baker, A. Cortes and M. Llinas, *A transcriptional switch underlies commitment to sexual development in malaria parasites.* Nature, 2014. **507**(7491): p. 248-52.
33. Cheeseman, I.H., N. Gomez-Escobar, C.K. Carret, A. Ivens, L.B. Stewart, K.K. Tetteh and D.J. Conway, *Gene copy number variation throughout the Plasmodium falciparum genome.* BMC Genomics, 2009. **10**: p. 353.
34. Jeffares, D.C., A. Pain, A. Berry, A.V. Cox, J. Stalker, C.E. Ingle, A. Thomas, M.A. Quail, K. Siebenthall, A.C. Uhlemann, S. Kyes, S. Krishna, C. Newbold, E.T. Dermitzakis and M. Berriman, *Genome variation and evolution of the malaria parasite Plasmodium falciparum.* Nat Genet, 2007. **39**(1): p. 120-5.
35. Dolan, S.A., J.L. Proctor, D.W. Alling, Y. Okubo, T.E. Wellems and L.H. Miller, *Glycophorin B as an EBA-175 independent Plasmodium falciparum receptor of human erythrocytes.* Mol Biochem Parasitol, 1994. **64**(1): p. 55-63.
36. Gilberger, T.W., J.K. Thompson, T. Triglia, R.T. Good, M.T. Duraisingh and A.F. Cowman, *A novel erythrocyte binding antigen-175 paralogue from Plasmodium falciparum defines a new trypsin-resistant receptor on human erythrocytes.* J Biol Chem, 2003. **278**(16): p. 14480-6.
37. Ribacke, U., K. Moll, L. Albrecht, H. Ahmed Ismail, J. Normark, E. Flaberg, L. Szekely, K. Hultenby, K.E. Persson, T.G. Egwang and M. Wahlgren, *Improved in vitro culture of Plasmodium falciparum permits establishment of clinical isolates with preserved multiplication, invasion and rosetting phenotypes.* PLoS One, 2013. **8**(7): p. e69781.

38. Otto, T.D., D. Wilinski, S. Assefa, T.M. Keane, L.R. Sarry, U. Bohme, J. Lemieux, B. Barrell, A. Pain, M. Berriman, C. Newbold and M. Llinas, *New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq.* Mol Microbiol, 2010. **76**(1): p. 12-24.
39. Dolan, S.A., L.H. Miller and T.E. Wellems, *Evidence for a switching mechanism in the invasion of erythrocytes by Plasmodium falciparum.* J Clin Invest, 1990. **86**(2): p. 618-24.
40. Henriques, G., R.L. Hallett, K.B. Beshir, N.B. Gadalla, R.E. Johnson, R. Burrow, D.A. van Schalkwyk, P. Sawa, S.A. Omar, T.G. Clark, T. Bousema and C.J. Sutherland, *Directional selection at the pfmdr1, pfcrt, pfubp1, and pfap2mu loci of Plasmodium falciparum in Kenyan children treated with ACT.* J Infect Dis, 2014. **210**(12): p. 2001-8.
41. Filarsky, M., S.A. Fraschka, I. Niederwieser, N.M.B. Brancucci, E. Carrington, E. Carrio, S. Moes, P. Jenoe, R. Bartfai and T.S. Voss, *GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing.* Science, 2018. **359**(6381): p. 1259-1263.
42. Carter, R., P.M. Graves, A. Creasey, K. Byrne, D. Read, P. Alano and B. Fenton, *Plasmodium falciparum: an abundant stage-specific protein expressed during early gametocyte development.* Exp Parasitol, 1989. **69**(2): p. 140-9.
43. Brancucci, N.M.B., N.L. Bertschi, L. Zhu, I. Niederwieser, W.H. Chin, R. Wampfler, C. Freymond, M. Rottmann, I. Felger, Z. Bozdech and T.S. Voss, *Heterochromatin protein 1 secures survival and transmission of malaria parasites.* Cell Host Microbe, 2014. **16**(2): p. 165-176.
44. Lu, X.M., G. Batugedara, M. Lee, J. Prudhomme, E.M. Bunnik and K.G. Le Roch, *Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum.* Nucleic Acids Res, 2017. **45**(13): p. 7825-7840.
45. Fraschka, S.A., M. Filarsky, R. Hoo, I. Niederwieser, X.Y. Yam, N.M.B. Brancucci, F. Mohring, A.T. Mushunje, X. Huang, P.R. Christensen, F. Nosten, Z. Bozdech, B. Russell, R.W. Moon, M. Marti, P.R. Preiser, R. Bartfai and T.S. Voss, *Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites.* Cell Host Microbe, 2018. **23**(3): p. 407-420 e8.
46. Peterson, D.S. and T.E. Wellems, *EBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses.* Mol Biochem Parasitol, 2000. **105**(1): p. 105-13.
47. Nacer, A., E. Roux, S. Pomel, C. Scheidig-Benatar, H. Sakamoto, F. Lafont, A. Scherf and D. Mattei, *Clag9 is not essential for PfEMP1 surface expression in non-cytoadherent Plasmodium falciparum parasites with a chromosome 9 deletion.* PLoS One, 2011. **6**(12): p. e29039.
48. Ntumngia, F.B., M.K. Bouyou-Akotet, A.C. Uhlemann, B. Mordmuller, P.G. Kremsner and J.F. Kun, *Characterisation of a tryptophan-rich Plasmodium falciparum antigen associated with merozoites.* Mol Biochem Parasitol, 2004. **137**(2): p. 349-53.
49. Morita, M., E. Takashima, D. Ito, K. Miura, A. Thongkukiatkul, A. Diouf, R.M. Fairhurst, M. Diakite, C.A. Long, M. Torii and T. Tsuboi, *Immunoscreening of Plasmodium falciparum proteins expressed in a wheat germ cell-free system reveals a novel malaria vaccine candidate.* Sci Rep, 2017. **7**: p. 46086.
50. Mata, J., A. Wilbrey and J. Bahler, *Transcriptional regulatory network for sexual differentiation in fission yeast.* Genome Biol, 2007. **8**(10): p. R217.
51. Cases, I., V. de Lorenzo and C.A. Ouzounis, *Transcription regulation and environmental adaptation in bacteria.* Trends Microbiol, 2003. **11**(6): p. 248-53.
52. Cestari, I. and K. Stuart, *Transcriptional Regulation of Telomeric Expression Sites and Antigenic Variation in Trypanosomes.* Curr Genomics, 2018. **19**(2): p. 119-132.

53. Awandare, G.A., P.B. Nyarko, Y. Aniweh, R. Ayivor-Djanie and J.A. Stoute, *Plasmodium falciparum strains spontaneously switch invasion phenotype in suspension culture.* Sci Rep, 2018. **8**(1): p. 5782.
54. Robinson, M.D. and G.K. Smyth, *Moderated statistical tests for assessing differences in tag abundance.* Bioinformatics, 2007. **23**(21): p. 2881-7.
55. Agyeman-Budu, A., C. Brown, G. Adjei, M. Adams, D. Dosoo, D. Dery, M. Wilson, K.P. Asante, B. Greenwood and S. Owusu-Agyei, *Trends in multiplicity of Plasmodium falciparum infections among asymptomatic residents in the middle belt of Ghana.* Malar J, 2013. **12**: p. 22.
56. Poran, A., C. Notzel, O. Aly, N. Mencia-Trinchant, C.T. Harris, M.L. Guzman, D.C. Hassane, O. Elemento and B.F.C. Kafsack, *Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites.* Nature, 2017. **551**(7678): p. 95-99.
57. Reid, A.J., A.M. Talman, H.M. Bennett, A.R. Gomes, M.J. Sanders, C.J.R. Illingworth, O. Billker, M. Berriman and M.K. Lawniczak, *Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites.* Elife, 2018. **7**.
58. Balaji, S., M.M. Babu, L.M. Iyer and L. Aravind, *Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains.* Nucleic Acids Research, 2005. **33**(13): p. 3994-4006.
59. Crosnier, C., L.Y. Bustamante, S.J. Bartholdson, A.K. Bei, M. Theron, M. Uchikawa, S. Mboup, O. Ndir, D.P. Kwiatkowski, M.T. Duraisingh, J.C. Rayner and G.J. Wright, *Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum.* Nature, 2011. **480**(7378): p. 534-7.
60. Kozarewa, I., Z. Ning, M.A. Quail, M.J. Sanders, M. Berriman and D.J. Turner, *Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes.* Nat Methods, 2009. **6**(4): p. 291-5.
61. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
62. Lawrence, M., W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M.T. Morgan and V.J. Carey, *Software for computing and annotating genomic ranges.* PLoS Comput Biol, 2013. **9**(8): p. e1003118.
63. Love, M.I., W. Huber and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
64. Salanti, A., T. Staalsoe, T. Lavstsen, A.T. Jensen, M.P. Sowa, D.E. Arnot, L. Hviid and T.G. Theander, *Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria.* Mol Microbiol, 2003. **49**(1): p. 179-91.

Figure 1 Increased replication improves discovery of differentially expressed genes. Boxplot of the true-positive (1 and 2) and false-positive (3 and 4) rates for comparisons of 3D7 and D10 strains using two, four, six or eight replicates, compared to ten-replicate comparisons. 1 and 3 depict rates calculated for comparisons of 3D7 and D10 strains, with all differentially expressed genes considered. 2 and 4 depict rates calculated for differentially expressed genes among the top quartile of expression levels for comparisons of 3D7 and D10 strains.

Figure 2 Differential gene expression between schizonts from laboratory-adapted and clinical isolates of malaria parasites. a) Chromosome positions of genes differentially expressed in clinical isolates compared to laboratory lines. Differentially expressed genes among the top quartile of expression values are indicated in red. b) Plots of counts (normalised to library size) for twelve genes

differentially expressed between schizonts of laboratory-adapted and clinical isolate lines (Table 2), plotted by strain.

Figure 3. Differentially expressed genes among schizonts from laboratory-adapted malaria parasite lines. a. Plot of maximum $\log_2$ FPKM values in any sample, and mean absolute $\log_2$ fold changes in any comparison of four laboratory isolates, for genes within the top quartile of FPKM values. Genes that satisfy an absolute $\log_2$ fold change > 2 and adjusted P-value < 0.01 are highlighted in orange; b. Histogram of absolute $\log_2$ fold changes for genes within the top quartile of $\log_2$ FPKM values that satisfy an absolute $\log_2$ fold change >2 and adjusted P-value < 0.01. Boxplot and whiskers indicate the inter-quartile range of expression values, and 1.5 x inter-quartile range limits, respectively, and outliers are indicated by circles; c. Plot of maximum and mean absolute $\log_2$ fold changes (in any comparison of four laboratory isolates) for genes in the top quartile FPKM values. Outlier genes in b are highlighted in blue; d. Summed $\log_2$ FPKM expression values of thirteen potentially deleted genes on chromosome 9 (PF3D7_0935400 to PF3D7_0936800) for replicates of each laboratory isolate.

Figure 4. Differential gene expression schizonts from clinical isolates of malaria parasites. a. Histogram of fold changes (in any comparison of six clinical isolates) for genes in the top quartile of FPKM expression values. Genes that satisfy an absolute $\log_2$ fold change >2 and adjusted P-value < 0.01 are highlighted in orange; b. Plot of $\log_2$ maximum FPKM expression value for genes in the top quartile of expression values, against the mean absolute $\log_2$ fold change for 15 pairwise comparisons of six clinical isolates. Genes that satisfy an absolute $\log_2$ fold change >2 and adjusted P-value < 0.01 are highlighted in orange; c. Plot of maximum and mean absolute $\log_2$ fold changes in expression in comparisons of six clinical isolates, for genes within the top quartile $\log_2$ FPKM expression values. Genes that satisfy an absolute $\log_2$ fold change >2 and adjusted P-value < 0.01 are highlighted in orange.

Figure 5. High correlations between RNA-seq and RT-qPCR expression measures. Eight genes identified as differentially expressed among clinical isolates by RNA-seq were validated by RT-qPCR for 49 of the 71 RNA preparations under study (Table 1). Scatter plots for each gene show the ($\log_2$-transformed) gene of interest (GOI) copies normalised to house-keeping gene copies (HKG; PF3D7_0717700, serine-tRNA ligase), against ($\log_2$-transformed) GOI FPKM values normalised to HKG FPKM values. Spearman's correlation coefficients are shown for the untransformed correlations within each plot. For PF3D7_1461700, a single library had an FPKM value of 0. This point was not plotted but is included in the calculation of correlation.

Figure 6. Expression levels of differentially expressed genes for clinical, *ex vivo* samples are consistent with laboratory and cultured clinical material. a) Transcriptome correlations of nine *ex vivo* clinical samples matured to schizont stage. Two isolates (INV020 and INV032) showed peak correlations at earlier than 40 hours and were not analysed further. b) For the eight genes identified as differentially expressed among clinical isolates by RNA-seq, distributions of RT-qPCR-derived

copies (normalised to HKG) for 30 laboratory isolate (blue) and 19 cultured clinical isolate RNA preparations (green) from the 71 preparations in Table 1, were compared to expression values for schizont-matured *ex vivo* clinical isolates derived by either RT-qPCR (green) or RNA-seq (black). Gene expression values were normalised to HKG copies as per Fig 5.

S1 Figure. Distribution of $\log_2$ fold changes in gene expression comparison between laboratory-adapted and clinical isolates. Plot of maximum $\log_2$ FPKM expression value against the absolute $\log_2$ fold change in expression between laboratory-adapted and clinical isolate samples. Genes that satisfy an absolute $\log_2$ fold change >2 and adjusted P-value < 0.01 are highlighted in orange. Genes subject to regulation by HP1 are blue circles. Black solid line reflects the cumulative density of maximum $\log_2$ FPKM expression values. Dashed line reflects the top quartile of expression values ($\log_2$ FPKM 6.93).

S2 Figure. Differential expression of invasion genes among schizonts from laboratory-adapted and clinical isolates of malaria parasites. Plots of counts (normalised to library size) for eight genes in [19] for replicated laboratory-adapted and clinical isolate samples, plotted by strain.

S3 Figure. Gene expression levels for laboratory-adapted and clinical isolate samples for genes differentially expressed among clinical isolates. Plots of counts (normalised to library size) for eight genes identified as being differentially expressed among clinical isolates for replicated laboratory-adapted and clinical isolate samples, plotted by strain.

Fig 1

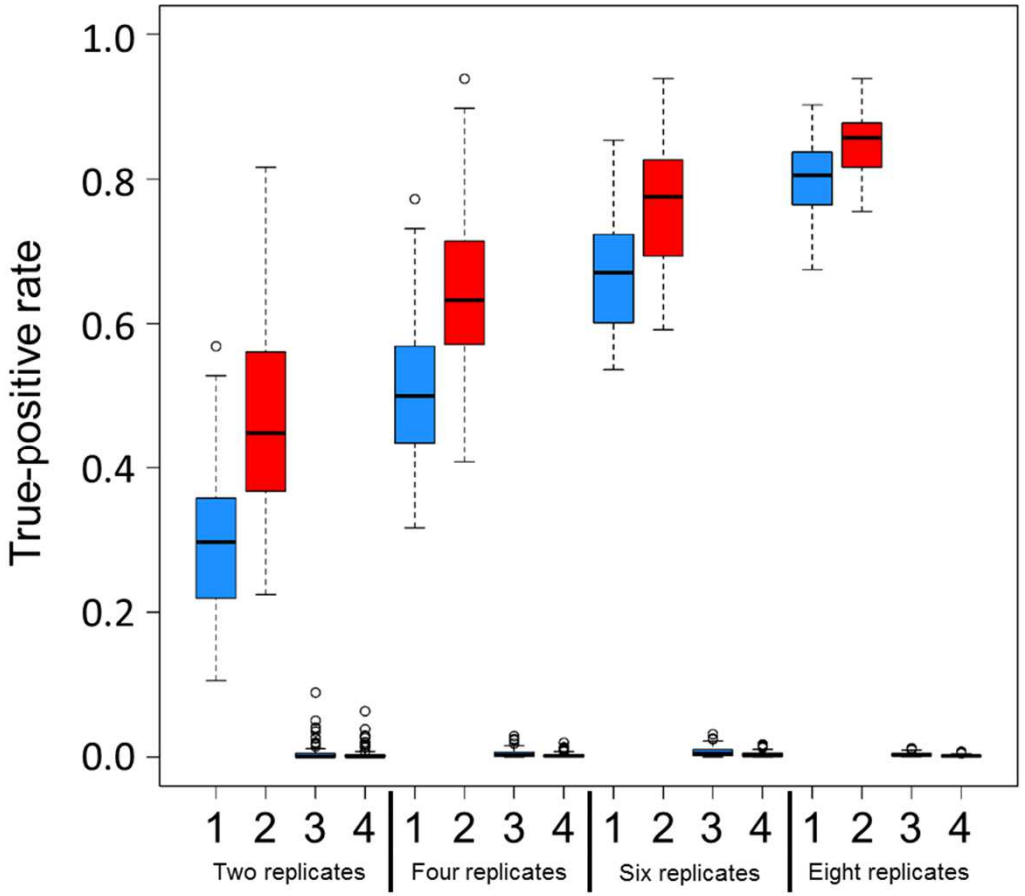


Fig 2

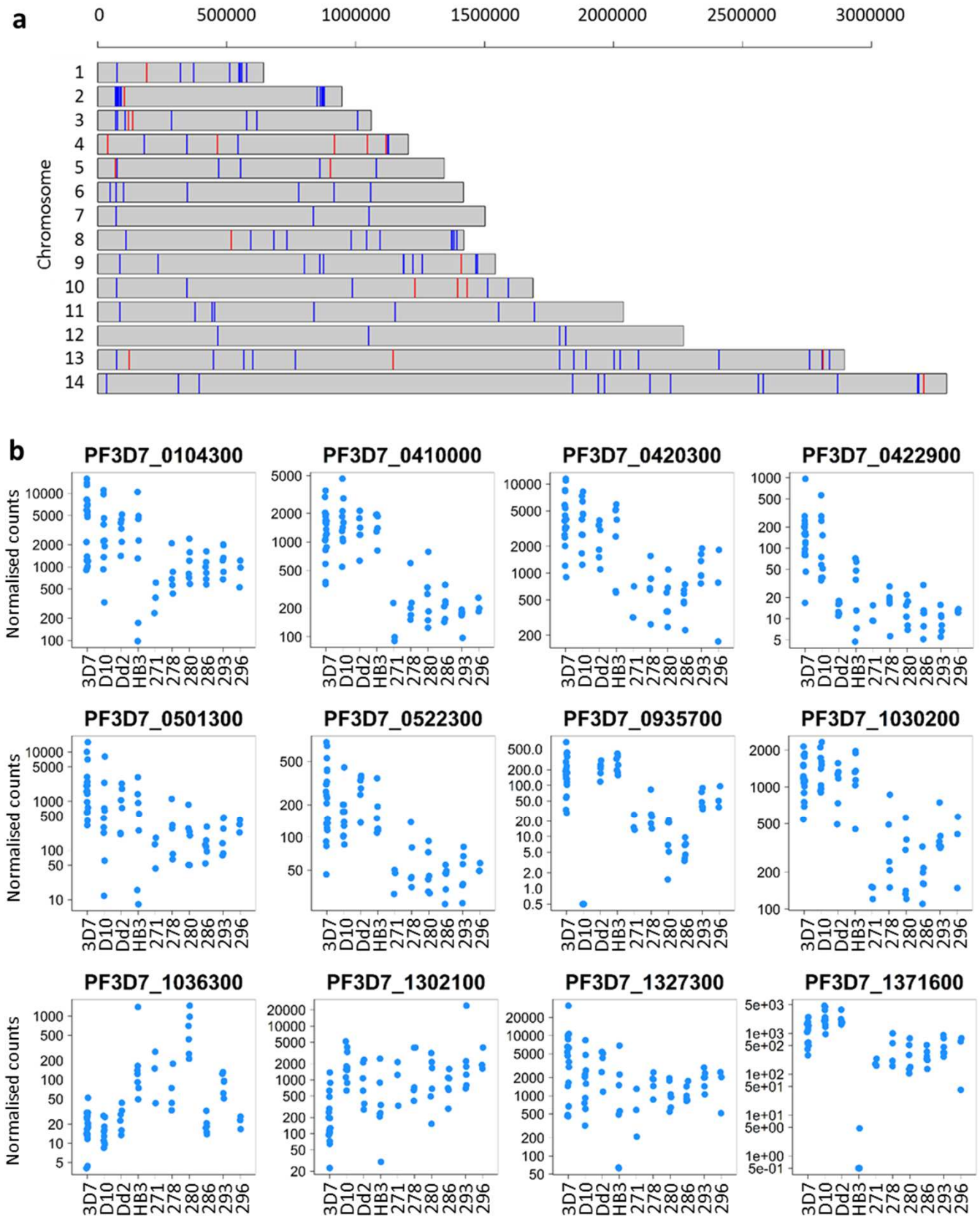


Fig 3

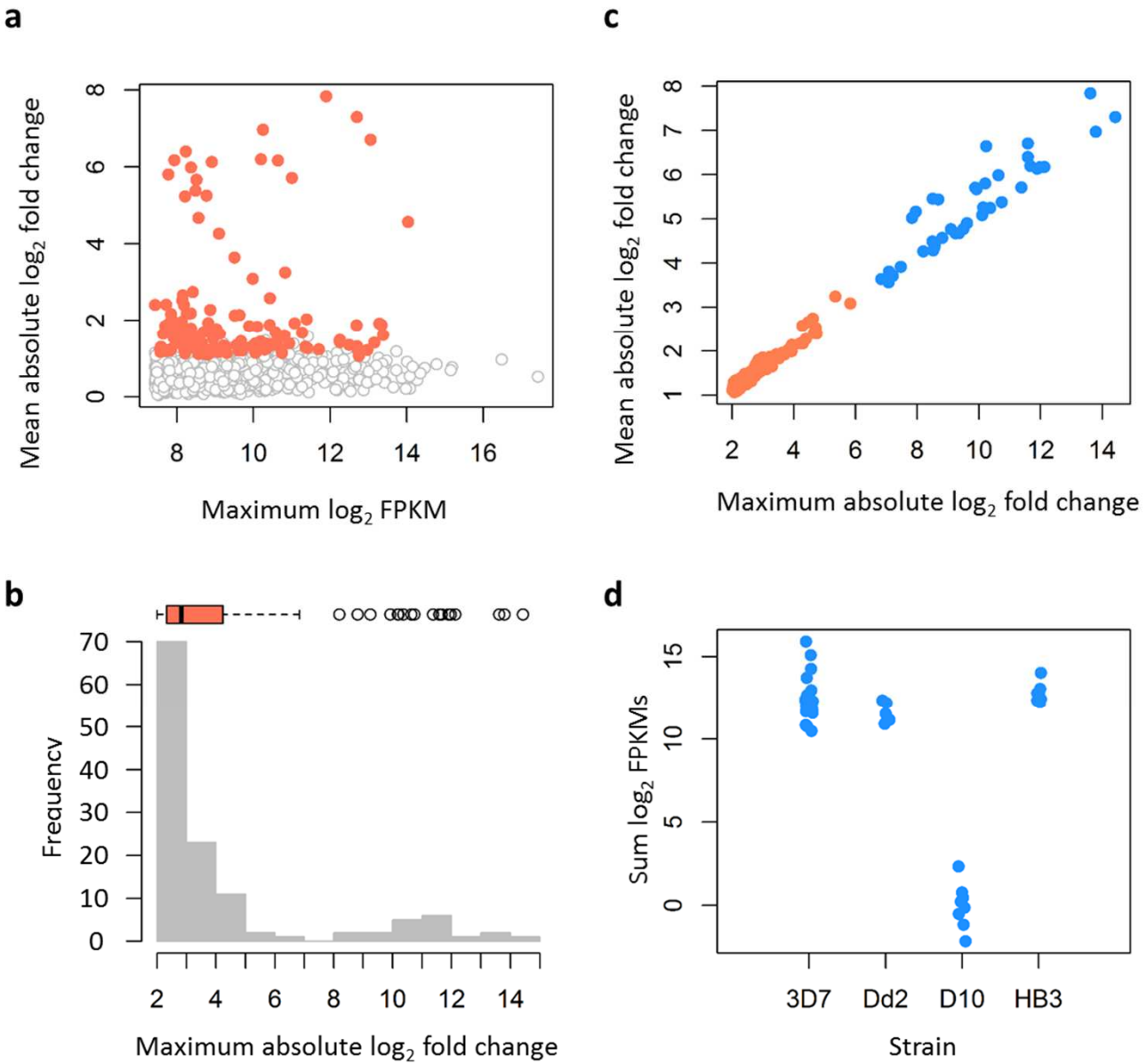
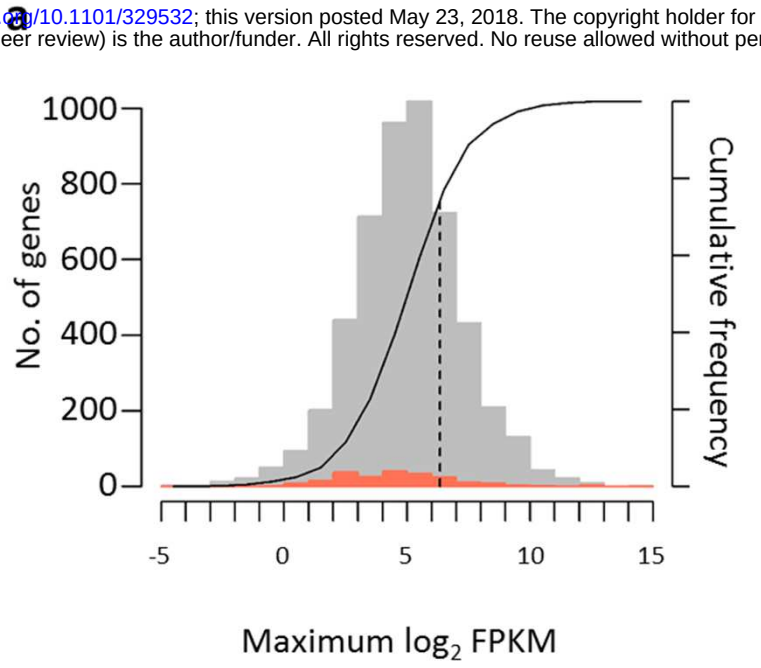
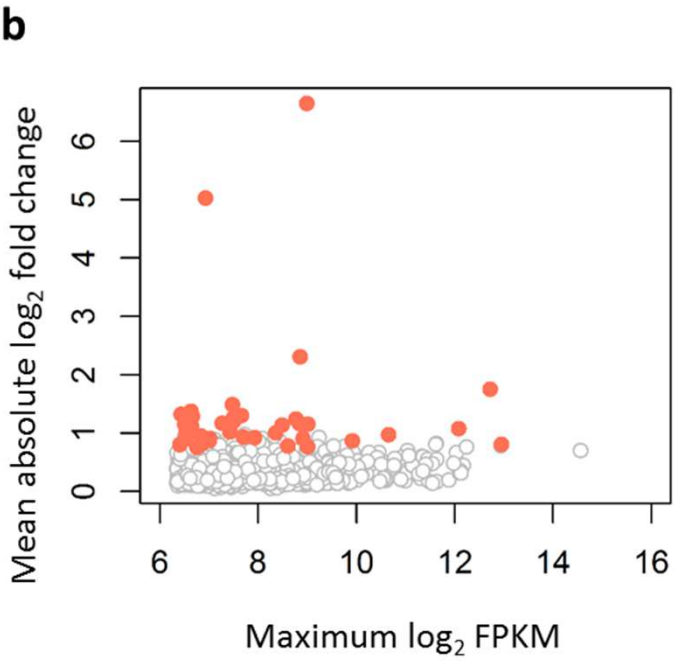


Fig 4





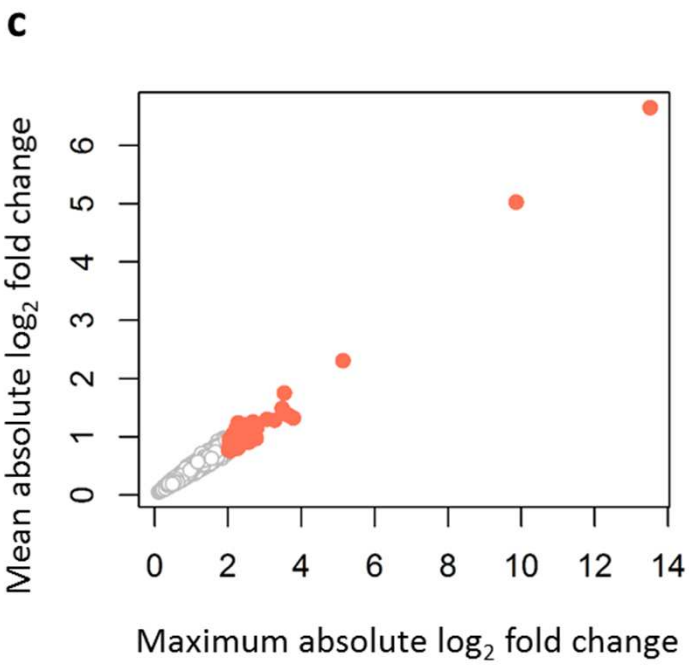


Fig 5

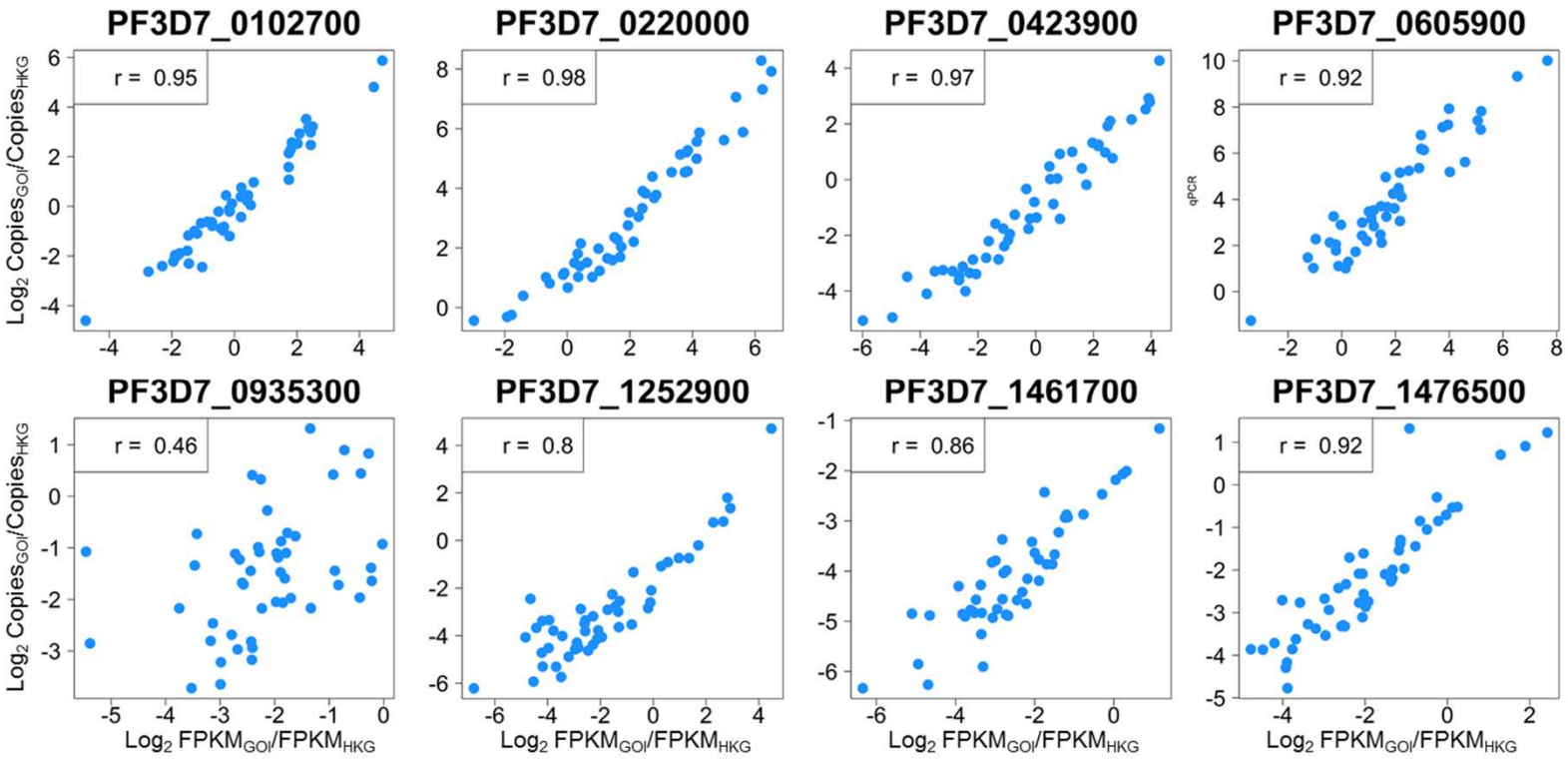


Fig 6