Color : Fig 1-9

SuppInfo : 1 table

Activation and clustering of a *Plasmodium falciparum var* gene are affected by subtelomeric sequences.

Michael F. Duffy, [1, 2]* Jingyi Tang, [1, 2]† Fransisca Sumardy, [1, 3]† Hanh H. T. Nguyen, [1, 2] Shamista A. Selvarajah, [1, 2] Gabrielle A. Josling, [1, 4] Karen P. Day, [2] Michaela Petter, [1, 2] Graham V. Brown, [5]

[1]Dept Medicine, Royal Melbourne Hospital, the University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Victoria, Australia

[2] The School of BioSciences, Bio21, The University of Melbourne, Victoria, Australia

[3]The Florey Institute of Neuroscience and Mental Health, 30 Royal Parade, Parkville, Australia

[4] Department of Biochemistry and Molecular Biology, W124 Millennium Science Complex, The Pennsylvania State University, State College, Pennsylvania 16802, USA

[5] The Nossal Institute for Global Health, The University of Melbourne, Victoria, Australia

*For correspondence. E-mail mduffy@unimelb.edu.au; Tel. 61383443262; †contributed equally

Running title: *cis* sequences affect *var* activity and location

keywords: *P. falciparum*, *var* genes, expression site, FISH, chromatin

Abbreviations:

PfEMP1: *P. falciparum* erythrocyte Membrane Protein 1

iRBC: infected red blood cell

H3: histone 3

K: lysine

ac: acetylation

me2, me3: di-, tri-methylation

TARE: telomere associated repeat element

Hi-C: Genome-wide chromosome conformation capture

FISH: fluorescence in-situ hybridisation

Sir3: silent information regulator 3

PfOrc1: *P. falciparum* origin recognition complex 1 protein

PfHP1: *P. falciparum* heterochromatin protein 1

PfRh4: *P. falciparum* reticulocyte binding protein homolog 4

ICAM-1: Intercellular Adhesion Molecule 1

CSA: chondroitin sulfate A

CD36: cluster determinant 36

ACS7: acyl-CoA synthetase 7

SNPs: single nucleotide polymorphisms

HA: hemagglutinin

ChIP: chromatin immunoprecipitation

Q-PCR: quantitative-PCR

hsp70: heat shock protein 70

RNA pol II CTD: RNA polymerase II C terminal domain

DAPI: 4',6-diamidino-2-phenylindole

Abstract

The *P. falciparum var* multigene family encodes the cytoadhesive, variant antigen PfEMP1. *P. falciparum* antigenic variation and cytoadhesion specificity are controlled by epigenetic switching between the single, or few, simultaneously expressed *var* genes. Most *var* genes are maintained in perinuclear clusters of heterochromatic telomeres. The active *var* gene(s) occupy a single, perinuclear *var* expression site. It is unresolved whether the *var* expression site forms *in situ* at a telomeric cluster or whether it is an extant compartment to which single chromosomes travel, thus controlling *var* switching. Here we show that transcription of a *var* gene did not require decreased co-localisation with clusters of telomeres, supporting *var* expression site formation *in situ.* However following recombination within adjacent subtelomeric sequences the same *var* gene was persistently activated and did co-localise less with telomeric clusters. Thus participation in stable, heterochromatic, telomere clusters and *var* switching are independent but are both affected by subtelomeric sequences. The *var* expression site co-localised with the euchromatic mark H3K27ac to a greater extent than it did with heterochromatic H3K9me3. H3K27ac was enriched within the active *var* gene promoter even when the *var* gene was transiently repressed in mature parasites and thus H3K27ac may contribute to *var* gene epigenetic memory.

Introduction

Each *P. falciparum* parasite possesses approximately 60 different *var* genes [1] that encode the immunodominant, variant surface antigen *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) [2-4]. At any one time only one [5], or a few [6], different PfEMP1s are expressed

at the surface of the infected red blood cell (iRBC) where they mediate adhesion to host receptors in the microvasculature [2, 5-7]. This allows parasites to sequester away from the circulation [8] and thus presumably avoid destruction in the spleen [9, 10]. The same *var* gene is transcribed in immature ring stage parasites over multiple generations, but parasites can switch to expression of different *var* genes altering host receptor specificity and causing clonal antigenic variation [11-13]. The tight regulation of *var* allelic exclusive expression and switching is believed to be critical for escaping immunity and maintaining chronic infection.

Expression of *var* genes is regulated by epigenetic mechanisms and the *var* promoter [5] or promoter and intron [14] are sufficient for participation in *var* allelic exclusive expression and switching. Silencing is associated with enrichment of the histone modifications H3K9me3 and H3K36me3 across the *var* coding sequence and promoter whilst activation is associated with enrichment of H3K9ac, H3K4me3 and the variant histones Pf H2A.Z and Pf H2B.Z at the promoter (reviewed in [15]). The active *var* gene is transiently repressed in the mature trophozoite and notably this corresponds with loss of the activation marks but enrichment in H3K4me2 at the promoter [16]. As the transiently repressed *var* gene is usually activated in the subsequent intraerythrocytic lifecycle, H3K4me2 has been proposed to confer epigenetic memory upon the transiently repressed *var* gene.

Var genes are grouped at subtelomeric and several chromosome internal sites within the genome [1]. Telomeric sequences tether *P. falciparum* telomeres to the nuclear periphery [17] in clusters [18] that are formed via interactions between telomere associated repeat elements (TAREs) [17] and uncharacterised proteins [19]. The sub-telomeric group A and B *var* genes are present in the perinuclear telomere clusters that hybridise to probes derived from telomeric repeats and the subtelomeric repeat sequences rep20 (TARE6) and TARE4 [5, 19-23]. Some *var* introns can also be directly recruited to the nuclear periphery [24]. Genome-wide chromosome conformation capture (Hi-C) [25, 26] and fluorescence in-situ hybridisation (FISH) [5] studies indicated that the chromosome internal group C *var* gene clusters are also associated with sub-telomeric *var* clusters. However another FISH study indicated that they form separate, poly-chromosomal clusters at the nuclear periphery [23]

Clustering of silent loci and the perinuclear tethering and clustering of telomeres [27-29] are widespread in eukaryotes. Recruitment to the periphery requires H3K9 methylation in *Caenorhabditis elegans* [30] and tethering involves a range of telomere and membrane associated proteins including Sir3 and Mps3 in yeast [31, 32]. *P. falciparum* sub-telomeric

regions are also enriched in methylated H3K9 [33] and PfOrc1 probably fulfils the tethering, and possibly cluster formation, functions of Sir3 [34, 35].

P. falciparum ring stage parasites have around nine transcription sites in their nucleus which are significantly enriched at the nuclear periphery in regions of lower chromatin density [36]. Multiple studies have shown that the expressed *var* gene occupies a transcriptionally permissive expression site at the nuclear periphery. These studies have used RNA FISH [22, 23, 37], DNA FISH [5, 19, 20, 22, 23, 37, 38], endogenous *var* loci [19, 20, 22, 23, 38] or plasmids or *trans* genes [5, 19 , 37, 38]. The *var* expression site has been reported to accommodate one [19, 20, 23, 37] or two [6, 22] expressed *var* genes simultaneously but can also simultaneously accommodate other non-*var,* active promoters [37, 38]. Thus active *var* promoters have a strong preference for a single perinuclear expression site that can accommodate multiple active *var* and non-*var* promoters.

For a *var* gene within a silent perinuclear cluster to be activated it could either leave the cluster and enter the active site as a solitary chromosome [20, 21, 23], or the active site could form *in situ* within the poly-chromosomal cluster of silent *var* loci [5, 19, 22]. These different observations were made using FISH with various probes. An orthogonal analysis using Hi-C data found no difference in the interchromosomal interactions between multiple *var* genes in their active or silent state, also suggesting that active *var* genes did not leave their poly-chromosomal clusters.

Aside from accommodating active *var* genes, the *var* perinuclear expression site is poorly defined. Upon activation *var* genes are associated with the periphery of a polar concentration of H3K4me3 [39] and are closer to the H3K4 methyltransferase PfSET10 and further from *P. falciparum* heterochromatin protein 1 (PfHP1) than when silent [40]. However the active *var* gene does not co-localise with nuclear pores [41], nor with subtelomeric long non-coding RNAs [42], nor with the euchromatic mark H3K79me3 [39].

Further insight into *P. falciparum* perinuclear expression sites was furnished by the study of *Pfrh4* [43]. This gene is also retained at perinuclear repressive, subtelomeric clusters when silent and when expressed it also co-localised with the same expressed episome used to define the *var* expression site [38]. Recruitment of *Pfrh4* to the expression site via expression of an adjacent, integrated trans-gene increased the frequency of *Pfrh4* activation but did not inevitably cause *Pfrh4* chromatin remodelling and promoter activation. Thus decreased

localisation of *Pfrh4* with telomere clusters was associated with its increased activation rate rather than activity *per se*.

Previously we showed that gene conversion from within the intron of the *var2csa* gene to the proximal telomere was associated with an increased *var2csa* on switch rate in the CS2 parasite line [44]. The recombination did not alter the *var2csa* promoter or first exon. The mechanism for the altered *var2csa* switch rate was unclear but we proved that it was not due to the recombined *var* intron sequence and thus concluded that sequences 3' of *var2csa* had affected the *var* gene switch rate. We isolated clones from the recombination event that had ceased to express *var2csa* and thus concluded that the recombination had increased the on switch rate but that *var2csa* could still be stably silenced by epigenetic mechanisms. Here we show that the recombination in *var2csa* appears to have caused an inescapable activation of *var2csa* and that this is associated with a decreased co-localisation of the *var2csa* locus with telomeric clusters. This work shows that subtelomeric sequence elements outside the *var* transcriptional unit affect its activation and its ability to participate in stable, silent, telomeric clusters. Whether these traits are functionally related is unknown but these findings do suggest an intriguing possible explanation for how *var* gene switch rates can differ independent of their promoters. We also show that the *var* expression site does not form a discrete concentration of *var*-associated euchromatic marks but does co-localise with H3K27ac to a greater extent than with H3K9me3. Furthermore we show that H3K27ac may play a role in *var* epigenetic memory.

Results

Recombined subtelomeric sequences cause inescapable activation of a cis var gene.

To study factors affecting *var* gene switching we analysed the spontaneous mutant CS2 parasites. CS2 parasites are a subclone of the E8B clone derived from the ItG strain. E8B parasites exhibit wild type *var* switching but CS2 activates *var2csa* at an atypically high rate leading to continual expression of the *var2csa* gene within a population of CS2 parasites [44]. We could select for CS2 parasites that had switched to expression of a different *var* gene by selecting CS2 infected red blood cells for adhesion to Intercellular Adhesion Molecule 1 (ICAM-1). Previously we analysed four clones of CS2 parasites that had been selected for adhesion to ICAM-1 and showed that three had reverted to adhesion to chondroitin sulfate A (CSA) and thus expression of *var2csa* (clones CS2isB10, CS2isC9 and CS2isD8) (fig 1). The

clone CS2isC8 however did not revert to expression of *var2csa* but instead maintained expression of a *var* gene(s) encoding adhesion to cluster determinant 36 (CD36) for 240 days [44]. Two other clones were recovered but not analysed at that time. One of these clones (CS2isB3) expressed solely *var2csa* by the time it was recovered, consistent with our previous report of an increased on-switch for *var2csa* (data not shown). The other clone (CS2isD4) however did not adhere to CSA but instead bound ICAM-1 and CD36 when recovered at 86 days post selection. Adhesion to ICAM-1 decreased after continuous culture for a further 100 days, indicative of switching to expression of *var* genes encoding adhesion to CD36 alone (fig 1). Thus, two of the six ICAM-1-selected CS2 clones that were recovered did not spontaneously revert to expression of *var2csa*.

The two clones that had switched to expression of *var* genes encoding adhesion to CD36 were analysed by Southern blot for the presence of the recombined *var2csa* locus. Both clones had undergone a further recombination event resulting in loss of *var2csa* (fig 2). A probe to the 3' end of the *var2csa* exon 1 hybridised to predicted *Btg*I and *Cla*I fragments in the CS2 parental clone, but not in the clonal progeny CS2isD4 and CS2isC8 (fig 2 A & B). A probe to the 3' end of acyl-CoA synthetase 7 (*ACS7*) (accession PF3D7_1200700), the adjacent gene towards the centromere, hybridised to all of the predicted restriction fragments from the CS2 parental clone but not the clonal progeny CS2isD4 and CS2isC8 (fig 2 A & C). The *ACS7* probe spanned an *Acl*I restriction site and in both clones hybridised to the 12,546 bp predicted fragment containing the 5' portion of *ACS7*. This indicated that the rearrangement that had resulted in loss of the *var2csa* gene occurred between the *Acl*I site in ACS7 and the *Acl*I site 5010 bp away towards *var2csa*. *ACS7* is a member of a partially conserved multigene family and the additional bands observed in CS2, CS2isD4 and CS2isC8 are presumably due to hybridisation with fragments from other members of this gene family. We concluded that the only CS2 sub-clones we obtained that could maintain silencing of *var2csa* had deleted the entire *var2csa* locus. That two of the six clones recovered from parasites that could be selected for adhesion to ICAM1 appeared to be siblings from the same rare deletion of an entire *var2csa* locus reinforced the very low frequency at which parasites not expressing *var2csa* were available in the CS2 population. We concluded that high frequency switching on of *var2csa* was an inescapable consequence of genetic difference between CS2 and the parental E8B, most probably due to the gene conversion that had replaced all sequences from the *var2csa* intron to the left hand telomere.

To identify other possible genetic causes of the CS2 parasite's persistent activation of *var2csa* we had previously used comparative genome hybridisation [44]. We confirmed those findings and extended them to detection of mutations predicted to cause loss of function and to identification of differentially expressed genes that could cause the activation of *var2csa*. Genetic variations between CS2 and the ECS and EIC phenotype selections of the E8B parental clone were analysed by sequencing cDNA. Genome wide association studies have repeatedly shown that variants associated with phenotypes are frequently outside of coding sequences, presumably in regulatory regions. We have previously observed that the high AT content and extreme low complexity of the *P. falciparum* non-coding sequences results in sequencing errors that don't impair mapping but could lead to many false positive variants. By sequencing cDNA we eliminated the need to interpret variants in non-coding sequence by directly determining whether genes were expressed, and thus whether regulatory sequences were functional. Although sequencing cDNA reduced our coverage of lowly expressed genes we also assumed that genes not expressed in ring stages were probably not important for *var* gene regulation. Reads mapped to 5327 of the 5699 annotated ItG strain genes in at least one of the samples. Pearson correlations of rlog transformed readcounts between any of two of EIC, ECS and CS2 were all greater than 0.98. Only three genes had log2 fold change greater than 2 in CS2, they were a rifin, a stevor pseudogene and a 28S rRNA and no genes had log2 fold change less than -1.8 in CS2 (fig 3, Table S1). We concluded that differences in the levels of expression of regulatory genes was not responsible for the differences in *var2csa* switching and co-localisation of the *var2csa* locus with telomere clusters observed in CS2.

For technical reasons identification of genetic variation was performed using the more complete 3D7 strain genome (see methods for details). Compared to the 3D7 genome CS2 parasites had 237 single nucleotide polymorphisms (SNPs), 78 insertions and 63 deletions that were absent from EIC and ECS and 33 of these were predicted to cause a loss of function. EIC had 521 SNPs, 59 insertions and 96 deletions that were absent from CS2 and ECS and 23 of these were predicted to cause loss of function. Unsurprisingly variant, multigene families contained many of the loss of function variants, as the 3D7 strain genome contains a different repertoire of these genes to the ItG strain, these reads will have mismapped, e.g. most of the 3D7 *var* genes identified as having loss of function variants had reads that mapped only to semi-conserved exon 2. These loss of function variants can therefore be disregarded. Inspection of all predicted loss of function variants revealed that all

bar one did not in fact affect function. The majority were incorrectly identified as insertions or deletions present in only CS2 or only EIC but were in fact due to clusters of SNPs that were present in both CS2 and EIC. The sole loss of function variant present in CS2 and absent from EIC and ECS was a frameshift deletion in the protein phosphatase gene *ppm1* (PF3D7_0410300, chromosome 4, position 473391, TC to T). Thus no clear candidate for mediating the altered switching of *var2csa* was identified, although a role for PPM1 cannot be excluded.

Recombined subtelomeric sequences altered the cis telomere's ability to participate in perinuclear, telomeric clusters.

If the recombination in *var2csa* caused an inescapable increase in activation at the locus in CS2 parasites we reasoned that there may be a detectable difference in nuclear localisation of the active *var2csa* locus between CS2 and wild type, parental E8B parasites. Nuclear mobility and residency of the *var* expression site have been linked to allelic exclusive expression of *var* genes and subtelomeric sequences are required for polychromosomal cluster formation and can exert repression on *var2csa* promoter activity [44]. To this end we repeatedly selected CS2 and E8B (ECS) parasites for adhesion to CSA until they both expressed primarily *var2csa* (fig 4A). We also selected parental E8B (EIC) parasites for adhesion to ICAM-1 and confirmed that they did not express *var2csa* (fig 4A). We then used these parasites for DNA FISH with probes to *var2csa* and to the subtelomeric repeat rep20 (TARE6). Rep20 has been repeatedly used to identify subtelomeric clusters in DNA FISH experiments [20].

DNA FISH for each parasite line was performed in biological triplicate. In total at least 66 fields were sampled for each parasite line and 854 CS2 nuclei, 957 ECS nuclei and 491 EIC nuclei were counted. The proportion of counted nuclei in which the *var2csa* FISH signal co-localised at all with a rep20 signal was determined for each field. The data were all normally distributed (Dagostino and Pearson normality test α=0.05). Significantly fewer of the CS2 nuclei (mean 0.4188 sd 0.191) had *var2csa* loci that co-localised with the rep20 probe than did the ECS nuclei that expressed *var2csa* (mean 0.5275 sd 0.1703) or the EIC that did not express *var2csa* (mean 0.5366 sd 0.2086) (unpaired t test $p \leq 0.0005$) (fig 5). However, there was no difference between ECS and EIC in the proportion of nuclei in which *var2csa* co-localised with rep20. We concluded that in CS2 parasites the recombined sequences from

var2csa towards the telomere caused an increased frequency of escape of the left hand chromosome 12 subtelomeric region from perinuclear telomeric clusters compared to the ECS parasites that also exclusively expressed *var2csa.* However, the mobility of the *var2csa* locus was independent of *var2csa* transcription *per se* as the wildtype parental clone lines EIC and ECS had similar, higher rates of co-localisation with rep20. Thus our DNA FISH data supported the model in which the *var* expression site can be accommodated within a poly-chromosomal sub-telomeric cluster [5, 19, 22].

The var gene expression site does not clearly associate with markers of euchromatin but is overlapped more by H3K27ac than by heterochromatic H3K9me3.

The composition of the *var* gene expression site remains largely undefined. We attempted to determine whether the *var* gene expression site was detectable as a gross sub-nuclear structure that associated with euchromatic histone modifications but not heterochromatic H3K9me3. We hypothesised that if a histone or modification was associated with the *var* expression site then it might be closer to, or have more overlap with, the active *var2csa* than did the heterochromatic mark H3K9me3. To this end we analysed 3D7 parasites expressing *var2csa* and hemagglutinin (HA) tagged PfH2B.Z by IFA; using antibodies to HA, PfH2A.Z and various histone modifications; combined with RNA FISH; using *in-vitro* transcribed *var2csa* anti-sense RNA probes. We combined IFA with RNA-FISH rather than DNA-FISH because the harsher conditions required for denaturation and hybridisation in DNA FISH degrade epitopes [45, 46] and because RNA FISH has been used previously in the analysis of *var* gene expression sites [22, 23, 37, 40]. All experiments were conducted in biological duplicate and analysed by confocal microscopy. As expected, the RNA FISH (fig 6A) generated a more diffuse signal than DNA FISH (fig 5). However in mammals, FISH signals from mRNAs are detectably concentrated at the active locus despite the presence of nuclear FISH signal due to mRNA diffusion [47, 48]. Therefore we employed two widely used measures to ascertain whether a histone or its modification associated with the *var* expression site: 1) the distance between peak fluorescence of the histone or modification and the *var2csa* RNA (as used previously in *P. falciparum* [40]) (examples in fig 6A, fourth column); 2) the overlap between the fluorescent signal from the histone or modification and the *var2csa* RNA [49].

The distance between peak fluorescence of the *var2csa* RNA and any of the histones or histone modifications tested were not significantly different (Kruskal Wallis test) (fig 7A). This suggests that none of these activation-associated histones or modifications were

particularly associated with, or restricted to the putative, active *var* gene expression site. This result is consistent with existing chromatin immunoprecipitation (ChIP) genomic datasets for H3K9ac, K3K4me3, PfH2A.Z, and PfH2B.Z, which all have a widespread distribution throughout the genome and thus, although they might be expected to be present at the *var* gene expression site, they would also be expected to be spread throughout the nucleus. Indeed only H3K9me3 showed the predicted restricted distribution at perinuclear foci.

Comparing Manders M2 coefficients for all histone modifications, less *var2csa* RNA co-localised with H3K9me3 than with H3K27ac (Kruskal Wallis test p<0.0001, Dunns multiple comparison test of all others to H3K9me3 p=0.0072) (fig 7C). This is consistent with the *var* expression site being in, or closer to, a region of the nucleus that is enriched in euchromatin rather than heterochromatin. Comparing Manders M1 coefficients for all histone modifications, less H3K27ac than H3K9me3 colocalised with *var2csa* RNA (Kruskal Wallis test p<0.0001, Dunns multiple comparison test of all others to H3K9me3 p=0.0125) (fig 7B). This is consistent with the observable, broader nuclear distribution of H3K27ac than H3K9me3 (fig 6B), therefore less H3K27ac overlaps with the *var* expression site.

H3K27ac marks active var promoters

In other eukaryotes H3K27ac is found at active promoters [50] and both H3K27ac [50-52] and H3K4me1 [50, 51, 53-56] are associated with enhancers of expressed genes. H3K27ac has been identified by mass spectrometry of blood stage *P. falciparum [57, 58]*. To determine whether H3K27ac and H3K4me1 were enriched at active *var* genes we performed ChIP using the *P. falciparum* 3D7 clone and its progeny 3C and 3D7-icam. The 3D7 clone is allogenic to E8B, but like E8B does not express detectable levels of *var2csa.* 3D7 parasites were selected for adhesion to CSA to generate 3C parasites that express high levels of *var2csa,* similar to ECS. 3D7 parasites were also separately selected for adhesion to ICAM1 to generate 3D7-icam parasites that expressed a range of genes including *PFL0020w* (fig 4B). 3D7, 3C and 3D7-icam are established in our laboratory as a model for studying changes in chromatin associated with *var* gene expression [59, 60]. Precipitated DNA was quantitated by quantitative-PCR (Q-PCR) and expressed as levels of (H3 modification ChIP relative to non-immune ChIP)/(H3 ChIP relative to non-immune ChIP) at the locus of interest relative to the same ChIP ratios at the hsp70 coding sequence. This approach enables biological replicates to be aggregated for statistical comparison [60] (see methods).

H3K27ac was enriched at the promoter (-1500, p=0.0023, -1000 p=0.0101) and intron (p=0.0174) of *var2csa* in 3C compared to the promoter (-1000) in 3D7 parasites (fig 8A) (Friedman test p<0.0001, Dunn's multiple comparison test, comparing 3D7 -1000 to all others). There was no significant enrichment of H3K27ac in *PFL0020w* in 3C or 3D7 when compared to the *PFL0020w* promoter (-450) in 3D7 parasites, although a trend to enrichment was apparent at the *PFL0020w* intron (p=0.0528) in 3C parasites (Friedman test p=0.0002, Dunn's multiple comparison test, comparing 3D7 -450 to all others).

H3K4me1 was marginally enriched in the *var2csa* promoter (-1000 p=0.0305, -575 p=0.0127) and start of the coding sequence (+75 p=0.0402) in 3C parasites compared to the promoter (-1000) in 3D7 parasites (fig 8A) (Friedman test p=0.0006, Dunn's multiple comparison test, comparing 3D7 -1000 to all others) but there was no enrichment of H3K4me1 at any position within *PFL0020w* in 3C compared to the *PFL0020w* promoter (-450) in 3D7 (Friedman test p=0.0137, Dunn's multiple comparison, comparing 3D7 -450 to all others).

H3K27ac levels were more variable across *PFL0020w* in 3D7-icam with much less enrichment in the 5'UTR at position -450 in rings compared to enrichment in the 5' UTR of *var2csa* in rings of 3C parasites (fig 8A). This is probably because multiple *var* genes encode ICAM1 adhesion leading to the mixed *var* gene expression profile of 3D7-icam parasites (fig 4B). This is unlike adhesion to CSA which is encoded solely by *var2csa* in 3C parasites (fig 4B). However greater enrichment of H3K27ac was still detected in 3D7-icam than in 3C and 3D7 combined when comparing the aggregated values for all positions in the 5' UTR of *PFL0020w* in either rings or schizonts (both p<0.0001 Mann Whitney test) (fig 8B). In contrast there was no significant difference between 3D7-icam and combined 3C and 3D7 in the levels of H3K27ac when the aggregated values of both positions in the coding sequence of *PFL0020w* were compared in either rings or schizonts (fig 8B).

Neither H3K27ac nor H3K4me1 enrichment differed between ring and schizont stages at any single position within either *var2csa* or *PFL0020w* in either 3C, 3D7 or 3D7-ICAM parasites (t-tests, two stage linear step up procedure of Benjamini, Krieger and Yeukuteli, all q>0.176).

The levels of H3K27ac were also assessed in non-*var* genes. Moderate enrichment upstream of genes was apparent in both rings and schizonts for genes expressed predominantly in rings (*hsp70, etramp10.1, sbp1, rex*) or at similar levels in rings and schizonts (*eba175 casein kinase 1*) (fig 9). However, for genes that were predominantly expressed in schizonts (*actin 1,*

msp2) H3K27ac enrichment was far more apparent upstream in the schizont stage but was not apparent in the ring stage (fig 9). H3K27ac was not enriched in either rings or schizonts upstream of genes that were not expressed during the intra-erythrocytic lifecycle (*csp, trap, slarp*).

The patterns of H3K27ac enrichment were consistent with constitutive enrichment of H3K27ac at the promoters of active *var* and other ring-stage expressed genes throughout the intraerythrocytic lifecycle, but with dynamic, expression associated enrichment at the promoters of genes expressed in schizont stages.

Discussion

We used RNAseq to identify differences in gene expression or genetic variants beyond the chromosome 12 recombination that may have caused the persistent activation of *var2csa* in CS2 parasites. Differences in the levels of the expression of other genes did not explain the persistent activation of *var2csa* in CS2. Impairment of the methyltransferase PfSET2 (PF3D7_1322100) was previously implicated in persistent activation of *var2csa* [61]. However in another study disruption of PfSET2 led to general de-repression of many *var* genes [62]. This was despite the prolonged *in vitro* cultivation post transfection required to recover PfSET2 mutants, which would have provided ample time for *var2csa*-specific effects of PfSET2 disruption to be detected. More recently, inhibition of other methyl-transferases also specifically activated *var2csa* at the chromosome 12 locus, but not *var2csa* transposed to another locus [63]. This reinforced the importance of sequences near *var2csa* on the chromosome 12 loci but outside of the *var2csa* gene and promoter for establishing proper chromatin mediated regulation of *var2csa.* We found that PfSET2 had several short deletions and several single nucleotide polymorphisms compared to the 3D7 sequence but these were conserved between CS2, EIC and ECS and were not predicted to affect PfSET2 function. No other genes have been implicated in the mechanism of impaired *var2csa* switching. PfSET2 binds dephosphorylated RNA polymerase II C terminal domain (RNA pol II CTD) which has been proposed to recruit PfSET2 to chromatin where it regulates *var* gene switching [61]. In other eukaryotes multiple protein phosphatases are able to dephosphorylate RNA pol II CTD and some affect the transcription of restricted groups of genes [64]. The protein phosphatase gene *ppm1* was the only gene in CS2 with a predicted loss of function that was not in EIC or ECS. *P. falciparum* has 29 protein phosphatases, PPM1 is not an ortholog of any known

RNA pol II CTD phosphatases and its deletion causes no detectable phenotype in asexual *P. berghei*, but it is essential for exflagellation of gametes [65]. Thus there is no evidence that that PPM1 dephosphorylates RNA pol II CTD but we cannot exclude that it plays a role in SET2 recruitment to *var2csa*.

Previously it was shown that loss of an entire subtelomeric region, including the *var* genes, prevented the affected telomere from joining a telomere cluster [17]. We have now shown that altering the composition of apparently complete subtelomeric sequences was associated with reduced ability of the affected sub-telomeric sequences to remain in telomere clusters and also with inescapable activation of *var2csa* independent of the *var2csa* promoter. A decrease in the participation in telomeric clusters was, however, not necessary for *var2csa* transcription in parental E8B parasites. These findings demonstrate that leaving a telomeric cluster is associated with, but not necessary for, *var* gene expression. The independence of *var* gene activation from leaving a telomeric cluster is also consistent with the observation that the release of the *Pfrh4* gene from telomeric clusters was associated with an increased on-switch rate but did not necessarily lead to transcription [43].

In mammals nascent coding and non-coding RNA from a single member of a gene family can nucleate nuclear bodies *in situ* by recruiting and concentrating *trans* factors required for specific transcription of the gene family [66]. The RNA pol II containing transcription factories then do not transcribe mRNA from all genes equally but preferentially transcribe the subset of genes that are regulated by the enriched specific transcription factors [67]. Thus, a gene only needs to escape silencing to create a nuclear body wherein related genes are transcribed. Our data support a similar model for *P. falciparum* wherein *var* genes are expressed at a singular gene expression site which does not exclude other silent *var* genes in *trans*. These findings are consistent with several other FISH studies which indicated that the *var* gene expression site forms *in situ* within a poly-chromosomal cluster [5, 19, 22] and Hi-C data that found no difference in the interchromosomal interactions between the *var* genes in their active or silent state [26],

The slight but significant increased tendency of *var2csa* to escape telomeric clusters in CS2 compared to EIC does not of itself explain the profound difference in *var2csa* transcription and may be of little functional consequence. However, if the altered subtelomeric sequences in CS2 were less effective at maintaining silent heterochromatin this might lead to increased transcription of *var2csa* and nucleation of the expression site *in situ* as well as loss of

heterochromatin associated *trans* factors required to maintain the *var2csa* locus within a telomeric cluster. This model is also consistent with the report of *Pfrh4* escape from a telomeric cluster independent of its transcription because the escape resulted from activation of a neighbouring *trans* gene, thus *PfRh4* could be maintained as silent heterochromatin within the expression site [43]. *Var2csa* is integrated into the program of *var* gene allelic exclusive expression and the *var* expression site has been largely defined by studies of the behaviour of *var2csa*. Thus this model is relevant to current understanding of the *var* gene expression site but it is possible that other *var* genes also employ other mechanisms of regulation.

The *var* gene expression site had greater overlap with the euchromatic histone modification H3K27ac than with heterochromatic H3K9me3. This is consistent with the *var* expression site being located within, or close to, the previously described perinuclear regions of euchromatin [20, 68]. The limitations of IFA-RNA-FISH and conventional confocal microscopy resolution may have hindered detection of *var2csa* RNA co-localisation with the other euchromatin-associated marks. Overall we concluded that a discrete nuclear *var* expression site could not be defined by a consistent euchromatic composition. This is in keeping with the widespread distribution of the euchromatic histone modifications across the genome which might prevent detection of a punctate depot of perinuclear euchromatin at the *var* expression site

The constitutive enrichment of H3K27ac upstream of inducible, ring stage-expressed genes, including active *var* genes, suggests that H3K27ac contributes to the demarcation of inducible euchromatin. In this context H3K27ac may play a role in *var* epigenetic memory through marking the erstwhile active *var* promoter when it is transiently repressed in schizonts. A role in *var* gene memory was previously proposed for H3K4me2 on the basis of its dynamic enrichment at transiently repressed *var* promoters in schizonts [16]. This pattern differs from the other euchromatic marks H3K9ac, H3K4me3, Pf H2A.Z and Pf H2B.Z which are dynamically enriched at the *var* promoter only when it is actively transcribed in ring stage parasites, and thus cannot transmit its activation state to mitotic progeny [16, 59, 60].

The *var* intron plays a role in regulating *var* gene transcription and was recently shown to interact with a protein shared with the promoter, and to transcribe antisense RNA associated with gene activation, which are both features suggestive of an enhancer/repressor like function [69, 70]. H3K27ac was elevated at the *var2csa* intron in 3C parasites and PfH2A.Z

and PfH2B.Z are also both elevated at *var* gene introns [59, 60]. This is intriguing as H2A.Z [50, 53, 54] and H3K27ac are enriched at enhancers in other eukaryotes and is suggestive of a role for H3K27ac in differentially marking regulatory elements outside the canonical promoter in *P. falciparum*.

The reason for the difference between constitutive H3K7ac enrichment at ring stage promoters and dynamic H3K27ac enrichment at schizont stage promoters is unclear. Perhaps it relates to schizont stage genes having a greater dynamic range of expression than ring stage genes, as indicated by the far higher levels of total gene expression observed in schizonts than in rings [71]. In other species H3K27 can also be methylated and H3K27me3 recruits the polycomb repressor complex in higher eukaryotes however neither H3K27me3 nor polycomb repressor complex has been reported in *P. falciparum*. *P. falciparum* H3K27 can also be formylated [72] but no functional association has been reported for this modification.

In conclusion we show that release from a telomeric cluster was not required for *var2csa* expression but was associated with the previously demonstrated increase in *var2csa* on-switch rate [44]. We propose that the altered subtelomeric sequences in CS2 had decreased ability to maintain the chromatin structure required for both maintenance of *var2csa* silencing and stable cluster formation. Furthermore, the set of histone modifications known to be enriched at the active *var* promoter can now be extended to include H3K27ac which is a candidate for mediating epigenetic memory of *var* activation across mitotic generations.

Experimental Procedures

Parasite culture

P. falciparum in vitro culture, adhesion phenotype selection and cloning was performed as described previously [44].

Southern blotting

Southern blotting of *P. falciparum* genomic DNA was performed as described previously [44]. Southern blots were hybridised with probes at 55 °C overnight and washed three times for 30 minutes with 0.5xSSC 0.1% SDS at 60 °C.

RNAseq, variant calling and differential gene expression analysis

CS2, ECS, EIC and 3D7 RNAseq libraries were prepared using NEBNext Ultra RNA library prep kits with 12 cycles of amplification using kappa hifi polymerase. Libraries were 126 bp paired end sequenced using an Illumina 2500 Hiseq as described previously [73]. Reads were

filtered for bases with a Phred Q>20 and reads >20 bp. After filtering there was 1.648 gigabases in 9,482,548 EIC reads, 1.344 gigabases in 11,657,290 CS2 reads and 0.422 gigabases in 3,474,526 EIC reads.

The filtered reads were analysed separately using the Broad Institutes recommended pipeline available at (https://software.broadinstitute.org/gatk/guide/article?id=3891). Briefly, STAR [74] was used for two pass mapping of filtered reads to the 3D7 version 12 genome from plasmoDB (plasmodb.org) . 3D7 was used because downstream variant annotation using snpEff [75] required a pre-built database that was not available for the CS2 clone's ItG parental strain. Duplicates were marked and readgroups added using Picard tools (http://picard.sourceforge.net). The GATK program was used to call variants [76]. Briefly, SplitNCigarReads was used to remove Ns, hard clip sequences overhanging introns and reassign MAPQ values of 255 from STAR output to 60 for subsequent GATK analysis. GATK Base recalibration was performed and then variants were called with the GATK HaplotypeCaller using the 3D7_v12 genome from plasmoDB and a Phred scaled confidence threshold of 20. GATK VariantFiltration was performed using the GATK recommended values for RNAseq of FS>30 and QD<2. Variants were then additionally filtered for a minimum depth (DP) of 10 filtered reads supporting an allele using vcf tools [77] and variants that were detected in CS2 but absent from EIC, and vice versa, were identified using vcftools isec. Variants were annotated including for predicted loss of function using snpEff [75] with the Pf3D7v91 annotations. Variants were filtered for predicted loss of function using snpSift [78]. For differential gene expression analysis the quality filtered reads were mapped by STAR to the ItG genome IT version 28 available from plasmoDB. Reads mapped to genes were then counted by featureCount [79]. Read counts were rlog transformed by DESeq2 [80] prior to analysis of correlation between samples. Raw readcounts were then analysed by DESeq2 for differential Gene Expression. 3D7 reads were mapped using Tophat2 [81] and fpkm determined using Cufflinks [82].

Q-RT-PCR

RNA extraction, gDNA digestion, reverse transcription and Quantitative (Q) RT-PCR using $2^{-\Delta\Delta Ct}$ analysis were performed as described previously [44]. Q-RT-PCR was performed using a previously published set of ItG *var* primers [83], skeleton binding protein 1 (accession PF3D7_0501300) as a normalising gene [84] and E8B gDNA as a calibrator.

DNA FISH

Indirect DNA FISH was performed using a modification of an existing method [85]. Roche Hi Prime biotin and digoxigenin labelling kits were used to random prime label 1 µg of purified dsDNA (rep20 or *var2csa*) as per the manufacturer's method. The labelled probes were precipitated with 0.1 volume 4 M LiCl and 3 volumes of cold 100% ethanol and washed with 70% ethanol, dissolved in 0.1 ml water and stored at -80°C. The *var2csa* probe template was a 2320 bp fragment of exon 1 spanning nucleotides 1133 to 3452. The rep20 probe template was a 509 bp sequence containing tandem repeats of rep20 that was initially amplified from 3D7 gDNA and used to demonstrate that rep20 recruited plasmids to telomeric clusters [86].

P. falciparum were suspended in nine cell pellet volumes of ice cold 0.15% saponin in PBS pH7.4 and incubated on ice for 5 min, then free parasites were pelleted at 3300 g for 10 minutes at 4°C and the cell pellet washed twice at 2500 g for 10 minutes at 4 °C in ice cold PBS prior to fixation in at least 5 pellet volumes of freshly made 4% paraformaldehyde in PBS for 16-20 hour at 4°C. Fixed parasites were washed twice with cold PBS as above and then resuspended at 5×10^8 parasites/ml in PBS, fixed parasites were stored up to 1 week at 4°C.

20 µl of fixed parasites were spotted into wells of geneframe hybridisation chambers that had been attached (strong adhesive side down) (Thermo Fisher Scientific) to Menzel Superfrost plus slides. Parasites were incubated with slides for 1 hour in a humidified chamber and then permeabilised with 0.1% TX100 in human tonicity-PBS for 5 min. Slides were washed three times with 100 µl/well PBS.

The hybridisation chamber wells were then blocked for 30 minutes at 42 °C in a humidified chamber with denatured (5 minutes at 95°C) hybridisation solution (50% formamide, 10% dextran sulfate (Sigma), 2xSSPE pH 7.4 (0.3M NaCl, 0.02M NaH_2PO_4, 0.002M EDTA), 250µg/ml Herring sperm DNA (Sigma).

Probes in hybridisation solution were denatured at 95°C for 10 minutes and then incubated in an ice water bath for 5 min. The blocking hybridisation solution in the hybridisation chamber wells was replaced with denatured probe and the wells were covered with the *in situ* frame coverslip (Thermo Fisher Scientific) prior to denaturing the gDNA on the slides on a heater block for 30 minutes at 80°C followed by hybridization at 37°C overnight. Initially probe

concentration was optimised empirically by making 1/54, 1/27 and 1/9 dilutions in a final volume 27 µl hybridisation solution (gene frames take 25 µl).

The next day the hybridisation solution and coverslip were carefully removed, avoiding simultaneous removal of the geneframe. Each well was rinsed with 100 µl 50% formamide 2xSSC at 50°C then the entire slide was washed in 30 ml 2xSSC 50% formamide for 30 minutes at 37°C; then in 30 ml 1xSSC for 15 minutes at 50°C; then in 30 ml 2xSSC for 15 minutes at 50°C; then in 30 ml 4xSSC for 15 minutes at 50°C; then in 30 ml PBS pH 7.4 for 1 minutes at room temperature.

The slide was then blocked with 50 ul/well 4% bovine serum albumin/PBS pH 7.4 for 30 minutes in a humid chamber at room temperature prior to incubation with 50 ul/well of antibodies in 4% bovine serum albumin/PBS for 30 minutes in a humid chamber at room temperature. Biotinylated probes were detected with 5 µg/ml goat anti-biotin diluted (#B3640 Sigma diluted with 0.135 M NaCl 15mM NaN3). Digoxigenin probes were detected with mouse anti digoxigenin monoclonal antibody diluted 1/200 in 4% bovine serum albumin in PBS. Slides were then washed in TNT (100 mM Tris-HCl, 150 mM NaCl, 0.5% Tween 20, pH 7.5) for 10 minutes and incubated with 50 ul/well secondary antibody in 4% bovine serum albumin in PBS for 30 minutes in a humid chamber at room temperature. Biotin probes were detected with donkey anti-goat IgG 594 diluted 1:200 and digoxigenin probes were detected with rabbit anti mouse 488 diluted 1/500. Slides were then washed three times with PBS and gene frames were removed prior to mounting slides with Vectashield + DAPI and sealing #1.5 coverslips with nail polish.

Slides were imaged using a Zeiss Axioskop fluorescence microscope with a PLAN-NEOFLUAR 100x 1.30 oil immersion objective and an AxioCam Mrm black and white camera to capture 1300x1030 µm at 0.0043 µm^2 per pixel and 1 x zoom. All image analysis was performed using ImageJ (Fiji) [87, 88].

IFA RNA FISH

All solutions were made using diethylpyrocarbonate treated water.

IFA. Parasites were fixed as for the DNA probes and then fixed cells were immediately immobilised on slides and permeabilisation was conducted for 10 min. Slides were blocked for 30 minutes in a humid chamber at room temperature with 50 µL/well 4% bovine serum

albumin in PBS pH7.4 with 0.4 U/µl RNA guard (Life technologies). Slides were incubated with 50 µL/well primary antibodies in 1% bovine serum albumin in PBS with 0.4 U/µl RNA guard for 60 minutes in a humid chamber at room temperature. Primary antibodies were 1/400 rabbit anti-Pf H2A.Z [59], 1/200 mouse monoclonal anti-HA 12CA5 (Sigma Aldrich) [60], 1/800 rabbit anti-H3K9me3 (Active Motif: cat. no. 39161), 1/200 rabbit anti-H3K9ac (Millipore cat. no. 06-942), 1/400 rabbit anti-H3K4me3 (Millipore cat. no.04-745), 1/800 rabbit anti-H3K4me1 (Abcam cat. no.Ab8895), 1/800 rabbit anti-H3K27ac (Abcam cat. no.Ab4729). Slides were then washed in PBS three times for five minutes each on a rocking platform and then incubated with 50µl/well secondary antibodies in 1% bovine serum albumin in PBS with 0.4 U/µl RNA guard for at least 1h in a humid chamber at room temperature in the dark. From this point onwards slides were shielded from exposure to light. Secondary antibodies were 1/500 rabbit anti-mouse Ig conjugated to Alexafluor 488 (Invitrogen) and 1/500 donkey anti-rabbit conjugated to Alexafluor 488 (Invitrogen). Slides were then washed in PBS three times for 5 minutes each and then the fluorescent labelled proteins were post-fixed with fresh 4% PFA, 0.0075% Glutaraldehyde in PBS for 30 minutes at room temperature, washed for 5 minutes with PBS and then fixation quenched with 0.1mg/ml NaBH4 in PBS for 10 minutes in 50 ml tubes on a roller. Slides were then washed once with PBS. Non-immune 1/200 mouse IgG and 1/400 rabbit IgG primary antibody controls were included with each assay.

Probe transcription. One µg of linearised plasmid containing the *var2csa* probe sequence upstream of a T7 promoter was *in vitro* transcribed at 37ºC for 4 hours in 20 µl using 2 µl Roche 10x digoxigenin labeling mix, 2 µl 10x reaction buffer 2 µl Ambion megascript T7 mix. The reaction was stopped with 1 ul 0.5 M EDTA pH 8 and ethanol precipitated, washed and stored at -80ºC as for the DNA probes.

RNA FISH. hybridisation solution was denatured at 95 °C for 5 minutes then placed on ice, 50 µl was used to block to each well for 30 minutes at 42 °C in a humidified chamber. The *var2csa* as RNA probe was diluted 1/27 in hybridisation solution and denatured at 65 °C for 5 minutes and then placed into an ice water bath for 5 min. The hybridisation solution blocking solution was aspirated from the wells on slides and replaced with the denatured probe, controls with denatured hybridisation solution only were also included. The wells were covered with the *in situ* frame coverslip and parasites were hybridised with the probe at 42 °C overnight. The next day the coverslip and probe were removed and slides were washed twice

in 30ml 2x SSC 50% formamide for 5 minutes at 45°C then washed twice in 30ml 2xSSC for 5 minutes at 37 °C then washed in 30ml PBS pH7.4 for 1 minutes at room temperature. The wells were then blocked with 50 μl each of 4%bovine serum albumin/PBS pH 7.4 with 0.4U/μl RNAguard for 30 minutes in a humid chamber at room temperature and then incubated with mouse anti digoxigenin diluted 1/200 in 4% bovine serum albumin/PBS and 0.4U/μl RNAguard for 30 minutes in a humid chamber at room temperature. Slides were washed with TNT twice for 5 minutes each and then incubated with rabbit anti mouse conjugated to Alexafluor 488 diluted 1/500 in 4% bovine serum albumin/PBS and with 0.4U/μl RNA guard for 30 minutes in a humid chamber at room temperature. Slides were washed three times with PBS for minutes each, the geneframes were removed and samples mounted with Vectashield (with DAPI) and #1.5 coverslips sealed with nail polish

Slides were imaged with an Olympus FV1000 microscope with a uplan SAPO ∞/0.1/FN26.5 100x 1.40 oil immersion objective. Images were acquired at 512x512 pixel field size with 2.5 x optical zoom using Kalman filtering with 3 repeats and sequential scans by line. All image analysis was performed using ImageJ (Fiji). Colocalisation for a region of interest containing the single nucleus analysed was performed using Coloc2 to determine Manders M1 and M2 coefficients with Costes automatic thresholds. For a defined nucleus M2 is the sum of the grey value of voxels of *var2csa* mRNA that colocalised with voxels of a histone modification/sum of the total grey value of voxels of *var2csa* mRNA. M1 is the sum of the grey value of voxels of a histone modification that colocalised with voxels of *var2csa* mRNA /sum of the total grey value of voxels of the histone modification [49]. Comparison between histone modifications of distance to *var2csa* FISH peak fluorescence was performed by Kruskal Wallis test. Manders coefficients were compared between histone(s)/modifications using the Kruskal Wallis test with Dunn's multiple comparison test to compare all other histone(s) modifications to H3K9me3.

Chromatin immunoprecipitation (ChIP). Antibodies used for ChIP were rabbit anti-H3 (Abcam Ab1791), rabbit anti-H3K4me1 (Abcam Ab8895), rabbit anti-H3K27ac (Abcam Ab4729) and rabbit control IgG (Abcam Ab46540).

Cross-linked ChIP was performed as previously described [60] with slight modifications. Chromatin was harvested at early ring (6-14 hpi) and late schizont (36-44 hpi) stages during the intraerythrocytic development cycle (IDC). Chromatin was sheared into 200-800 bp fragments by sonication in a Bioruptor UCD-200 (Diagenode) for 28 minutes on high at 30 s

intervals. ChIP was performed with the EZ ChIP Kit (Millipore) with exclusion of the beads. For each IP, approximately 5.5×10^8 ring or 1×10^8 schizont stage parasites, 8 μg antibody and 15 μl protein A/G beads (GE Life Sciences 17-0618-01 & 17-5280-01) were used. Cross-linking of the immune complexes was reversed overnight at 45°C after addition of 500 mM NaCl. DNA was purified using the MinElute® PCR purification kit (Qiagen).

Quantitative real time PCR (qPCR) method and the primers used were described previously (Petter *et al.*, 2013). For each PCR, enrichment was calculated as $2^{-\Delta Ct} = 2^{-(Ct\ histone\ ChIP\ -\ Ct\ non-immune\ ChIP)}$. Enrichment ratios for each PCR were normalised for nucleosomal occupancy, which is lower in the higher AT content non-coding sequence than in the coding sequence, using H3 ChIP ($2^{-(\Delta Ct\ H3K27ac\ or\ H3K4me1)}/2^{-\Delta Ct\ H3}$). Enrichment values were then normalised for ChIP efficiency by dividing by the ($2^{-(\Delta Ct\ H3K27ac\ or\ H3K4me1)}/2^{-\Delta Ct\ H3}$) value for the hsp70 coding sequence PCR for that ChIP [60]. This enabled replicates from different ChIPs to be aggregated for statistical comparison despite the usual variation between experiments in the percentage of input material precipitated. It is important to note that as a consequence of this transformation the resulting y-axis values are a ratio relative to H3 levels in the coding sequence of Hsp70 rather than a fold enrichment relative to ChIP with non-immune serum. Difference between ring and schizont stages in histone modification levels at single PCRs was tested across each gene and parasite line separately using a table of t tests using the Two-stage linear step-up procedure of Benjamini, Krieger and Yeuketeli to determine the q value from multiple comparison testing. Because there was no difference between rings and schizonts for either modification at any position, the two ring stage and two schizont stage replicates for each parasite line were combined for subsequent analysis. Differences in levels of histone modifications between different positions in genes in 3C and 3D7 parasites was then tested across each gene separately by Friedman test with Dunn's test for multiple comparisons, where each PCR for the gene in 3C and 3D7 was compared with the 3D7 promoter PCR, either -1000 for *var2csa* or -450 for PFL0020w. Analysis was performed using GraphPad Prism.

Author contributions

MFD planned experiments, MFD, JT, FS, HN, SS, GAJ, MP performed experiments and analysed data, GVB and KPD contributed reagents or other essential material, MFD, JT, FS, HN, SS, GAJ, MP, KPD and GVB wrote the paper.

References

1. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Elsen, J. A., Rutherford, K., Saizberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angluoll, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Valdya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. D., Subramanlan, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. & Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature.* **419**, 498-511.

2. Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I. & Miller, L. H. (1995) Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes, *Cell.* **82**, 101-110.

3. Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A. & Wellems, T. E. (1995) The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes, *Cell.* **82**, 89-100.

4. Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F. & Howard, R. J. (1995) Cloning the *P.falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes, *Cell.* **82**, 77-87.

5. Voss, T. S., Healer, J., Marty, A. J., Duffy, M. F., Thompson, J. K., Beeson, J. G., Reeder, J. C., Crabb, B. S. & Cowman, A. F. (2006) A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria, *Nature.* **439**, 1004-1009.

6. Joergensen, L., Bengtsson, D. C., Bengtsson, A., Ronander, E., Berger, S. S., Turner, L., Dalgaard, M. B., Cham, G. K., Victor, M. E., Lavstsen, T., Theander, T. G., Arnot, D. E. & Jensen, A. T. (2010) Surface co-expression of two different PfEMP1 antigens on single plasmodium falciparum-infected erythrocytes facilitates binding to ICAM1 and PECAM1, *PLoS Pathog.* **6**, e1001083.

7. Biggs, B. A., Anders, R. F., Dillon, H. E., Davern, K. M., Martin, M., Petersen, C. & Brown, G. V. (1992) Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*, *J Immunol.* **149**, 2047-2054.

8. Miller, L. H. (1969) Distribution of mature trophozoites and schizonts of *Plasmodium falciparum* in the organs of Aotus trivirgatus, the night monkey., *Am J Trop Med Hyg.* **18**, 860-865.

9. Bach, O., Baier, M., Pullwitt, A., Fosiko, N., Chagaluka, G., Kalima, M., Pfister, W., Straube, E. & Molyneux, M. (2005) Falciparum malaria after splenectomy: a prospective controlled study of 33 previously splenectomized Malawian adults, *Trans R Soc Trop Med Hyg.* **99**, 861-7.

10. Bachmann, A., Esser, C., Petter, M., Predehl, S., von Kalckreuth, V., Schmiedel, S., Bruchhaus, I. & Tannich, E. (2009) Absence of erythrocyte sequestration and lack of multicopy gene family expression in Plasmodium falciparum from a splenectomized malaria patient, *PLoS One.* **4**, e7459.

11. Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J. & Lanzer, M. (1998) Antigenic variation in malaria: *in situ* switching, relaxed and mutally exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*, *EMBO J.* **17**, 5418-5426.

12. Frank, M., Dzikowski, R., Amulic, B. & Deitsch, K. (2007) Variable switching rates of malaria virulence genes are associated with chromosomal position, *Mol Microbiol.* **64**, 1486-1498.

13. Biggs, B. A., Goozé, L., Wycherley, K., Wollish, W., Southwell, B., Leech, J. H. & Brown, G. V. (1991) Antigenic variation in *Plasmodium falciparum*, *Proc Natl Acad Sci USA.* **88**, 9171-9174.

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

15. Petter, M. & Duffy, M. F. (2015) Antigenic Variation in Plasmodium falciparum, *Results Probl Cell Differ.* **57**, 47-90.

16. Lopez-Rubio, J. J., Gontijo, A. M., Nunes, M. C., Issar, N., Hernandez Rivas, R. & Scherf, A. (2007) 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites, *Mol Microbiol.* **66**, 1296-1305.

17. Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C. & Scherf, A. (2002) A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation, *EMBO J.* **21**, 815-824.

18. Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T. E. & Scherf, A. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *Plasmodium falciparum*, *Nature.* **407**, 1018-1022.

19. Marty, A. J., Thompson, J. K., Duffy, M. F., Voss, T. S., Cowman, A. F. & Crabb, B. S. (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.* **62**, 72-83.

20. Ralph, S. A., Scheidig-Benatar, C. & Scherf, A. (2005) Antigenic variation in *Plasmodium falciparum* is associated with movement of *var* loci between subnuclear locations, *Proc Natl Acad Sci U S A.* **102**, 5414-5419.

21. Mok, B. W., Ribacke, U., Rasti, N., Kironde, F., Chen, Q., Nilsson, P. & Wahlgren, M. (2008) Default Pathway of *var2csa* switching and translational repression in *Plasmodium falciparum*, *PLoS ONE.* **3**, e1982.

22. Brolin, K. J., Ribacke, U., Nilsson, S., Ankarklev, J., Moll, K., Wahlgren, M. & Chen, Q. (2009) Simultaneous transcription of duplicated var2csa gene copies in individual Plasmodium falciparum parasites, *Genome Biol.* **10**, R117.

23. Lopez-Rubio, J. J., 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**, 179-90.

24. Zhang, Q., Huang, Y., Zhang, Y., Fang, X., Claes, A., Duchateau, M., Namane, A., Lopez-Rubio, J. J., Pan, W. & Scherf, A. (2011) A critical role of perinuclear filamentous actin in spatial repositioning and mutually exclusive expression of virulence genes in malaria parasites, *Cell Host Microbe.* **10**, 451-63.

25. Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J. P., Noble, W. S. & Le Roch, K. G. (2014) Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression, *Genome Res.* **24**, 974-88.

26. Lemieux, J. E., Kyes, S. A., Otto, T. D., Feller, A. I., Eastman, R. T., Pinches, R. A., Berriman, M., Su, X. Z. & Newbold, C. I. (2013) Genome-wide profiling of chromosome

interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation, *Mol Microbiol*.

27. Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H. & Gasser, S. M. (1996) The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae, *J Cell Biol.* **134**, 1349-63.

28. Wesolowska, N., Amariei, F. L. & Rong, Y. S. (2013) Clustering and protein dynamics of Drosophila melanogaster telomeres, *Genetics.* **195**, 381-91.

29. Crabbe, L., Cesare, A. J., Kasuboski, J. M., Fitzpatrick, J. A. & Karlseder, J. (2012) Human telomeres are tethered to the nuclear envelope during postmitotic nuclear assembly, *Cell Rep.* **2**, 1521-9.

30. Gonzalez-Sandoval, A., Towbin, B. D., Kalck, V., Cabianca, D. S., Gaidatzis, D., Hauer, M. H., Geng, L., Wang, L., Yang, T., Wang, X., Zhao, K. & Gasser, S. M. (2015) Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos, *Cell.* **163**, 1333-47.

31. Kueng, S., Oppikofer, M. & Gasser, S. M. (2013) SIR proteins and the assembly of silent chromatin in budding yeast, *Annu Rev Genet.* **47**, 275-306.

32. Antoniacci, L. M., Kenna, M. A. & Skibbens, R. V. (2007) The nuclear envelope and spindle pole body-associated Mps3 protein bind telomere regulators and function in telomere clustering, *Cell Cycle.* **6**, 75-9.

33. Salcedo-Amaya, A. M., van Driel, M. A., Alako, B. T., Trelle, M. B., van den Elzen, A. M., Cohen, A. M., Janssen-Megens, E. M., van de Vegte-Bolmer, M., Selzer, R. R., Iniguez, A. L., Green, R. D., Sauerwein, R. W., Jensen, O. N. & Stunnenberg, H. G. (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*, *Proc Natl Acad Sci U S A.* **106**, 9655-60.

34. Deshmukh, A. S., Srivastava, S., Herrmann, S., Gupta, A., Mitra, P., Gilberger, T. W. & Dhar, S. K. (2012) The role of N-terminus of Plasmodium falciparum ORC1 in telomeric localization and var gene silencing, *Nucleic Acids Res.* **40**, 5313-31.

35. Varunan, S. M., Tripathi, J., Bhattacharyya, S., Suhane, T. & Bhattacharyya, M. K. (2013) Plasmodium falciparum origin recognition complex subunit 1 (PfOrc1) functionally complements Deltasir3 mutant of Saccharomyces cerevisiae, *Mol Biochem Parasitol.* **191**, 28-35.

36. Moraes, C. B., Dorval, T., Contreras-Dominguez, M., Dossin Fde, M., Hansen, M. A., Genovesio, A. & Freitas-Junior, L. H. (2013) Transcription sites are developmentally regulated during the asexual cycle of Plasmodium falciparum, *PLoS One.* **8**, e55539.

37. Dzikowski, R., Li, F., Amulic, B., Eisberg, A., Frank, M., Patel, S., Wellems, T. E. & Deitsch, K. W. (2007) Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites, *EMBO Rep.* **8**, 959-65.

38. Duraisingh, M. T., Voss, T. S., Marty, A. J., Duffy, M. F., Good, R. T., Thompson, J. K., Freitas-Junior, L. H., Scherf, A., Crabb, B. S. & Cowman, A. F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*, *Cell.* **121**, 13-24.

39. Issar, N., Ralph, S. A., Mancio-Silva, L., Keeling, C. & Scherf, A. (2009) Differential sub-nuclear localisation of repressive and activating histone methyl modifications in *P. falciparum*, *Microbes Infect.* **11**, 403-7.

40. Volz, J. C., Bartfai, R., Petter, M., Langer, C., Josling, G. A., Tsuboi, T., Schwach, F., Baum, J., Rayner, J. C., Stunnenberg, H. G., Duffy, M. F. & Cowman, A. F. (2012) PfSET10, a *Plasmodium falciparum* Methyltransferase, Maintains the Active var Gene in a Poised State during Parasite Division, *Cell Host Microbe.* **11**, 7-18.

41. Guizetti, J., Martins, R. M., Guadagnini, S., Claes, A. & Scherf, A. (2013) Nuclear pores and perinuclear expression sites of var and ribosomal DNA genes correspond to physically distinct regions in Plasmodium falciparum, *Eukaryot Cell.* **12**, 697-702.

42. Sierra-Miranda, M., Delgadillo, D. M., Mancio-Silva, L., Vargas, M., Villegas-Sepulveda, N., Martinez-Calvillo, S., Scherf, A. & Hernandez-Rivas, R. (2012) Two long non-coding RNAs generated from subtelomeric regions accumulate in a novel perinuclear compartment in Plasmodium falciparum, *Mol Biochem Parasitol.* **185**, 36-47.

43. Coleman, B. I., Ribacke, U., Manary, M., Bei, A. K., Winzeler, E. A., Wirth, D. F. & Duraisingh, M. T. (2012) Nuclear repositioning precedes promoter accessibility and is linked to the switching frequency of a Plasmodium falciparum invasion gene, *Cell Host Microbe.* **12**, 739-50.

44. Duffy, M. F., Byrne, T. J., Carret, C., Ivens, A. & Brown, G. V. (2009) Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate, *J Mol Biol.* **389**, 453-69.

45. Chaumeil, J., Augui, S., Chow, J. C. & Heard, E. (2008) Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation, *Methods Mol Biol.* **463**, 297-308.

46. Namekawa, S. H. & Lee, J. T. (2011) Detection of nascent RNA, single-copy DNA and protein localization by immunoFISH in mouse germ cells and preimplantation embryos, *Nat Protoc.* **6**, 270-84.

47. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. (2006) Stochastic mRNA synthesis in mammalian cells, *PLoS Biol.* **4**, e309.

48. Vargas, D. Y., Raj, A., Marras, S. A., Kramer, F. R. & Tyagi, S. (2005) Mechanism of mRNA transport in the nucleus, *Proc Natl Acad Sci U S A.* **102**, 17008-13.

49. Manders, E. M. M., Verbeek, F. J. & Aten, J. A. (1993) Measurement of co-localisation of objects in dual-colour confocal images, *Journal of Microscopy.* **169**, 375-382.

50. Dunham, I.Kundaje, A.Aldred, S. F.Collins, P. J.Davis, C. A.Doyle, F.Epstein, C. B.Frietze, S.Harrow, J.Kaul, R.Khatun, J.Lajoie, B. R.Landt, S. G.Lee, B. K.Pauli, F.Rosenbloom, K. R.Sabo, P.Safi, A.Sanyal, A.Shoresh, N.Simon, J. M.Song, L.Trinklein, N. D.Altshuler, R. C.Birney, E.Brown, J. B.Cheng, C.Djebali, S.Dong, X.Ernst, J.Furey, T. S.Gerstein, M.Giardine, B.Greven, M.Hardison, R. C.Harris, R. S.Herrero, J.Hoffman, M. M.Iyer, S.Kelllis, M.Kheradpour, P.Lassman, T.Li, Q.Lin, X.Marinov, G. K.Merkel, A.Mortazavi, A.Parker, S. C.Reddy, T. E.Rozowsky, J.Schlesinger, F.Thurman, R. E.Wang, J.Ward, L. D.Whitfield, T. W.Wilder, S. P.Wu, W.Xi, H. S.Yip, K. Y.Zhuang, J.Bernstein, B. E.Green, E. D.Gunter, C.Snyder, M.Pazin, M. J.Lowdon, R. F.Dillon, L. A.Adams, L. B.Kelly, C. J.Zhang, J.Wexler, J. R.Good, P. J.Feingold, E. A.Crawford, G. E.Dekker, J.Elinitski, L.Farnham, P. J.Giddings, M. C.Gingeras, T. R.Guigo, R.Hubbard, T. J.Kellis, M.Kent, W. J.Lieb, J. D.Margulies, E. H.Myers, R. M.Starnatoyannopoulos, J. A.Tennebaum, S. A.Weng, Z.White, K. P.Wold, B.Yu, Y.Wrobel, J.Risk, B. A.Gunawardena, H. P.Kuiper, H. C.Maier, C. W.Xie, L.Chen, X.Mikkelsen, T. S., et al. (2012) An integrated encyclopedia of DNA elements in the human genome, *Nature.* **489**, 57-74.

51. Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. & Jaenisch, R. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state, *Proc Natl Acad Sci U S A.* **107**, 21931-6.

52. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions, *Genome Res.* **21**, 1273-83.

53. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome, *Cell.* **129**, 823-37.

54. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q. & Zhao, K. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome, *Nat Genet.* **40**, 897-903.

55. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. & Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, *Nat Genet.* **39**, 311-8.

56. Hon, G., Wang, W. & Ren, B. (2009) Discovery and annotation of functional chromatin signatures in the human genome, *PLoS Comput Biol.* **5**, e1000566.

57. Miao, J., Fan, Q., Cui, L. & Li, J. (2006) The malaria parasite *Plasmodium falciparum* histones: Organization, expression, and acetylation, *Gene.* **369**, 53-65.

58. Trelle, M. B., Salcedo-Amaya, A. M., Cohen, A. M., Stunnenberg, H. G. & Jensen, O. N. (2009) Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum*, *J Proteome Res.* **8**, 3439-50.

59. Petter, M., Lee, C. C., Byrne, T. J., Boysen, K. E., Volz, J., Ralph, S. A., Cowman, A. F., Brown, G. V. & Duffy, M. F. (2011) Expression of *P. falciparum* var genes involves exchange of the histone variant H2A.Z at the promoter, *PLoS Pathog.* **7**, e1001292.

60. Petter, M., Selvarajah, S. A., Lee, C. C., Chin, W. H., Gupta, A. P., Bozdech, Z., Brown, G. V. & Duffy, M. F. (2013) H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*, *Mol Microbiol*.

61. Ukaegbu, U. E., Kishore, S. P., Kwiatkowski, D. L., Pandarinath, C., Dahan-Pasternak, N., Dzikowski, R. & Deitsch, K. W. (2014) Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in P. falciparum, *PLoS Pathog.* **10**, e1003854.

62. Jiang, L., Mu, J., Zhang, Q., Ni, T., Srinivasan, P., Rayavara, K., Yang, W., Turner, L., Lavstsen, T., Theander, T. G., Peng, W., Wei, G., Jing, Q., Wakabayashi, Y., Bansal, A., Luo, Y., Ribeiro, J. M., Scherf, A., Aravind, L., Zhu, J., Zhao, K. & Miller, L. H. (2013) PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum, *Nature.* **499**, 223-7.

63. Ukaegbu, U. E., Zhang, X., Heinberg, A. R., Wele, M., Chen, Q. & Deitsch, K. W. (2015) A Unique Virulence Gene Occupies a Principal Position in Immune Evasion by the Malaria Parasite Plasmodium falciparum, *PLoS Genet.* **11**, e1005234.

64. Mayfield, J. E., Burkholder, N. T. & Zhang, Y. J. (2016) Dephosphorylating eukaryotic RNA polymerase II, *Biochim Biophys Acta.* **1864**, 372-87.

65. Guttery, D. S., Poulin, B., Ramaprasad, A., Wall, R. J., Ferguson, D. J., Brady, D., Patzewitz, E. M., Whipple, S., Straschil, U., Wright, M. H., Mohamed, A. M., Radhakrishnan, A., Arold, S. T., Tate, E. W., Holder, A. A., Wickstead, B., Pain, A. & Tewari, R. (2014) Genome-wide functional analysis of Plasmodium protein phosphatases reveals key regulators of parasite development and differentiation, *Cell Host Microbe.* **16**, 128-40.

66. Shevtsov, S. P. & Dundr, M. (2011) Nucleation of nuclear bodies by RNA, *Nat Cell Biol.* **13**, 167-73.

67. Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S., Eskiw, C. H., Luo, Y., Wei, C. L., Ruan, Y., Bieker, J. J. & Fraser, P. (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells, *Nat Genet.* **42**, 53-61.

68. Weiner, A., Dahan-Pasternak, N., Shimoni, E., Shinder, V., von Huth, P., Elbaum, M. & Dzikowski, R. (2011) 3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum, *Cell Microbiol.* **13**, 967-77.

69. Avraham, I., Schreier, J. & Dzikowski, R. (2012) Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite Plasmodium falciparum, *Proc Natl Acad Sci U S A.* **109**, E3678-86.

70. Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E. & Dzikowski, R. (2015) Antisense long noncoding RNAs regulate var gene activation in the malaria parasite Plasmodium falciparum, *Proc Natl Acad Sci U S A.* **112**, E982-91.

71. Sims, J. S., Militello, K. T., Sims, P. A., Patel, V. P., Kasper, J. M. & Wirth, D. F. (2009) Patterns of gene-specific and total transcriptional activity during the Plasmodium falciparum intraerythrocytic developmental cycle, *Eukaryot Cell.* **8**, 327-38.

72. Saraf, A., Cervantes, S., Bunnik, E. M., Ponts, N., Sardiu, M. E., Chung, D. W., Prudhomme, J., Varberg, J. M., Wen, Z., Washburn, M. P., Florens, L. & Le Roch, K. G. (2016) Dynamic and Combinatorial Landscape of Histone Modifications during the Intraerythrocytic Developmental Cycle of the Malaria Parasite, *J Proteome Res.* **15**, 2787-801.

73. Josling, G. A., Petter, M., Oehring, S. C., Gupta, A. P., Dietz, O., Wilson, D. W., Schubert, T., Längst, G., Gilson, P. R., Crabb, B. S., Moes, S., Jenoe, P., Lim, S. W., Brown,

G. V., Bozdech, Z., Voss, T. S. & Duffy, M. F. (2015) A *Plasmodium Falciparum* Bromodomain Protein Regulates Invasion Gene Expression, *Cell Host & Microbe.* **17**, 741-751.

74. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner, *Bioinformatics.* **29**, 15-21.

75. Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3, *Fly (Austin).* **6**, 80-92.

76. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res.* **20**, 1297-303.

77. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & Genomes Project Analysis, G. (2011) The variant call format and VCFtools, *Bioinformatics.* **27**, 2156-8.

78. Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M. & Lu, X. (2012) Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift, *Front Genet.* **3**, 35.

79. Liao, Y., Smyth, G. K. & Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics.* **30**, 923-30.

80. Love, M. I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* **15**, 550.

81. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biol.* **14**, R36.

82. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. & Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat Biotechnol.* **28**, 511-5.

83. Viebig, N. K., Gamain, B., Scheidig, C., Lepolard, C., Przyborski, J., Lanzer, M., Gysin, J. & Scherf, A. (2005) A single member of the *Plasmodium falciparum var* multigene family

determines cytoadhesion to the placental receptor chondroitin sulphate A, *EMBO Rep.* **6**, 775-81.

84. Duffy, M. F., Caragounis, A., Noviyanti, R., Kyriacou, H. M., Choong, E. K., Boysen, K., Healer, J., Rowe, J. A., Molyneux, M. E., Brown, G. V. & Rogerson, S. J. (2006) Transcribed *var* genes associated with placental malaria in Malawian women, *Infect Immun.* **74**, 4875-4883.

85. Contreras-Dominguez, M., Moraes, C. B., Dorval, T., Genovesio, A., Dossin Fde, M. & Freitas-Junior, L. H. (2010) A modified fluorescence in situ hybridization protocol for Plasmodium falciparum greatly improves nuclear architecture conservation, *Mol Biochem Parasitol.* **173**, 48-52.

86. O'Donnell, R. A., Freitas-Junior, L. H., Preiser, P. R., Williamson, D. H., Duraisingh, M., McElwain, T. F., Scherf, A., Cowman, A. F. & Crabb, B. S. (2002) A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes, *EMBO J.* **21**, 1231-9.

87. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis, *Nat Methods.* **9**, 671-5.

88. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis, *Nat Methods.* **9**, 676-82.

Supporting Information

Table S1) Differential gene expression comparison of CS2 compared to EIC and ECS using the IT genome version 28. baseMean are mean read counts reported using feature counts and log2 fold change was calculated using DEseq2.

Figure legends

Fig. 1 A) Derivation of the parasite lines E8B, ECS, EIC, CS2 and the CS2 subclones. Arrows labelled with CSA or ICAM1 indicate that the parasites lines were selected for adhesion to these receptors, arrows labelled with clone indicate that the parasites were cloned at this step. The table below indicates the relative level of binding of each parasite line/subclone to immobilised CSA, ICAM1 and CD36 at recovery from selection (approximately 14 days post selection) or cloning (approximately 80 days post selection).

The table summarises previously published data for CS2isB10, CS2isC8, CS2isC9 and CS2isD8 [44]. B) The adhesion profile over time of sub-clone CS2isD4.

Fig. 2A) Diagram of *var2csa* and neighbouring genes in the left hand subtelomeric region of chromosome 12 in CS2 parasites, arrows indicate exons in 5' to 3' direction. The positions of the *var2csa* exon 1 probe and the *acyl-coA synthetase 7* probe that were used to probe Southern blots is indicated as blue bars above the diagram and the positions of restriction enzymes used to digest *P. falciparum* gDNA are indicated below the diagram. The positions and sizes of the predicted restriction fragments from CS2 that would hybridise to the probes are indicated by bars below the diagram. The colour of the bars indicates that probes hybridised to these fragments in Southern blot lanes labelled with the same coloured text in panels B and C. B) Southern blot probed with *var2csa* exon 1 probe. C) Southern blot probed with *acyl-coA synthetase 7* probe. Lanes in autoradiographs of Southern blots are labelled with the names of the parasite lines from which gDNA was obtained, CS2, and the CS2 subclones CS2isD4 (D4) and CS2isC8 (C8) and with the restriction enzymes used to digest the gDNA in those lanes.

Fig. 3) MA plot of RNAseq differential gene expression analysis of CS2 compared to EIC and ECS treated as replicates.

Fig. 4A) Quantitative reverse transcription PCR (Q-RT-PCR) of CS2 selected for adhesion to CSA and parental clone E8B selected for adhesion to CSA (ECS) or ICAM-1 (EIC). Shown are values from $2^{-\Delta\Delta Ct}$ analysis normalised between samples using cDNA levels of the skeleton binding protein 1 gene and with the amount of each transcript presented relative to the amount of that gene found in a sample of E8B gDNA. B) The proportion each *var* gene contributed to total *var* transcripts. CSA selected 3C parasites were analysed by RNAseq read counts and ICAM1 selected 3D7-icam parasites were analysed by quantitative reverse transcription PCR (Q-RT-PCR) with gene levels determined relative to the amount of each gene in 3D7 gDNA.

Fig. 5 A) Representative DNA FISH images of early ring stage parasites showing loci hybridising with *var2csa* and rep20 probes and both probes merged with the DNA stain DAPI. B) The proportion of nuclei with *var2csa* FISH signals that co-localised with rep20 FISH signals were determined for each field of view in biological triplicate experiments. Error bars indicate standard deviation, *** p<0.001 (unpaired t test).

Fig. 6 A) Representative IFA combined with RNA FISH: immunostained histone variants and histone modifications in 3D7 ring stage parasites expressing HA epitope tagged Pf H2B.Z are in panels in the first column, panels in the second column are FISH *var2csa* asRNA probe signals, the third column are FISH and IFA signals merged with DAPI, lines indicate the regions of interest that were drawn between regions of greatest fluorescent intensity in IFA and FISH signals, scale bar is one micron. The fluorescence profiles for FISH, IFA and DAPI along the regions of interest for each image are shown in the fourth column, y axis is fluorescence intensity. B) Zoomed out view of IFA RNA FISH as for panel A.

Fig. 7 A) Distance between peaks of fluorescence for IFA and RNA-FISH signals. The line is median, the box is interquartile range, and the whiskers are maximum and minimum values. Red is H3K27ac, blue is H3K9me3.B) Manders M1 and C) Manders M2 coefficients for proportion of IFA signal co-localising with *var2csa* RNA (M1) and the proportion of *var2csa* RNA co-localising with the IFA signal (M2). The line is median, the box is interquartile range, and the whiskers are minimum and maximum, * p<0.05, ** p<0.01 (Kruskal Wallis test with Dunn's multiple comparisons test). For both analyses equal numbers of nuclei from biological duplicate experiments were analysed with the total number of nuclei analysed for each histone variant or modification indicated in parentheses after the histone/modification label.

Fig. 8 Chromatin immunoprecipitation (ChIP) of 3C parasites that expressed *var2csa* but not *PFL0020w*, 3D7 parasites that did not express *var2csa* or *PFL0020w* and 3D7-icam parasites that expressed some *PFL0020w*. Chromatin from all cell lines was immunoprecipitated with anti-H3K27ac and anti-H3K4me1, anti-H3 and non-immune antibodies. The precipitated DNA was analysed by quantitative PCR, shown are the fold change enrichment relative to non-immune ChIP and normalised for input, nucleosomal occupancy (H3 levels) and ChIP efficiency (levels in *hsp70* coding sequence) (see methods). ChIPs were performed in at least biological duplicate. A) The level of enrichment of each histone modification at each position tested. The x axes are labelled to indicate position of the Q-PCR primers within the *var2csa* or *PFL0020w* genes. Numbers are relative to the start codons, DBL refers to the exon 1 DBL domains. Bars are means and error bars are SD B) The level of enrichment for the aggregated values for positions -800, -600, -450 and -100 in the *PFL0020w* 5' untranslated region (5'UTR) and for the aggregated values for positions DBL1 and DBL5 in the *PFL0020w*

coding sequence (CDS). The lines are medians, the boxes are interquartile ranges, and the whiskers are maximum and minimum values (**** p<0.0001, Mann Whitney test).

Fig. 9 A) Chromatin immunoprecipitated from 3D7 parasites using antibodies to H3K27ac and H3 was analysed by quantitative PCR, shown are the fold change enrichment relative to non-immune ChIP and normalised for input and nucleosomal occupancy (H3 levels) for biological duplicates (mean and standard deviation). B) The levels of expression of the genes analysed by ChIP was determined by RNAseq of 3D7 parasites (fpkm is Fragments Per Kilobase of transcript per Million mapped reads).

1A)

	ECS	EIC		CS2	CS2 isB10	CS2 isC8	CS2 isC9	CS2 isD8	CS2 isD4	CS2 isB3
CSA	+++	-		+++	++	-	++	++	-	+++
ICAM1	-	++		-	+++	+	++	++	+	-
CD36	-	+++		-	+++	+++	+++	+++	+++	-

CS2isD4

1B)
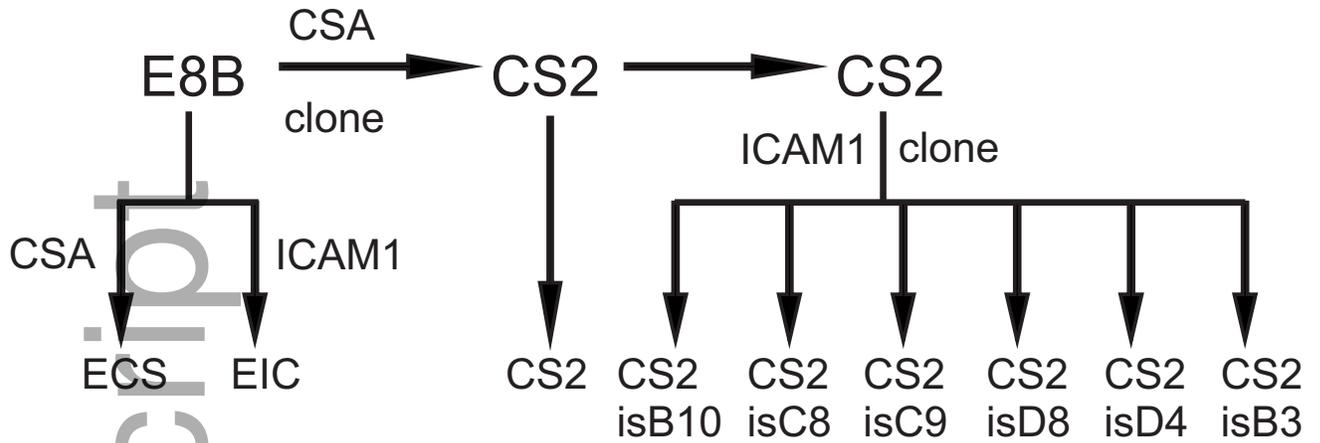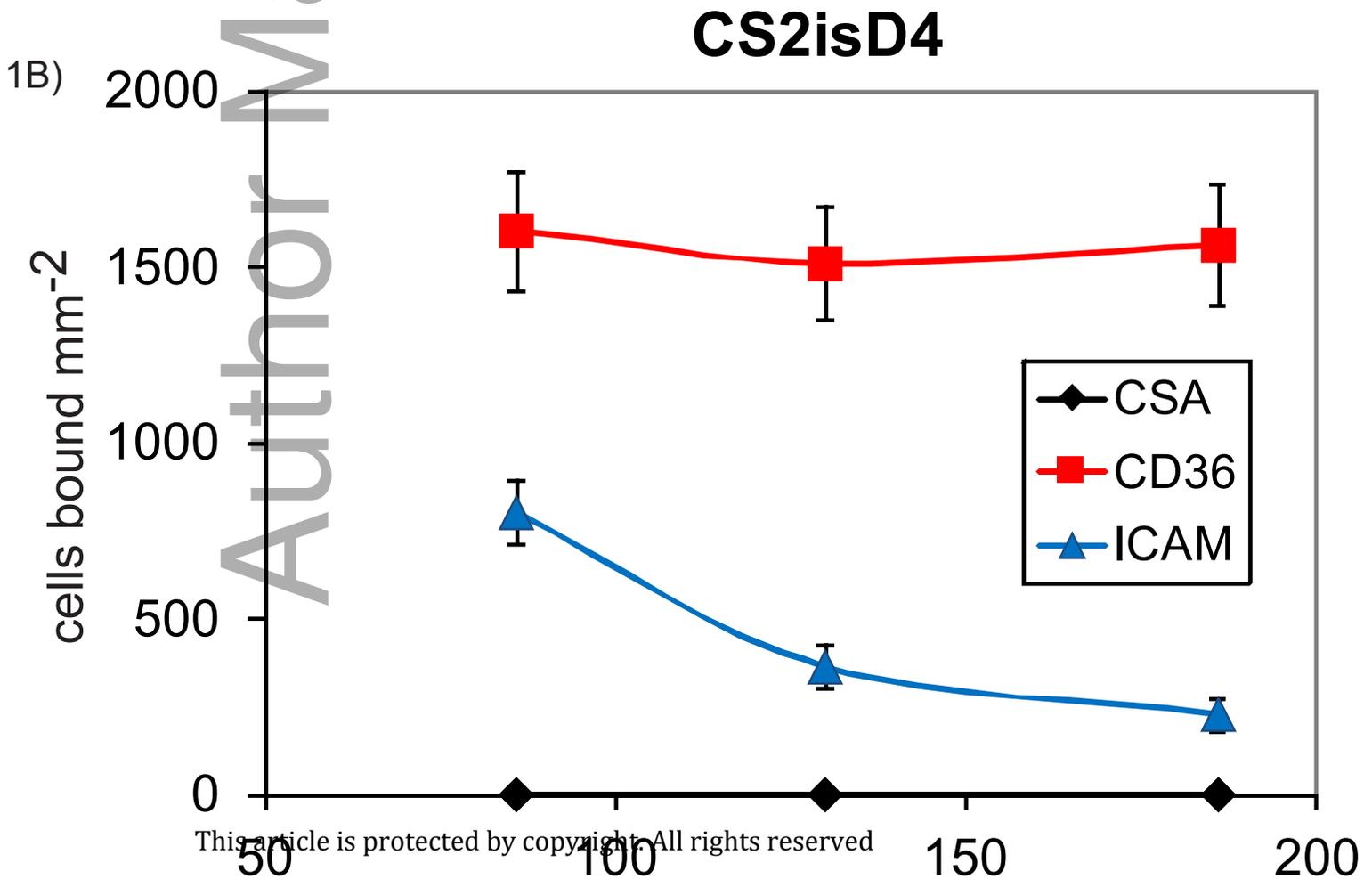
2A)

LH telomere   *It4var44*   *var2csa*   *acyl-coA synthetase 7 (ACS7)*

var2csa exon 1 probe

acyl-coA synthetase 7 probe

exon1   exon2   exon2   exon1

*Btg*I   *Cla*I   *Btg*I   *Cla*I   *Bsg*I *Acl*I   *Sca*I *Eco*RI   *Acl*I   *Bsg*I *Eco*RV   *Acl*I

*Btg*I 9,498 bp

*Cla*I 6.816 bp

*Eco*RI *Eco*RV 5,900 bp

*Sca*I *Eco*RV 6986 bp

*Bsg*I 7,910 bp

*Acl*I 5,010 bp   *Acl*I 12,546 bp

B)

*Btg*I   *Cla*I

CS2   D4   C8   CS2   D4   C8

bp

23130 —

9416 —

6557 —

4361 —

2322 —

2027 —

C)

CS2   C8   D4

*Eco*RI/*Eco*RV *Sca*I/*Eco*RV *Bsg*I *Acl*I   *Eco*RI/*Eco*RV *Sca*I/*Eco*RV *Bsg*I *Acl*I   *Eco*RI/*Eco*RV *Sca*I/*Eco*RV *Bsg*I *Acl*I

bp

23130 —

9416 —

6557 —

4361 —

2322 —

2027 —

564 —

CS2 compared to EIC and ECS

log fold change

4

2

0

-2

-4

1e-01   1e+01   1e+03   1e+05

A)

CS2 | ECS | EIC

var46
var44
var4/VAR2CSA5'
var4/VAR2CSA3'

var16/ICAM1
var14/A4var
var11
var10

relative expression

B)

3C | 3D7-icam

var2csa
PFL0020w

relative expression

febs_13967_f4.eps

A)

var2csa

rep20

merge
with
DAPI

not colocalised   colocalised

B)

A)

	IFA	*var2csa* RNA FISH	merge DAPI	Fluorescence profile (x axis is distance in μm)

H3K9me3

PfH2B.Z

PfH2A.Z

H3K9ac

H3K4me3

H3K4me1

H3K27ac

B)

IFA	*var2csa* RNA FISH	merge DAPI

H3K9me3

H3K27ac

A)

B)

C)

A)

H3K27ac

3C
var2csa-on
PFL0020w-off

3D7
var2csa-off
PFL0020w-off

3D7-icam
var2csa-off
PFL0020w-on

var2csa   *PFL0020w*

H3K4me1

var2csa   *PFL0020w*

B)

Rings

Schizonts

A)

H3K27ac ChIP

Enrichment ratio H3K27ac/H3

- rings
- schizonts

hsp70 utr, hsp70 cds, etramp 10.1 5'utr, etramp 10.1 cds, sbp1 5'utr, sbp1 cds, rex 5'utr, rex cds, eba-175 5'utr, eba-175 cds, casein kinase I 5'utr, casein kinase I cds, actin 1 5'utr, actin 1 cds, msp2 5'utr, msp2 cds, csp 5'utr, csp cds, trap 5'utr, trap cds, slarp 5'utr, slarp cds

B)

expression

fpkm

- rings
- schizonts

hsp70, tramp 10.1, sbp1, rex, eba-175, casein kinase I, actin 1, msp2, csp, trap, slarp

Author/s:
Duffy, MF; Tang, J; Sumardy, F; Nguyen, HHT; Selvarajah, SA; Josling, GA; Day, KP; Petter, M; Brown, GV

Title:
Activation and clustering of a Plasmodium falciparum var gene are affected by subtelomeric sequences

Date:
2017-01-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/292217

File Description:
Accepted version