Heterochromatin *de novo* formation and maintenance in *P. falciparum*

Alba Pérez-Cantero[1,2,#], Oriol Llorà-Batlle[1,2,#], Íngrid Peláez-Conde[1,2], César Martínez-Guardiola[1,2] and Alfred Cortés[1,2,3,*]

[1] ISGlobal, Barcelona, Catalonia, Spain
[2] Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona (UB), Barcelona, Catalonia, Spain
[3] ICREA, Barcelona, Catalonia, Spain

[#] Equal contribution as first authors

* Correspondence: alfred.cortes@isglobal.org (A.C.)

ABSTRACT

In the malaria parasite *Plasmodium falciparum,* the expression of many genes is regulated by heterochromatin (HC) based on the histone mark tri-methylation of histone H3 lysine 9 (H3K9me3). HC assembly involves three distinct steps: *de novo* nucleation, spreading and maintenance. Nucleation, which consists in formation of HC in a previously euchromatic region, determines at which specific regions of the genome HC occurs, but this process is not well understood in malaria parasites. Here we investigated the DNA sequence *cis* determinants of HC nucleation in *P. falciparum*, using a screening approach based on integration of fragments from different heterochromatic genes into an euchromatic locus, followed by chromatin immunoprecipitation (ChIP). We found that fragments of *var* gene upstream regions nucleated HC efficiently, whereas fragments from the *pfap2-g* upstream region or from the *mspdbl2* locus did not nucleate HC. Fragments from the beginning of the coding sequence (CDS) of *pfap2-g* nucleated HC with low efficiency, as evidence by nucleation requiring long fragments of ~2 Kb and occurring only in a fraction of the parasites. These results demonstrate that the primary DNA sequence is a main determinant of HC nucleation in *P. falciparum*. We also studied HC maintenance at the *pfap2-g* locus, which demonstrated that specific parts of the upstream region, different from the regions competent for HC nucleation, are required for maintenance. Together, our results provide initial insight into how HC is directed to specific loci and maintained in *P. falciparum*.

INTRODUCTION

In the nucleus of eukaryotic cells, DNA is packed as chromatin, a nucleoprotein complex formed mainly by DNA and histone proteins. Chromatin modulates the accessibility of the transcription machinery to the DNA, which constitutes one of the main mechanisms for the regulation of gene expression. Chromatin can be broadly classified into euchromatin, which is accessible and permissive to transcription, or heterochromatin (HC), which is less accessible and generally prevents transcription. In addition to its major role in the regulation of gene expression, HC is also important for genome stability, silencing of transposable elements and centromere function, among several other roles. A key property of HC is that it can be inherited by epigenetic mechanisms during cell division, such that different patterns of HC distribution in genetically identical cells are inherited in *cis* across multiple cell generations. This makes HC a truly epigenetic trait [1, 2].

HC can be constitutive or facultative. Constitutive HC typically occurs in repeat-rich regions of the genome, such as subtelomeric and pericentromeric repeats, which are in a permanently repressive state. Facultative HC is more dynamic and typically affects gene-rich regions that can be in either an euchromatic or a heterochromatic state in different cells with the same genome [1-3]. While HC occurs in different varieties, its formation always involves three distinct steps: nucleation, spreading and maintenance. Post-translational modification of histone tails, mainly methylation of specific lysine residues, lies at the basis of all steps of HC assembly. The molecular machinery responsible for HC formation includes enzymes that modify histone tails, typically referred to as writers or erasers, and enzymes that recognize these modifications, known as readers. The coupling of writer and reader protein domains, together with protein-protein interactions that mediate the recruitment of other regulatory factors, underlies all steps of HC formation [1, 2, 4].

The first step for HC assembly is *de novo* nucleation, which consists in converting a region that was previously euchromatic into HC. This step determines at which

specific regions of the genome HC is assembled. Multiple HC nucleation mechanisms have been described, which can be broadly divided between RNA-dependent mechanisms and those based on recognition of specific DNA sequences by DNA binding factors. The former, which involve transcripts generated at the site of HC nucleation to recruit chromatin modifying activities, can be divided in RNAi-dependent or RNAi-independent mechanisms [1, 2]. Regarding mechanisms based on DNA binding proteins, examples include ADNP or Pax family transcription factors that recognize motifs within repetitive DNA [5, 6], ATF family proteins [7, 8] and ATRX or KRAB zinc finger proteins that recognize transposable elements [9, 10].

After *de novo* nucleation, HC can spread from the nucleation sites to adjacent regions in a sequence-independent manner. Full spreading, which is slow and may require several cell generations, is restricted by limiting amounts of HC factors and by barrier elements (also known as barrier insulators or boundary elements). The latter prevent HC spreading into regions of the genome that need to be accessible [1, 11, 12]. Finally, inheritance of HC patterns as an epigenetic trait requires maintenance of HC during cell division. Clonal inheritance of HC in *cis* occurs in a self-templated manner, such that the epigenetic memory is stored in HC itself. The molecular mechanism of inheritance involves distribution of nucleosomes with HC-specific histone marks between the sister chromatids during DNA replication. These nucleosomes, located at the same original genome position, are used as templates to add the same marks in the histones of the newly formed nucleosomes [1, 2].

The most conserved type of HC across eukaryotic evolution, from yeast to humans, is based on the histone mark di- or tri-methylation of histone H3 lysine 9 (H3K9me2 or H3K9me3). Constitutive HC at repetitive regions is almost invariably associated with this histone mark, whereas facultative HC affecting protein-coding genes can be based on this mark [13, 14] or on methylation of lysine 27 (H3K27me3). In particular, facultative HC associated with development of multicellular organisms is often based on H3K27me3 and Polycomb-group proteins [1, 3]. Among the key players for the establishment of H3K9me2/3-based HC are lysine methyl

transferases such as Suv39 (Clr4 in yeast), which contain a SET domain that mediates methylation of H3K9 and a chromo-domain that recognizes methylated H3K9. The presence of the two domains in the same protein facilitates positive feedback loops needed for HC nucleation, spreading and maintenance, illustrating the importance of coupling reader and writer domains in HC biology. Another key player for the assembly of H3K9me2/3-based HC is heterochromatin protein 1 (HP1), which binds H3K9me2/3 via its chromo-domain and contributes to HC structure through dimerization and recruitment of many enzymatic activities, including histone deacetylases [1-4, 13].

In the malaria parasite *P. falciparum*, both constitutive and facultative HC are based on H3K9me3. Constitutive HC is present in subtelomeric repeats, but not in pericentromeric regions. Facultative HC occupies clonally variant gene (CVG) loci, which are located in subtelomeric regions and in a few chromosome-internal islands [15-19]. At CVGs, chromatin can adopt either a transcriptionally permissive (euchromatin) or repressive (HC) conformation. Once established, both states can be stably propagated for many generations (bistable chromatin), with switches between the two states occurring infrequently [16, 20-23]. This chromatin dynamics results in variant expression of CVGs, which play a fundamental role in host-parasite interactions by mediating processes such as immune evasion, solute transport, erythrocyte invasion, erythrocyte remodeling or sexual conversion, among others [19, 24, 25]. Within an isogenic parasite population, differences in CVG expression generate transcriptional and phenotypic diversity, which constitutes the basis of a 'bet-hedging' adaptive strategy [19, 26], where cell heterogeneity increases the chances of survival of the population in front of unexpected environmental changes [27, 28]. In addition to stage-independent variation between individual parasites, the distribution of HC changes during transmission stages [15, 29-31], whereas it remains constant during all asexual blood stages.

The putative *P. falciparum* H3K9 methyltransferase is the SET domain-containing protein SET3 [18, 32], although it has not been characterized in detail. Assembly of

H3K9me3-based HC in *P. falciparum* also requires the evolutionarily conserved HP1 [15, 17, 18, 33]. However, how HC forms only at some specific loci in the *P. falciparum* genome, resulting in variant expression of the genes affected, and does not form at other sites, still remains unknown. *P. falciparum* does not have a functional RNAi pathway [34], which suggests that HC nucleation uses an RNAi-independent mechanism in malaria parasites. Several DNA binding factors of the ApiAP2 family [35] are associated with HC [36, 37], but none of them has been directly involved in the formation of HC at specific genomic sites.

Here, we aimed to gain initial insight on the mechanism that directs HC formation to specific parts of the genome in malaria parasites. To test if the primary DNA sequence is a major determinant of *de novo* HC nucleation in *P. falciparum* and to identify putative DNA regions with an intrinsic nucleation capacity, we assessed the ability of fragments of different genes to trigger HC formation when ectopically integrated in an euchromatic region of the genome. Our primary analysis used fragments from the *pfap2-g* gene [38], which is located in an internal HC island [17, 18]. This gene encodes the master regulator of sexual conversion, *i.e.*, the conversion of asexual blood stage parasites into transmissible sexual forms called gametocytes. We also tested the ability to nucleate HC of fragments of another CVG located in a HC island, *mspdbl2* [17, 18, 39, 40], and from several *var* genes, located in large HC domains. The latter are involved in antigenic variation and pathogenesis [41]. Furthermore, we analyzed the role of specific *pfap2-g* sequences in HC maintenance by deleting different parts of this locus.

RESULTS

Identification of *pfap2-g* transcription start sites (TSSs)

To guide the design of the experiments to assess HC nucleation, ensuring that fragments containing the *pfap2-g* TSSs were included, we first mapped the TSSs of this gene using 5’ Rapid Amplification of cDNA Ends (5’RACE). This was performed with the E5 line [38] (a 3D7 subclone with a high sexual conversion rate) and the

transgenic E5-PfAP2-G-DD line, which has a destabilization domain appended to PfAP2-G and shows unusually high sexual conversion rates when the stabilizing ligand Shield-1 (Shld1) is added [38, 42]. In both parasite lines, we consistently identified two TSSs blocks in the *pfap2-g* upstream region, Block 1 and Block 2, located ~1,900 and ~1,600 bp from the start codon, respectively (Fig. 1). Of note, previous genome-wide studies on transcription initiation did not identify any TSSs for *pfap2-g* [43, 44], likely because the gene was only expressed in a small fraction of the cells in the parasite lines used.

To validate the predicted TSSs identified by 5'RACE, we performed a reverse transcriptase coupled to quantitative PCR (RT-qPCR) analysis using primers located upstream, between or downstream of the two blocks. This analysis confirmed that essentially no transcripts contained sequences upstream of Block 1 and the majority of transcripts contained sequences located between the two blocks (Fig. S1). The irregular pattern observed in the transcript levels measured at different positions may be explained by the occurrence of different mRNA species or by technical limitations associated with the extreme AT-richness of this region.

Identification of regions of the *pfap2-g* locus with capacity to nucleate HC *de novo*

To identify specific regions of the *pfap2-g* locus able to nucleate HC *de novo*, we designed a screening approach based on the integration of *pfap2-g* fragments into the euchromatic PF3D7_1144400 locus [16] (Fig. S2), which encodes a gene dispensable for asexual blood stages [45]. The resulting transgenic lines were analyzed using H3K9me3 chromatin immunoprecipitation (ChIP) coupled to quantitative PCR (ChIP-qPCR) to evaluate HC levels at the site of integration. The analysis of each transgenic line also included heterochromatic and euchromatic genes used as ChIP-qPCR positive and negative controls, respectively (Fig. S2).

To detect HC formation within the ectopically integrated fragments without interference from the endogenous locus, we first deleted the upstream regulatory

region and part of the coding sequence (CDS) of the endogenous *pfap2-g* gene in the 3D7-A subclone 1.2B. This subclone does not form gametocytes due to a non-sense mutation in *pfgdv-1* [46], an upstream activator of *pfap2-g* [47]. The resulting parasite line, which we termed Δ5'ap2-g, was subsequently edited to integrate the DNA fragments under analysis into the PF3D7_1144400 locus (Fig. 2A).

We analyzed by ChIP-qPCR five transgenic lines in which we integrated overlapping ~1 kb fragments (F0-F4) of the *pfap2-g* upstream region and beginning of the CDS (Fig. 2B), which are heterochromatic in the endogenous *pfap2-g* locus [15-18]. Fragments from the CDS (F0 and F1) and, to a lesser extent, from the proximal upstream region (F2) promoted HC formation, although with a lower coverage than in the ChIP-qPCR positive controls (Fig. 2C). HC was detected within the fragment itself and in the adjacent regions of the PF3D7_1144400 locus, indicating that the newly formed HC was able to spread from the nucleation site into flanking regions. In contrast, fragments from the distal part of the *pfap2-g* upstream region (F3 and F4) did not promote HC formation, as evidenced by similar H3K9me3 levels as in the ChIP-qPCR negative controls (Fig. 2C). Additionally, integration of a fragment of similar size from the upstream region of the euchromatic gene *ama1* (Fig. S2) did not promote HC formation, whereas integration of a fragment from the heterochromatic *var* gene PF3D7_1240300 (Fig. S2) resulted in prominent HC formation, in this case with a coverage similar to the ChIP-qPCR positive controls (Fig. 2B-C). To further dissect the sequences that can nucleate HC, we generated additional transgenic lines with shorter fragments of the *pfap2-g* F0 and F1 regions integrated, but clear HC formation was not observed for any of them (Fig. S3).

Unexpectedly, qPCR and Southern blot analysis of genomic DNA (gDNA) of the transgenic lines revealed integration of multiple copies of the fragment in all cases except for the lines with the *pfap2-g* F4 or *var* gene ~1 kb fragments (Fig. 2D, S4) and two of the lines with shorter fragments (Fig. S3). Although the qPCR results are compatible with either integration of concatemers (containing multiple copies of the fragments) or presence of episomal copies of the donor plasmid, the complex pattern

of bands observed in the Southern Blot analysis (Fig. S4) suggests integration of concatemers. While we cannot exclude the possibility that episomal copies of the plasmids were also present and accounted for part of the HC signal detected, HC formation at the F0 and F1 fragments integrated in the genome is demonstrated by the detection of H3K9me3 with the external primer pairs 5'ext. and 3'ext., which recognize sequences adjacent to the site of integration that are not present in the transfection plasmids (Fig. 2A,C). On the other hand, we cannot rule out the possibility that integration of concatemers containing multiple copies of the fragments affects their ability to nucleate HC formation, although the different ability of different fragments to form HC is suggestive of sequence specificity.

Long fragments are needed for efficient HC nucleation by *pfap2-g* sequences

To characterize the potential effect of integration of multiple copies of the fragments on *de novo* HC nucleation, the *pfap2-g* fragments F0 and F1 were integrated at the PF3D7_1144400 locus using donor plasmids that carried the *yfcu* negative selection marker. This enabled elimination of parasites carrying either episomal plasmid or concatemer integration. ChIP-qPCR analysis of the new transgenic lines containing single copies of fragments F0 or F1 revealed absence of HC at the integrated fragments (Fig. 3A-B).

Formation of HC only in parasite lines with multiple copies of F0 or F1 may be explained by the larger total size of putative HC-promoting DNA or by sequence repetitiveness (occurrence of the same sequence more than one time), as repetitiveness itself may promote HC formation. In *Drosophila* and other organisms, presence of multiple copies of a transgene are sufficient to trigger HC-based silencing [48, 49]. To distinguish between these possibilities, we assessed HC formation at the PF3D7_1144400 locus after integration of a single copy of longer fragments (~2 kb) from the *pfap2-g* gene (Fig. 3C-E). This revealed formation of HC with a single copy of the *pfap2-g* F1+F0 fragment, including the 5' end of the CDS and a small part of the upstream region, albeit at lower levels than in ChIP-qPCR positive controls. In contrast, no HC was detected after integration of the F2+F1

pfap2-g fragment or a ~2 kb fragment of the euchromatic *ama1* gene. These results indicate that single copies of fragments of the *pfap2-g* locus can nucleate *de novo* HC formation in a previously euchromatic region, without requiring repetitiveness.

HC nucleation by *pfap2-g* fragments occurs in only a subset of parasites

H3K9me3 coverage at the HC nucleation-positive *pfap2-g* fragments integrated at the PF3D7_1144400 locus was lower than in ChIP-qPCR positive HC controls (Fig. 2C, 3D). To determine whether this reflects that HC is formed in only a subset of the parasites, we generated subclones of the transgenic line carrying the *pfap2-g* F1 fragment (multiple copies). In two of the subclones (5E and 11A), no HC was observed at the site of fragment integration, whereas in two other subclones (3E and 9B) HC formation was apparent (Fig. 4A). HC-positive and -negative subclones contained a similar number of copies of the integrated fragment (Fig. 4B). Likewise, analysis of four subclones of the transgenic line with a single copy of the *pfap2-g* F1+F0 fragment integrated showed that HC formed in two of the subclones at levels similar to the ChIP-qPCR positive controls, but was absent in the other two (Fig. 4C-D). These results confirm that the intermediate levels of H3K9me3 observed in transgenic lines with HC-positive *pfap2-g* fragments reflect population heterogeneity rather than partial H3K9me3 occupancy in individual parasites.

Additionally, to map the extension of the HC formed at the PF3D7_1144400 locus after integration of nucleation-positive *pfap2-g* fragments, we assessed the distribution of H3K9me3 using ChIP-seq. HC spanned over the full PF3D7_1144400 CDS, such that the limits of the H3K9me3-enriched domain roughly coincided with the beginning and end of the gene (Fig. 4E-F). In the *pfap2-g* F1 fragment 3E subclone, we also analyzed the distribution of H3K9ac, which was severely depleted throughout the heterochromatic region (Fig. 4E).

HC formation by sequences from other CVGs

Next, we tested the ability of sequences from other CVGs to nucleate HC ectopically. Following the same approach as for *pfap2-g*, we integrated at the euchromatic locus

PF3D7_1144400 fragments from the *mspdbl2* locus, which is silenced in the majority of parasites and located in a small internal HC island [16, 18, 39, 40] (Fig. S2), similar to *pfap2-g*, and from three *var* gene loci. These included two subtelomeric *var* genes, PF3D7_0733000 and PF3D7_1041300, and a *var* gene located in a large internal HC island, PF3D7_1240300. We used plasmids containing the *yfcu* negative selective marker to enable selection for single copy integrations.

None of the *mspdbl2* fragments tested, including ~1 kb and ~2 kb fragments from the coding and upstream sequences, formed HC at the ectopic site (Fig. 5A-B). In contrast, all ~1 kb fragments of *var* genes (including upstream regions and in some cases also the beginning of the CDS) robustly formed HC, generally at levels similar to the ChIP-qPCR positive control (Fig. 5C-D). All transgenic lines with *mspdbl2* or *var* fragments had a single copy of the fragment integrated, with the exception of the line with the PF3D7_1041300 prom+CDS fragment (Fig. 5E). However, subclones of this transgenic line with a single copy of the fragment also nucleated HC efficiently, and a small deletion occurring in some parasites with the PF3D7_1240300 fragment did not affect HC nucleation (Fig S5). Together, these results show that *var* gene sequences can nucleate HC efficiently, whereas *pfap2-g* sequences can nucleate HC with lower efficiency and *mspdbl2* sequences cannot.

Analysis of the role of different regions of *pfap2-g* in HC maintenance

The H3K9me3 ChIP-seq analysis of Δ5'ap2-g-derived transgenic lines, in which the upstream region and part of the CDS of the *pfap2-g* locus were deleted (Fig. 2A), revealed that HC was completely absent from the remaining part of the endogenous *pfap2-g* locus (Fig. 6A). This unexpected observation indicates that the region deleted is necessary to maintain HC at the full *pfap2-g* locus. In contrast, ChIP-seq analysis of the E5-Δap2-g line, in which most of the *pfap2-g* CDS was deleted [38], maintained HC in the remaining part of the locus (Fig. 6A). These results suggest that specific parts of the *pfap2-g* locus are needed for HC maintenance, which prompted us to investigate the determinants of HC maintenance in more detail.

We systematically tested the impact of deleting different regions of the *pfap2-g* locus on the presence of HC at the rest of the locus. The deleted regions were replaced by a non-functional *gfp* gene using CRISPR/Cas9 and HC levels were determined using H3K9me3 ChIP-qPCR with a collection of primers at different positions of the locus (Fig. 6B). The ΔCDS transgenic line, in which the deleted region spanned most of the *pfap2-g* CDS, had H3K9me3 levels at the rest of the locus similar to the wild type 1.2B line (Fig. 6C-D), consistent with the ChIP-seq results for the E5-Δap2-g line (Fig. 6A). In contrast, deletion of the *pfap2-g* upstream region (ΔU) resulted in complete HC loss from the full locus (Fig. 6D), similar to the Δ5'ap2-g-derived lines (Fig. 6A). Together, these results confirm that the *pfap2-g* upstream region but not the CDS is essential for the maintenance of HC.

To narrow down the region responsible for HC maintenance, we generated the ΔU5' and ΔU3' lines, in which we removed the distal or the proximal part of the *pfap2-g* upstream region, respectively. While the distal part was not required for HC maintenance at the rest of the locus, deletion of the proximal part (ΔU3') affected HC levels. One ΔU3' subclone had HC at the rest of the locus, whereas another subclone completely lost it (Fig. 6E). These results suggest that the proximal part of the *pfap2-g* upstream region plays an important role for HC maintenance, although it is not absolutely essential. We also generated and analyzed several additional transgenic lines, in which different parts of the locus were deleted. The analysis of these additional lines supported the conclusion that the proximal upstream region is important for HC maintenance whereas the other regions are dispensable (Fig. 7, S6). Of note, the majority of these transgenic lines had a single copy of the *gfp* gene integrated. In a few cases, plasmid concatemers were integrated in some of the parasites (ΔU5', ΔF1 and ΔF2-3), but analysis of subclones showed that the increased separation between the regions upstream and downstream of the deletion in lines with concatemer integration did not impair HC maintenance (Fig. 6-7, S6).

Reestablishment of HC at the *pfap2-g* locus is inefficient

Our experiments based on ectopic fragment integration revealed that *pfap2-g* fragments promote *de novo* HC nucleation inefficiently. This made us predict that if a parasite line had the endogenous *pfap2-g* locus in an euchromatic state, it would take several generations for HC to form again. To test this hypothesis, we used the E5-AP2-G-DD parasite line [38], in which the PfAP2-G protein is fused to an FKBP destabilization domain (DD) that in the absence of the Shld1 ligand drives protein degradation. In this parasite line, when cultured without Shld1, absence of HC at the *pfap2-g* locus would not result in massive sexual conversion and therefore would not pose a fitness cost. Indeed, the E5-AP2-G-DD line shows an unusually high sexual conversion rate when Shld1 is added [38, 42], consistent with accumulation of parasites with *pfap2-g* in an euchromatic state during culture without Shld1.

We observed sexual conversion rates of ~45% after a first PfAP2-G stabilization with Shld1 for one cycle, but subsequent stabilization two cycles later resulted in a much lower conversion rate (Fig. 8A). After induction with Shld1 for one cycle, culturing in the absence of Shld1 for up to 60 generations led to progressive, partial recovery of high sexual conversion rates when Shld1 was added again (Fig. S7). Together, these results suggest that in cultures always maintained without Shld1, parasites with the *pfap2-g* locus in an euchromatic state progressively accumulated, because this did not pose a fitness cost, and these parasites underwent sexual conversion upon Shld1 addition. The remaining parasites had HC at the *pfap2-g* locus and a regular conversion rate. Consistent with this model, subclones of the E5-AP2-G-DD line (obtained from cultures not exposed to Shld1) had dramatically different sexual conversion rates when Shld1 was added, ranging from <5% to >50% (Fig. 8B).

ChIP-seq analysis of cultures maintained without Shld1 revealed a lower H3K9me3 coverage in the parental E5-AP2-G-DD line and a subclone with high sexual conversion compared with a subclone with low conversion, consistent with the gene being in an euchromatic state in the approximately 50% of parasites that convert upon Shld1 addition (Fig. 8C-D). No subclone was identified with 100% conversion upon Shld1 addition, suggesting that in the subclone originated from a single

parasite with euchromatic *pfap2-g*, HC formed in some parasites during the time between subcloning and ChIP-seq analysis, albeit at low speed consistent with inefficient nucleation. The reduction in H3K9me3 coverage in the high conversion subclone was homogeneous across the full locus, in contrast with activation of the gene during natural sexual conversion that appears to involve HC depletion only at the putative promoter region [30, 50] . Together, these experiments suggest that the *pfap2-g* locus can be in an euchromatic state when the protein is non-functional, the euchromatic or heterochromatic state can be transmitted as for other CVGs, and spontaneous reestablishment of HC is inefficient and requires multiple generations.

DISCUSSION

Transcriptional regulation of key biological processes in *P. falciparum* relies on transitions between HC and euchromatin at loci encoding CVGs [24, 25]. However, the molecular determinants for HC formation exclusively in some regions of the genome (*i.e.*, those containing subtelomeric repeats or CVGs) and not others remain largely unexplored. Here, we systematically analyzed how DNA sequence determines the ability of a fragment to nucleate HC *de novo* when ectopically integrated in an euchromatic chromosomal site. We found large differences between fragments of similar size but different sequence in their ability to nucleate HC, which clearly indicates that the primary DNA sequence is a major determinant of where in the genome HC is formed. We also found that, at least for *pfap2-g* sequences, a large total size of nucleation-competent DNA is needed for efficient HC nucleation. Furthermore, maintenance of HC also depends on *cis* sequence elements, as only some specific regions of a HC domain were needed to stably maintain pre-existing HC at the rest of the domain. Of note, the regions needed to maintain pre-existing HC are different from those capable to nucleate HC *de novo*, consistent with observations in other organisms showing that nucleation and maintenance are separable processes that depend on different factors [1, 2].

We analyzed the ability to nucleate HC of fragments from an euchromatic gene and from several CVGs located either in large HC blocks (*var* genes) or in small HC islands (*pfap2-g* and *mspdbl2*) [15-18], with a main focus on *pfap2-g*, which encodes the master regulator of sexual conversion [38]. We reasoned that for genes located within large HC blocks, presence of HC may simply reflect HC expansion from neighbor regions or a favorable HC-prone chromosomal context at the nuclear periphery, whereas in small internal chromatin islands the ability to nucleate HC may be strictly conferred by the primary sequence of the gene. However, all *var* promoter sequences tested nucleated HC efficiently, which indicates that these sequences have an intrinsic capacity to nucleate HC *de novo*, even when placed in an ectopic, previously euchromatic region separated from the subtelomeric or central *var* clusters. In contrast, we found that fragments from the beginning of the *pfap2-g* CDS were only able to nucleate HC inefficiently, and fragments from *mspdbl2* and from the *pfap2-g* upstream region were unable to nucleate. The low efficiency of HC nucleation by *pfap2-g* CDS sequences was reflected in nucleation occurring only in a fraction of the parasites and requiring ~2 kb fragments or multiple copies of ~1 kb fragments, whereas ~1 kb *var* fragments robustly nucleated HC in the vast majority of parasites. These results argue against nucleation at the *pfap2-g* locus being a deterministic event in which the presence of specific sequences always results in formation of HC, and rather suggest that, at least for this locus, nucleation is a probabilistic event. Of note, the contribution of different regions to HC maintenance also appears to be probabilistic, as suggested by differences in HC maintenance between genetically identical subclones. Probabilistic events are a common theme in HC biology, as realized many years ago with the variant expression associated with position effect variegation (PEV) in *Drosophila* [1].

A recent study reported that the *pfap2-g* upstream region was able to form ectopic HC in the context of an artificial chromosome, albeit with an H3K9me3 coverage much lower than in endogenous heterochromatic genes [51]. While these results cannot be directly compared with our study because of the artificial chromosome environment and because different regions were tested, they support our conclusion

that *pfap2-g* sequences can nucleate ectopic HC formation, although inefficiently. The ability of *P. falciparum* sequences to recapitulate the epigenetic makeup of their endogenous locus when integrated at an ectopic locus was also reported for H3K48ac [52]. However, this likely operates by a completely different mechanism because, unlike HC, this modification may be a direct consequence of active transcription [23] (associated with the promoter activity of the underlying sequence).

The inefficient *de novo* nucleation of HC by *pfap2-g* fragments and inability to nucleate by *mspdbl2* fragments raises the intriguing possibility that *de novo* nucleation at these (and other) loci may be a very rare event *in vivo*, at least during asexual blood stages. In line with this view, comparative analysis of HC distribution between parasite lines that differ in the expression of many CVGs revealed that the chromatin transitions that result in activation or silencing of these genes typically involve expansion or retraction of pre-existing HC domains, rather than *de novo* formation or complete dismantling [16]. Furthermore, we show that in a parasite line expressing PfAP2-G fused to a DD domain, in which activation of *pfap2-g* expression does not impair asexual proliferation, re-assembly of HC in parasites that lost HC at the *pfap2-g* locus requires multiple generations and does not occur in all parasites. This again indicates that HC formation by *pfap2-g* sequences is not deterministic and the HC state of this gene typically depends on inheritance from previous generations rather than *de novo* formation. However, *de novo* HC formation may occur efficiently at other stages. Epigenetic HC patterns are reset during transmission stages [53-57], and it remains unclear if the reset involves complete dismantling of HC, which would make *de novo* HC nucleation essential, or some parts of each HC domain always persist, dispensing the need for *de novo* nucleation. Altogether, existing data supports a scenario in which HC domains are perpetuated for many generations by inheritance through cell division, and *de novo* nucleation is either not regularly required or required only after transmission.

Although here we did not directly aim to investigate HC spreading or to identify barrier elements that limit spreading, we observed clear evidence for HC spreading

in some of the transgenic lines. This affected the *gfp* sequence used to replace fragments of the *pfap2-g* locus, which was transfected as naked DNA and after integration became heterochromatic, and the PF3D7_1144400 locus, which was euchromatic but became heterochromatic after integration of fragments competent for HC nucleation. Together with previous reports for *clag3* or *var* genes [58, 59], these results indicate that in *P. falciparum* HC can spread into adjacent regions, which is an evolutionarily conserved property of HC [1, 2]. Spreading into the PF3D7_1144400 locus had similar limits in different transgenic parasite lines, which is suggestive of the occurrence of barrier elements [11] at these positions. In fact, the limits of naturally-occurring HC domains associated with the active or silenced state of essentially all CVGs are well-defined and conserved in different parasite lines [15, 16, 20, 30], suggesting that barrier elements play a prominent role in the organization of *P. falciparum* HC. How the boundaries of HC domains are established in malaria parasites has not been characterized, but a recent preprint proposed that the 3'end of *pfap2-g* acts as a barrier element [60].

A limitation of our study is that we could not test the ability to nucleate HC of a larger number of fragments covering the entire loci of the genes studied and additional heterochromatic and control euchromatic genes, because it was not feasible to use our screening approach in a high-throughput manner. This is attributable to the well-known difficulties generating *P. falciparum* transgenic lines, especially when using extremely AT-rich non-coding regions, and also to the frequent integration of concatemers, which was found to be a confounding factor for our HC nucleation experiments. Moreover, the low efficiency and non-deterministic nature of HC nucleation by some of the fragments further complicated the analysis, as it implied having to insert very long DNA fragments and to generate and analyze multiple subclones for many of the transgenic lines. The non-feasibility of testing fragments at a larger scale reduced the power to discover potential short DNA motifs with the intrinsic capacity to promote *de novo* HC nucleation, which would have paved the way for the identification of putative DNA binding factors or non-coding RNAs driving nucleation. However, the need of long fragment for efficient nucleation suggests that

a single specific short DNA motif able to drive HC nucleation may in fact not exist in *P. falciparum*. Thus, it is possible that multiple different binding sites for putative DNA binding factors or specific DNA physical properties over a long sequence stretch are needed for efficient nucleation. In spite of these limitations, our results demonstrate that the primary sequence of some specific DNA regions contains the information needed to seed HC formation at specific locations of the genome and different regions are needed to maintain preexisting HC domains.

METHODS

Parasite cultures

The 3D7 subclones E5 (derived from the 3D7-B stock) and 1.2B (derived from the 3D7-A stock) and the transgenic lines E5-PfAP2-G-DD and E5-Δap2-g have been previously described [38, 61]. Parasites were cultured in B+ erythrocytes at 3% hematocrit under standard conditions, with RPMI-1640-based culture medium supplemented with 0.5% Albumax II (Invitrogen). Routine synchronization was carried out using 5% D-sorbitol lysis to eliminate late asexual stages (trophozoites and schizonts). For the analysis of transcripts by RT-qPCR to validate TSSs, cultures were tightly synchronized by purification of mature schizonts using 63% Percoll (Sigma-Aldrich) gradients followed by sorbitol lysis 5 h later [62]. For the stabilization of PfAP2-G in the E5-AP2-G-DD line, 0.5 µM AquaShield-1 (Shld1, Cheminpharma) was added to cultures ~10 h after sorbitol synchronization (ring / early trophozoite stage). To measure sexual conversion rates, at the following cycle after stabilization cultures at the ring stage were treated with 50 mM N-acetylglucosamine (Sigma-Aldrich, A3286) for ≥4 days to eliminate asexual parasites. Light microscopy quantification of Giemsa-stained smears was used to determine sexual conversion rates, calculated as the gametocytemia at day ≥4 divided by the initial parasitemia before adding N-acetylglucosamine [42].

5’RACE and RT-qPCR analysis to map TSSs

542 TSSs were identified using the FirstChoice 5′ RLM-RACE kit (Thermo Fisher)
543 following the manufacturer's instructions, as previously described [20]. In brief, total
544 RNA from non-synchronized E5 and E5-PfAP2-G-DD (treated with Shld1) cultures
545 was extracted using the TRIzol (Invitrogen) method [62]. Next, RNA was treated with
546 calf intestinal alkaline phosphatase to remove free 5'-phosphate groups from non-
547 mRNA molecules. Then, samples were treated with the tobacco acid
548 pyrophosphatase to remove the 5'-cap structure of full-length mRNAs, allowing the
549 ligation of an RNA adapter with T4 RNA ligase. We used a redesigned RNA adapter
550 because the one provided with the kit resulted in artefacts that incorporated
551 exogenous sequences of another chromosome into the *pfap2-g* PCR-amplified
552 product (Table S1). cDNA synthesis was performed by reverse transcription using
553 random hexamers followed by nested PCR using *pfap2-g* specific primers and
554 primers specific for the RACE adapter (Table S1). Nested PCR was performed either
555 with the *LA* Taq DNA Polymerase (Takara), with 25 cycles of amplification for each
556 PCR reaction and an annealing temperature of 60ºC, or with the KAPA HiFi PCR Kit
557 (KAPA Biosystems, KK2103), with 20 cycles of amplification in the first PCR and 30
558 in the second, and an annealing temperature of 62ºC. PCR products were cloned
559 into the pCR-2.1 plasmid using the TA Cloning Kit (Invitrogen) and sequenced with
560 M13 forward and reverse primers.
561
562 For RT-qPCR validation of the TSSs, RNA was extracted using the TRIzol method
563 (Invitrogen), DNAse I-treated (Qiagen), purified using the RNeasy MinElute Cleanup
564 kit (Qiagen) and reverse-transcribed using the AMV Reverse Transcription Kit
565 (Promega) with a mixture of random primers and oligo (dT). Transcript abundance
566 was measured by real-time quantitative PCR (qPCR) in triplicate wells using the
567 standard curve method, with the Power SYBR Green Master Mix (Applied
568 Biosystems) and a 7900HT Fast Real-Time PCR System, as previously described
569 [62]. Transcript levels of the serine-tRNA ligase (*serrs*, PF3D7_0717700) were used
570 for normalization. All primers used for RT-qPCR analysis are described in Table S1.
571
572 **Plasmids**

The pL7-KO-ap2g plasmid (Fig. S8) used to delete the *pfap2-g* upstream region and part of the CDS (Δ5'ap2-g line) was derived from the pL6-egfp-yfcu plasmid [63]. The original HR2 was removed by digestion with *EcoR*I and *Nco*I, followed by treatment with T4 DNA polymerase (NEB) to generate blunt ends and religation with T4 DNA ligase (Roche). The *pfap2-g* HR1 (positions -3929 to -3521 bp relative to the start codon) was PCR amplified and cloned into *Not*I/*Pst*I sites, disrupting the *yfcu* cassette. The *pfap2-g* HR2 (positions +991 to +1973 bp) was amplified and cloned into *Spe*I/*Afl*II sites, replacing the original HR1. To add the sgRNA, two annealed oligonucleotides containing the guide sequence (recognizing a sequence located at the *pfap2-g* -3503 to -3522 bp position) were cloned into a *Btg*ZI site using the In-Fusion HD Cloning Kit (Clontech). Last, an incomplete *gfp* sequence (gfp*), lacking the start codon, was amplified and cloned between HR1 and HR2, removing the remaining part of the *yfcu* cassette, using *Pst*I/*Spe*I sites. To generate plasmid pDC2-Cas9-hDHFR-ap2g-3' (Fig. S8) expressing a second sgRNA in addition to Cas9 and the *hdhfr* selectable marker, a guide RNA covering positions +921 to +940 bp was cloned into a *Bbs*I site of plasmid pDC2-Cas9-hDHFR [64] using the InFusion system.

For the integration of the fragments into the PF3D7_1144400 locus, pL1144400-Fx plasmids ("x" indicates each different fragment; Fig S4A) were generated from the pL6-egfp-yfcu plasmid [63]. The original HR2 was removed as described above. The PF3D7_1144400 HR1 (positions +1144 to +1562 bp from the start codon) was amplified with primers incorporating *Nco*I+*Spe*I sites in the reverse primer and a *Sac*II site in the forward primer. The PF3D7_1144400 HR2 (positions +1646 to +2199 bp) was amplified incorporating *Spe*I and *Afl*II sites, and cloned together with HR1 into *Sac*II/*Afl*II sites, replacing the original HR1. The original *hdfhr* and sgRNA expression cassettes were removed using restriction sites *Aat*II and *Afl*II, blunt ending and religation. The *yfcu* cassette was removed using restriction sites *Not*I and *Sac*II, blunt ending and religation. However, for all fragments except ~1Kb fragments in Fig. 2 and shorter fragments (Fig S3), the selection cassette was reintroduced using the same restriction sites (pL1144400-Fx-yFCU plasmids) (Fig

S9). Fragments (Fx) from different genes were PCR amplified and cloned into *Nco*I/*Spe*I sites with the primers listed in Table S1. To generate the plasmid pDC2-Cas9-hDHFRyFCU-1144400 (Fig S4A, S9), used for transfections with pL1144400-Fx and pL1144400-Fx-yFCU plasmids, a guide RNA (covering positions +1566 to +1585 bp from the PF3D7_1144400 start codon) was cloned into a *Bbs*I site of plasmid pDC2-Cas9-hDHFRyFCU [65] using the InFusion system.

To generate the plasmid pL-ap2g_Del1_GFP_yfcu, used to obtain the ΔATG-1 line, the original HR1, *ama1* fragment and HR2 regions were removed from the pL11444000-ama1-yFCU plasmid (Fig. S9) by digestion with *Sac*II and *EcoR*I. The HR1 and HR2 were cloned by InFusion, introducing *Hind*III and *Spe*I restriction sites between them to clone the gfp* sequence by ligation (Fig S10A). For the ΔATG-2, ΔCDS, ΔU, ΔU5', ΔU3' and ΔF2-3'-A lines, HRs for each construct were amplified, introducing *Afl*II and *Hind*III restriction sites between HR1 and HR2, and cloned into *Sac*II/*Nco*I sites of a plasmid derived from pL6-eGFP-yFCU [63] using the InFusion system. Digestion of the plasmid with *Sac*II and *Nco*I removed the original HRs and the *hdhfr* sequence. Next, a different sgRNA for each transgenic line was cloned into a *BtgZ*I site. Finally, the gfp* sequence was cloned into the *Afl*II/*Hind*III sites between HR1 and HR2 by ligation, resulting in pL-7-ap2g_Delx_GFP_yfcu plasmids (where "x" refers to each fragment deleted) (Fig S10A). Finally, a second sgRNA used for each transgenic line was cloned into a *Bbs*I site of plasmid pDC2-Cas9-hDHFR [64], resulting in pDC2-Cas9-hDHFR-sgRNA-Delx plasmids (Fig S10A, Table S1).

For deletion ΔF1 and to generate a second transgenic line with the deletion ΔF2-3' (ΔF2-3'-B line), we followed an alternative three-plasmid strategy (Fig. S10B). Donor plasmids pUC19-KO-Fx were generated by simultaneously cloning the respective HR1 and HR2 for each construct and the *gfp** sequence into a *Bam*HI site of the pUC19 vector, using the InFusion system. Guides were cloned into the pDC2-Cas9-hDHFR plasmid as described above.

All guides were designed and selected using the EuPaGDT web-based tool [66]. PCR amplification from *P. falciparum* gDNA was performed using LA Taq DNA Polymerase (Takara). For plasmid cloning and amplification, we used *Escherichia coli* DH5α or Stellar Competent Cells (Clontech) for difficult cloning. Oligonucleotides were purchased from Integrated DNA Technologies (IDT). All primers and oligonucleotides are described in Table S1.

Generation of transgenic lines

Unless otherwise stated, transfections were performed according to standard procedures [67], electroporating cultures at the ring stage with a BioRad GenePulser Xcell electroporator. Starting ~15 h after transfection, cultures were selected for 4 days with 10 nM WR99210 (Jacobus Pharmaceuticals). When parasites were observed, for transfections with donor plasmids containing the *yfcu* negative selection marker, cultures where selected with 1 µM 5-fluorocytosine (5-FC; clinical grade Ancotil, Mylan N.V.) for 8 days, to remove parasites carrying episomal copies of the plasmids or integration of concatemers containing the full plasmid. Note that in the case of the *var* 1240300 line, cultures were analyzed without 5-FC selection (for comparison with transgenic lines generated with plasmids lacking the *yfcu* selection marker, Fig. 2), and after 5-FC selection (Fig. 5). Diagnostic PCR to assess correct edition was performed using the *LA* Taq DNA Polymerase (Takara), typically with primers external to the region edited. Diagnostic qPCR using the standard curve method was performed to assess the presence of plasmid regions not intended to be integrated and to determine the relative copy number of HRs and integrated fragments. Analysis by qPCR was performed in triplicate wells as described above for RT-qPCR. We used gDNA from the W4-2 (containing a single copy of the *hdhfr* gene) and the wt 3D7-A parasite lines [68] for the standard curve included in each plate for each primer pair. In all cases, copy number was determined relative to a single copy gene (*serrs*). The primers used for diagnostic PCR and qPCR are described in Table S1.

To obtain the Δ5'ap2-g line, 1.2B cultures were transfected with 60 µg of plasmid pDC2-Cas9-HDHFR-ap2g-3' and 12 µg of pL7-KO-ap2g linearized using a *Sca*I site located in the backbone of the plasmid. Subclones were obtained by limiting dilution and assessed by diagnostic PCR (with primers external from the edited region) and qPCR analysis to assess presence of plasmid sequences not intended to be integrated (*hdhfr* marker). The H11 subclone was selected for further work because it contained no episomal copy of the transfection plasmids or multiple integrations (Fig. S8).

For the integration of fragments at the PF3D7_1144400 locus, Δ5'ap2-g (H11 subclone) cultures were transfected with 60 µg of pDC2-Cas9-yFCU-hDHFR-1144400 and 12 µg of either pL1144400-Fx (Fig. S4) or pL1144400-Fx-yFCU (Fig. S9), linearized using the *Sca*I or the *Xmn*I restriction sites.

To generate transgenic lines with deletions at the *pfap2-g* locus, 1.2B cultures were transfected with 60 µg of pDC2-Cas9-yFCU-hDHFR-sgRNA-Delx and 12 µg of either pL-ap2g_Del1_GFP_yfcu (ΔATG-1, the only transgenic line that was generated using only one guide RNA) or pL-7-ap2g_Delx_GFP_yfcu (ΔATG-2, ΔCDS, ΔU and ΔU5') plasmids, linearized with *Xmn*I, *ScaI* or *Pvu*I (Fig. S10A). The exception were the lines carrying the ΔF2-3'-B and ΔF1 deletions, which were generated using a three-plasmids transfection with 12 µg of pUC19-KO-Fx (linearized using *Xmn*I) and 60 µg of each of two different pDC2-Cas9-hDHFR-sgRNA-Delx plasmids containing different guides (Fig. S10B). This was followed by subcloning by limiting dilution, when required, to obtain pure populations of edited parasites. In two cases in which we could not obtain edited parasites using regular transfections (ΔU3' and ΔF2-3-A, the latter is equivalent to ΔF2-3-B, with the same F2-3 region deleted, but generated differently), we used Nucleofactor transfection of schizonts. For this, cultures were tightly synchronized to a 2 h age window using Percoll purification followed by sorbitol lysis 2 hours later. When most parasites reached the mature schizont stage, a new Percoll purification was performed, followed by transfection using the AMAXA P3 primary cell 4D Nucleofector X Kit L (Lonza), according to manufacturer's

instructions, with 60 µg of pDC2-Cas9-hDHFR-gRNA-Delx plasmid and 60 µg of pL-7-ap2g_Delx_GFP_yfcu, linearized with *Sca*I or *Pvu*I.

Southern Blot

Southern blot was performed according to standard procedures [68]. gDNA was digested with the restriction enzymes *Aat*II and *Hinc*II, plus *Afl*II to cleave episomal plasmids. Digestions were resolved in 1.25% agarose gels and transferred to Amersham Hybond N+ nylon membranes (GE Healthcare) for detection with a specific probe labelled with α-^{32}P dATP (Perkin Elmer). The primers used to amplify the probe were the same as the ones used for the amplification of the HR1 of PF3D7_1144400 (Table S1). Hybridization and washes were performed at 62°C.

ChIP-qPCR and ChIP-seq

Chromatin extraction from cultures at the late trophozoite/schizont stage was performed as previously described, using the MAGnify™ Chromatin Immunoprecipitation System (Life Technologies) [30, 46, 58], with minor modifications. Briefly, for ChIP-qPCR, formaldehyde-crosslinked samples were sonicated using a Bioruptor® Plus sonication device (Diagenode) at high power, with 6 cycles of 30 seconds ON / 30 seconds OFF followed by centrifugation and recovery of the aqueous phase, which was sonicated for 2 additional cycles of 30 seconds ON / 30 seconds OFF to obtain sheared chromatin with a size around 400-500 bp. Immunoprecipitation was performed overnight at 4ºC with 0.5 µg of chromatin and 1 µg of antibodies against H3K9me3 (Diagenode), previously coupled to protein A/G magnetic beads provided in the MAGnify kit. Washing, de-crosslinking and elution were performed following the kit recommendations. Eluted DNA was diluted 1:2 for further downstream analysis by qPCR. For qPCR analysis, samples were analysed in triplicate wells using the standard curve method as described above for RT-qPCR, with a standard curve in each plate for each primer pair (primers described in Table S1).

For ChIP-seq [46], samples were sonicated using an M220 sonicator (Covaris) at 10% duty factor, 200 cycles per burst, 140 W of peak incident power for 10 min. Immunoprecipitation was performed overnight at 4ºC with 2 µg of chromatin and 4 µg of antibodies against H3K9me3 (Diagenode) or H3K9ac (Diagenode), previously coupled to protein A/G magnetic beads provided in the MAGnify kit. Washing, de-crosslinking and elution were performed following the kit recommendations, but avoiding high temperatures that may result in denaturation of extremely AT-rich intergenic regions: de-crosslinking, proteinase K treatment and elution were performed at 45ºC (for 2 h, overnight, and 1.5 h, respectively).

Libraries for Illumina sequencing were prepared from 5 ng of immunoprecipitated DNA using a protocol adapted for extremely AT-rich genomes [44, 46]. In brief, after end-repair and addition of 3′ A-overhangs, NEBNext Multiplex Oligos for Illumina (NEB, E7500) were ligated. Purification steps were performed with Agencourt AMPure XP beads (Beckman Coulter, A63880). Libraries were amplified using the KAPA HiFi PCR Kit (KAPA Biosystems, KK2103) in KAPA HiFi Fidelity Buffer with the following conditions: 95°C for 3 min; 9 cycles at 98°C for 20 s and 62°C for 2 min 30 s; and 62 °C for 5 min. Amplified libraries were purified using 0.9X AMPure XP beads to remove adapter dimers. The library size was analysed in a 4200 TapeStation System (Agilent Technologies). We obtained 12-16 million 125 bp paired-end reads per sample using the HiSeq2500 or the NextSeq 500 System (Illumina).

The analysis of ChIP-seq data was performed as previously described [16, 30, 46]. After initial processing, the reads where aligned to the reference 3D7 genome (PlasmoDB versions 39 and above) modified when required to include the alterations introduced after genome editing (i.e., deletion of the *pfap2-g* upstream region and part of the CDS, deletion of the CDS, integration of fragments at the PF3D7_1144400 locus or integration of a plasmid at the *pfap2-g* locus to add a C-terminal DD tag). After full processing, data was visualized using IGV (v2.4.10) [69]. To calculate genome-wide HC occupancy at genes, H3K9me3 coverage was

quantified at the -1000 to + 500 bp from the start codon of each gene, as previously reported [16, 30].

ACKNOWLEDGEMENTS

We thank L. Michel-Todó for his contribution to the bioinformatic analysis, C. Bancells for providing cDNA samples of E5 and E5-PfAP2-G-DD, N. Rovira-Graells for technical assistance with Southern blot analysis, J. J. López-Rubio (University of Montpellier) for plasmid pL6-egfp-yfcu, M. Lee (Wellcome Sanger Institute) for plasmid pDC2-Cas9-HDHFR and E. Knuepfer (The Francis Crick Institute) for plasmid pDC2-Cas9-hDHFRyFCU. This work was supported by grants SAF2016-76190-R, PID2019-107232RB-I00 and PID2022-137863OB-I00 to A.C. from the Spanish Ministry of Science and Innovation (MCI)/ Agencia Estatal de Investigación (AEI, 10.13039/501100011033), co-funded by the European Regional Development Fund (ERDF, European Union). A.P.-C. is the recipient of a Juan de la Cierva fellowship (FJC2021-047724-I) from the Spanish Ministry of Science, Innovation and Universities. O.L.-B. was supported by a FPU fellowship from the Spanish Ministry of Education, Culture and Sports (FPU014/02456). This research is part of ISGlobal's Program on the Molecular Mechanisms of Malaria, which is partially supported by the Fundación Ramón Areces. We acknowledge support from the grant CEX2023-0001290-S funded by MCIN/AEI/ 10.13039/501100011033, and support from the Generalitat de Catalunya through the CERCA Program.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol. 2018;19(4):229-44. PubMed PMID: 29235574.

2. Grewal SIS. The molecular basis of heterochromatin assembly and epigenetic inheritance. Mol Cell. 2023;83(11):1767-85. PubMed PMID: 37207657.
3. Trojer P, Reinberg D. Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell. 2007;28(1):1-13. PubMed PMID: 17936700.
4. Wang J, Jia ST, Jia S. New Insights into the Regulation of Heterochromatin. Trends Genet. 2016;32(5):284-94. PubMed PMID: 27005444.
5. Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, et al. A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol. 2012;19(10):1023-30. PubMed PMID: 22983563.
6. Thorn GJ, Clarkson CT, Rademacher A, Mamayusupova H, Schotta G, Rippe K, et al. DNA sequence-dependent formation of heterochromatin nanodomains. Nat Commun. 2022;13(1):1861. PubMed PMID: 35387992.
7. Seong KH, Li D, Shimizu H, Nakamura R, Ishii S. Inheritance of stress-induced, ATF-2-dependent epigenetic change. Cell. 2011;145(7):1049-61. PubMed PMID: 21703449.
8. Jia S, Noma K, Grewal SI. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science. 2004;304(5679):1971-6. PubMed PMID: 15218150.
9. Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 2017;543(7646):550-4. PubMed PMID: 28273063.
10. Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, et al. Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep. 2015;16(7):836-50. PubMed PMID: 26012739.
11. Phillips-Cremins JE, Corces VG. Chromatin insulators: linking genome organization to cellular function. Mol Cell. 2013;50(4):461-74. PubMed PMID: 23706817.
12. Hamali B, Amine AAA, Al-Sady B. Regulation of the heterochromatin spreading reaction by trans-acting factors. Open Biol. 2023;13(11):230271. PubMed PMID: 37935357.
13. Nicetto D, Zaret KS. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr Opin Genet Dev. 2019;55:1-10. PubMed PMID: 31103921.
14. Becker JS, Nicetto D, Zaret KS. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. 2016;32(1):29-41. PubMed PMID: 26675384.
15. Fraschka SA, Filarsky M, Hoo R, Niederwieser I, Yam XY, Brancucci NMB, et al. Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites. Cell Host Microbe. 2018;23(3):407-20. PubMed PMID: 29503181.
16. Michel-Todó L, Bancells C, Casas-Vila N, Rovira-Graells N, Hernández-Ferrer C, González JR, et al. Patterns of Heterochromatin Transitions Linked to Changes in the Expression of *Plasmodium falciparum* Clonally Variant Genes. Microbiol Spectr. 2023;11(1):e0304922. PubMed PMID: 36515553.
17. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, et al. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog. 2009;5(9):e1000569. PubMed PMID: 19730695.
18. Lopez-Rubio JJ, Mancio-Silva L, Scherf A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe. 2009;5(2):179-90. PubMed PMID: 19218088.
19. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, et al. Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res. 2012;22(5):925-38. PubMed PMID: 22415456.
20. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol. 2011;80(2):391-406. PubMed PMID: 21306446.
21. Jiang L, Lopez-Barragan MJ, Jiang H, Mu J, Gaur D, Zhao K, et al. Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. Proc Natl Acad Sci USA. 2010;107(5):2224-9. PubMed PMID: 20080673.
22. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A. 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. Mol Microbiol. 2007;66(6):1296-305. PubMed PMID: 18028313.

23. Cortés A, Crowley VM, Vaquero A, Voss TS. A view on the role of epigenetics in the biology of malaria parasites. PLoS Pathog. 2012;8(12):e1002943. PubMed PMID: 23271963.
24. Cortés A, Deitsch KW. Malaria Epigenetics. Cold Spring Harb Perspect Med. 2017;7(7):a025528. PubMed PMID: 28320828.
25. Voss TS, Bozdech Z, Bartfai R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. Curr Opin Microbiol. 2014;20:88-95. PubMed PMID: 24945736.
26. Llora-Batlle O, Tinto-Font E, Cortes A. Transcriptional variation in malaria parasites: why and how. Brief Funct Genomics. 2019;18(5):329-41. PubMed PMID: 31114839.
27. Satory D, Gordon AJ, Halliday JA, Herman C. Epigenetic switches: can infidelity govern fate in microbes? Curr Opin Microbiol. 2011;14(2):212-7. PubMed PMID: 21496764.
28. Starrfelt J, Kokko H. Bet-hedging--a triple trade-off between means, variances and correlations. Biol Rev Camb Philos Soc. 2012;87(3):742-55. PubMed PMID: 22404978.
29. Bunnik EM, Cook KB, Varoquaux N, Batugedara G, Prudhomme J, Cort A, et al. Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages. Nat Commun. 2018;9(1):1910. PubMed PMID: 29765020.
30. Nhim S, Tintó-Font E, Casas-Vila N, Michel-Todó L, Cortés A. Heterochromatin dynamics during the initial stages of sexual development in *Plasmodium falciparum*. Sci Rep. 2024;14(1):23180. PubMed PMID: 39369041.
31. Gomez-Diaz E, Yerbanga RS, Lefevre T, Cohuet A, Rowley MJ, Ouedraogo JB, et al. Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in Anopheles gambiae. Sci Rep. 2017;7:40655. PubMed PMID: 28091569.
32. Cui L, Fan Q, Cui L, Miao J. Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. Int J Parasitol. 2008;38(10):1083-97. PubMed PMID: 18299133.
33. Perez-Toledo K, Rojas-Meza AP, Mancio-Silva L, Hernandez-Cuevas NA, Delgadillo DM, Vargas M, et al. *Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of *var* genes. Nucleic Acids Res. 2009;37(8):2596-606. PubMed PMID: 19270070.
34. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, et al. Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. Nucleic Acids Res. 2009;37(11):3788-98.
35. Singhal R, Prata IO, Bonnell VA, Llinás M. Unraveling the complexities of ApiAP2 regulation in *Plasmodium falciparum*. Trends Parasitol. 2024;40(11):987-99. PubMed PMID: 39419713.
36. Carrington E, Cooijmans RHM, Keller D, Toenhake CG, Bártfai R, Voss TS. The ApiAP2 factor PfAP2-HC is an integral component of heterochromatin in the malaria parasite *Plasmodium falciparum*. iScience. 2021;24(5):102444. PubMed PMID: 33997710.
37. Shang X, Wang C, Fan Y, Guo G, Wang F, Zhao Y, et al. Genome-wide landscape of ApiAP2 transcription factors reveals a heterochromatin-associated regulatory network during *Plasmodium falciparum* blood-stage development. Nucleic Acids Res. 2022;50(6):3413-31. PubMed PMID: 35288749.
38. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature. 2014;507(7491):248-52. PubMed PMID: 24572369.
39. Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genet. 2012;8(11):e1002992.
40. Freville A, Stewart LB, Tetteh KKA, Treeck M, Cortes A, Voss TS, et al. Expression of the MSPDBL2 antigen in a discrete subset of *Plasmodium falciparum* schizonts is regulated by GDV1 but may not be linked to sexual commitment. mBio. 2024;15(5):e0314023. PubMed PMID: 38530030.
41. Smith JD, Rowe JA, Higgins MK, Lavstsen T. Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. Cell Microbiol. 2013;15(12):1976-83. PubMed PMID: 23957661.
42. Bancells C, Llora-Batlle O, Poran A, Notzel C, Rovira-Graells N, Elemento O, et al. Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium falciparum*. Nat Microbiol. 2019;4(1):144-54. PubMed PMID: 30478286.

43. Adjalley SH, Chabbert CD, Klaus B, Pelechano V, Steinmetz LM. Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*. Cell Rep. 2016;14(10):2463-75. PubMed PMID: 26947071.
44. Kensche PR, Hoeijmakers WA, Toenhake CG, Bras M, Chappell L, Berriman M, et al. The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. Nucleic Acids Res. 2016;44(5):2110-24. PubMed PMID: 26578577.
45. Zhang M, Wang C, Otto TD, Oberstaller J, Liao X, Adapa SR, et al. Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. Science. 2018;360(6388). PubMed PMID: 29724925.
46. Llorà-Batlle O, Michel-Todó L, Witmer K, Toda H, Fernández-Becerra C, Baum J, et al. Conditional expression of PfAP2-G for controlled massive sexual conversion in *Plasmodium falciparum*. Sci Adv. 2020;6(24):eaaz5057. PubMed PMID: 32577509.
47. Filarsky M, Fraschka SA, Niederwieser I, Brancucci NMB, Carrington E, Carrio E, et al. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. Science. 2018;359(6381):1259-63. PubMed PMID: 29590075.
48. Dorer DR, Henikoff S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. Cell. 1994;77(7):993-1002. PubMed PMID: 8020105.
49. Wolffe AP. When more is less. Nat Genet. 1998;18(1):5-6. PubMed PMID: 9425885.
50. Jeninga MD, Tang J, Selvarajah SA, Maier AG, Duffy MF, Petter M. *Plasmodium falciparum* gametocytes display global chromatin remodelling during sexual differentiation. BMC Biol. 2023;21(1):65. PubMed PMID: 37013531.
51. Mori T, Nakashima M. Sequence-dependent heterochromatin formation in the human malaria parasite *Plasmodium falciparum*. Heliyon. 2023;9(9):e19164. PubMed PMID: 37681121.
52. Gupta AP, Chin WH, Zhu L, Mok S, Luah YH, Lim EH, et al. Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*. PLoS Pathog. 2013;9(2):e1003170. PubMed PMID: 23468622.
53. Bachmann A, Petter M, Krumkamp R, Esen M, Held J, Scholz JA, et al. Mosquito Passage Dramatically Changes var Gene Expression in Controlled Human *Plasmodium falciparum* Infections. PLoS Pathog. 2016;12(4):e1005538. PubMed PMID: 27070311.
54. Mira-Martínez S, van Schuppen E, Amambua-Ngwa A, Bottieau E, Affara M, Van Esbroeck M, et al. Expression of the *Plasmodium falciparum* Clonally Variant *clag3* Genes in Human Infections. J Infect Dis. 2017;215(6):938-45. PubMed PMID: 28419281.
55. Pickford AK, Michel-Todó L, Dupuy F, Mayor A, Alonso PL, Lavazec C, et al. Expression Patterns of *Plasmodium falciparum* Clonally Variant Genes at the Onset of a Blood Infection in Malaria-Naive Humans. mBio. 2021;12(4):e0163621. PubMed PMID: 34340541.
56. Spence PJ, Brugat T, Langhorne J. Mosquitoes Reset Malaria Parasites. PLoS Pathog. 2015;11(7):e1004987. PubMed PMID: 26133171.
57. Spence PJ, Jarra W, Levy P, Reid AJ, Chappell L, Brugat T, et al. Vector transmission regulates immune control of *Plasmodium* virulence. Nature. 2013;498(7453):228-31. PubMed PMID: 23719378.
58. Rovira-Graells N, Crowley VM, Bancells C, Mira-Martínez S, Ribas de Pouplana L, Cortés A. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. Nucleic Acids Res. 2015;43(17):8243-57. PubMed PMID: 26202963.
59. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, et al. A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. Nature. 2006;439(7079):1004-8. PubMed PMID: 16382237.
60. Cordon-Obras C, Barcons-Simon A, Scheidig-Benatar C, Claes A, Sabatet V, Loew D, et al. An exon DNA element modulates heterochromatin spreading in the master regulator for sexual commitment in malaria parasites. bioRxiv. 2020;https://doi.org/10.1101/2020.06.26.163105.
61. Cortés A. A chimeric *Plasmodium falciparum Pfnbp2b*/*Pfnbp2a* gene originated during asexual growth. Int J Parasitol. 2005;35(2):125-30. PubMed PMID: 15710432.
62. Casas-Vila N, Pickford AK, Portugaliza HP, Tintó-Font E, Cortés A. Transcriptional Analysis of Tightly Synchronized *Plasmodium falciparum* Intraerythrocytic Stages by RT-qPCR. Methods Mol Biol. 2021;2369:165-85. PubMed PMID: 34313989.

63. Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A, Lopez-Rubio JJ. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. Nat Biotechnol. 2014;32(8):819-21. PubMed PMID: 24880488.
64. Lim MY, LaMonte G, Lee MC, Reimer C, Tan BH, Corey V, et al. UDP-galactose and acetyl-CoA transporters as *Plasmodium* multidrug resistance genes. Nat Microbiol. 2016;1:16166. PubMed PMID: 27642791.
65. Knuepfer E, Napiorkowska M, van Ooij C, Holder AA. Generating conditional gene knockouts in *Plasmodium* - a toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9. Sci Rep. 2017;7(1):3881. PubMed PMID: 28634346.
66. Peng D, Tarleton R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. Microb Genom. 2015;1(4):e000033. PubMed PMID: 28348817.
67. Crabb BS, Rug M, Gilberger TW, Thompson JK, Triglia T, Maier AG, et al. Transfection of the human malaria parasite *Plasmodium falciparum*. Methods Mol Biol. 2004;270:263-76. PubMed PMID: 15153633.
68. Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. PLoS Pathog. 2007;3(8):e107. PubMed PMID: 17676953.
69. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92. PubMed PMID: 22517427.

FIGURES

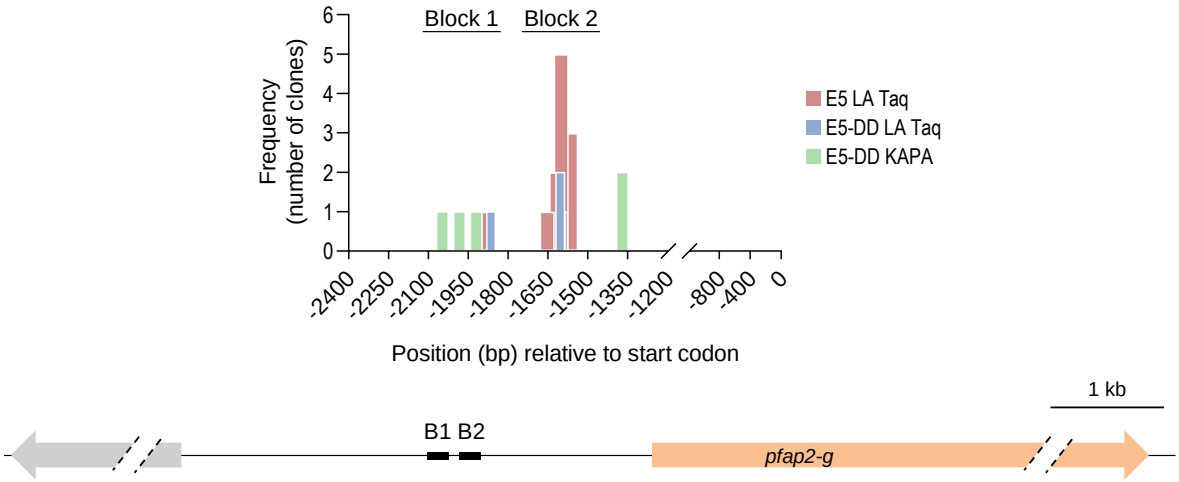

Figure 1. Identification of *pfap2-g* transcription start sites (TSSs)

Position relative to the start codon and frequency (number of clones) of the TSSs identified by the 5'RACE analysis of *pfap2-g* in the E5 and E5-AP2-G-DD (E5-DD) lines. The E5-DD line was cultured in the presence of Shld1. TSSs were identified from experiments in which PCR was performed with either LA Taq or KAPA DNA polymerases, as indicated. The position of the two main TSS blocks (B1 and B2) is indicated in the schematic (at scale) below.

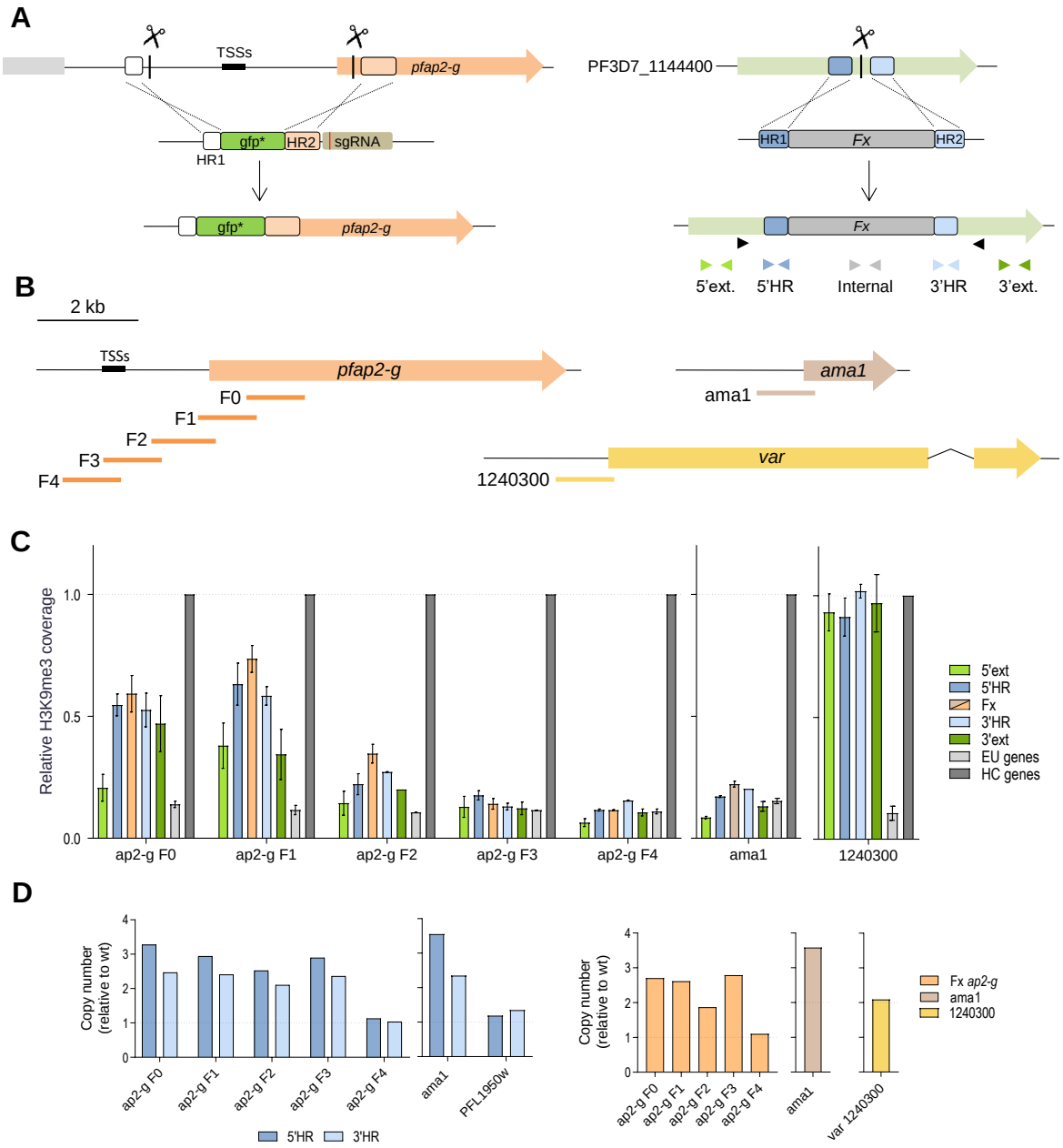

Figure 2. Identification of regions of the *pfap2-g* locus that can nucleate HC *de novo*

(A) Left: Strategy for the deletion of the upstream region and part of the CDS of *pfap2-g* using CRISPR/Cas9 (Δ5'ap2-g line). The scissors indicate the positions targeted by single guide RNAs (sgRNA). Homology regions (HR1 and HR2) are indicated as boxes. A non-functional *gfp* (gfp*) gene was used to replace the deleted region. Right: Strategy for the integration of different fragments into the euchromatic locus PF3D7_1144400 using CRISPR/Cas9. Fx refers to the fragment to be

integrated. Black arrowheads indicate the position of the primers used for diagnostic PCR, and coloured arrowheads the position of the primers used for ChIP-qPCR analyses.

(B) Schematic (at scale) of the position of the ~1 kb *pfap2-g, ama1* and *var* (PF3D7_1240300) fragments tested for HC nucleation capacity. The transgenic line with the *var* fragment was generated with plasmids containing the *yfcu* marker, but for the experiments presented here, it was analyzed before negative selection, which makes it analogous to the other transgenic lines.

(C) H3K9me3 ChIP-qPCR analysis of the transgenic lines with the integrated fragments (position of primers shown in panel A). Values are the % of the DNA recovered in the H3K9me3 IP (% input) at each position relative to the average % input in the positive control heterochromatic genes *var* PF3D7_1240300 (or *mspdbl2* for the 1240300 transgenic line) and *clag3.2* ("HC genes"). "EU genes" are euchromatic genes actin I (*actI*) and serine-tRNA ligase (*serrs*) used as negative controls (average of the two is shown). The dashed line indicates the level of enrichment in the positive control heterochromatic genes. Data are presented as the average and s.e.m. of two biological replicates.

(D) qPCR analysis of the number of copies of the PF3D7_1144400 HRs (5'HR and 3'HR) and the fragments under analysis in the gDNA of the transgenic lines. Copy number was calculated relative to 1.2B (wild type) gDNA. The dashed lines indicate the expected value for single, correct integration.

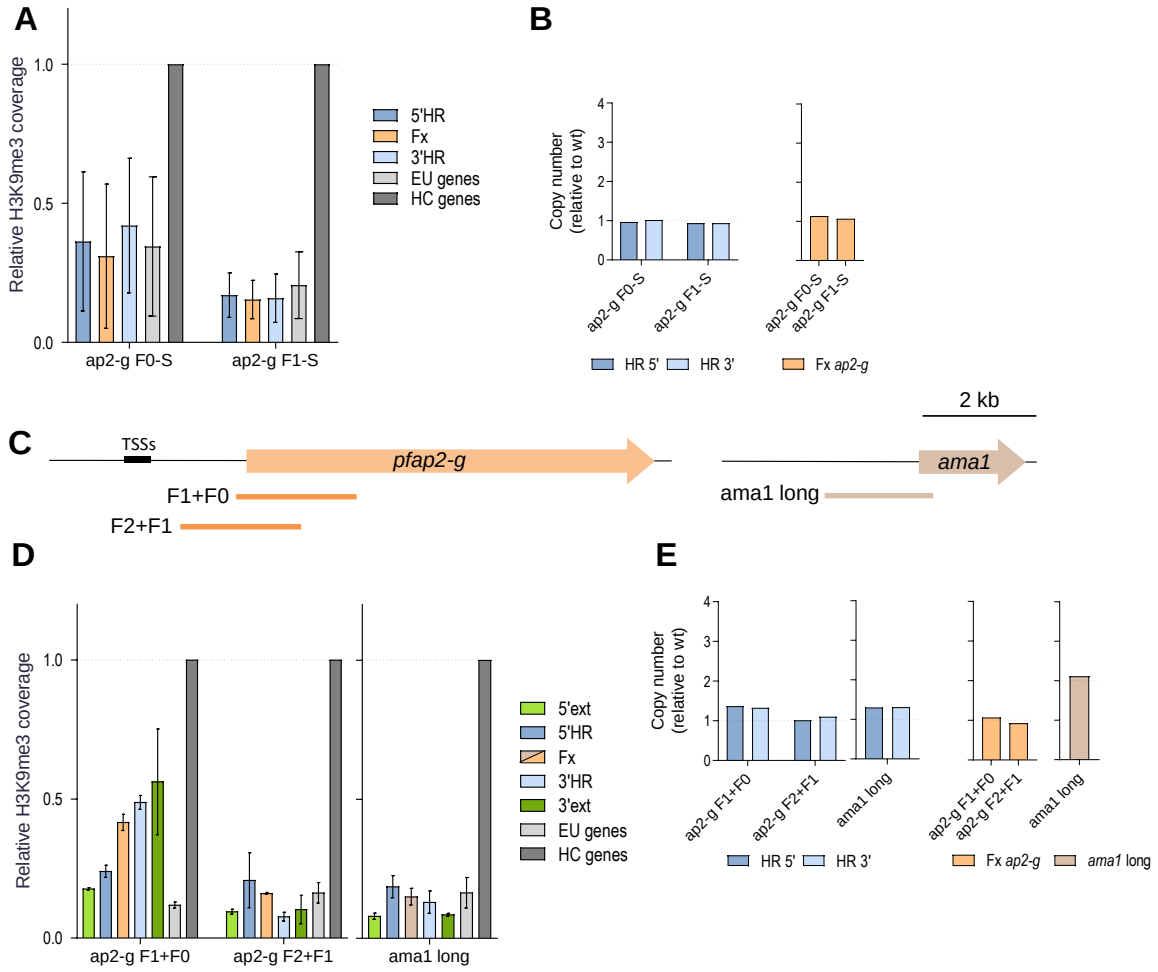

Figure 3. Effect of fragment size on HC nucleation

(A) H3K9me3 ChIP-qPCR analysis, as in Fig. 2C, of transgenic lines carrying single copies of *pfap2-g* fragments F0 and F1 (ap2-g F0-S and ap2-g F1-S). Data are presented as the average and s.e.m. of two biological replicates.

(B) qPCR analysis of copy number, as in Fig. 2D.

(C) Schematic (at scale) of the position of the ~2 kb *pfap2-g* and *ama1* fragments.

(D) H3K9me3 ChIP-qPCR analysis, as in panel A, of transgenic lines carrying single copies of ~2 kb *pfap2-g* and *ama1* fragments.

(E) qPCR analysis of copy number, as in panel B.

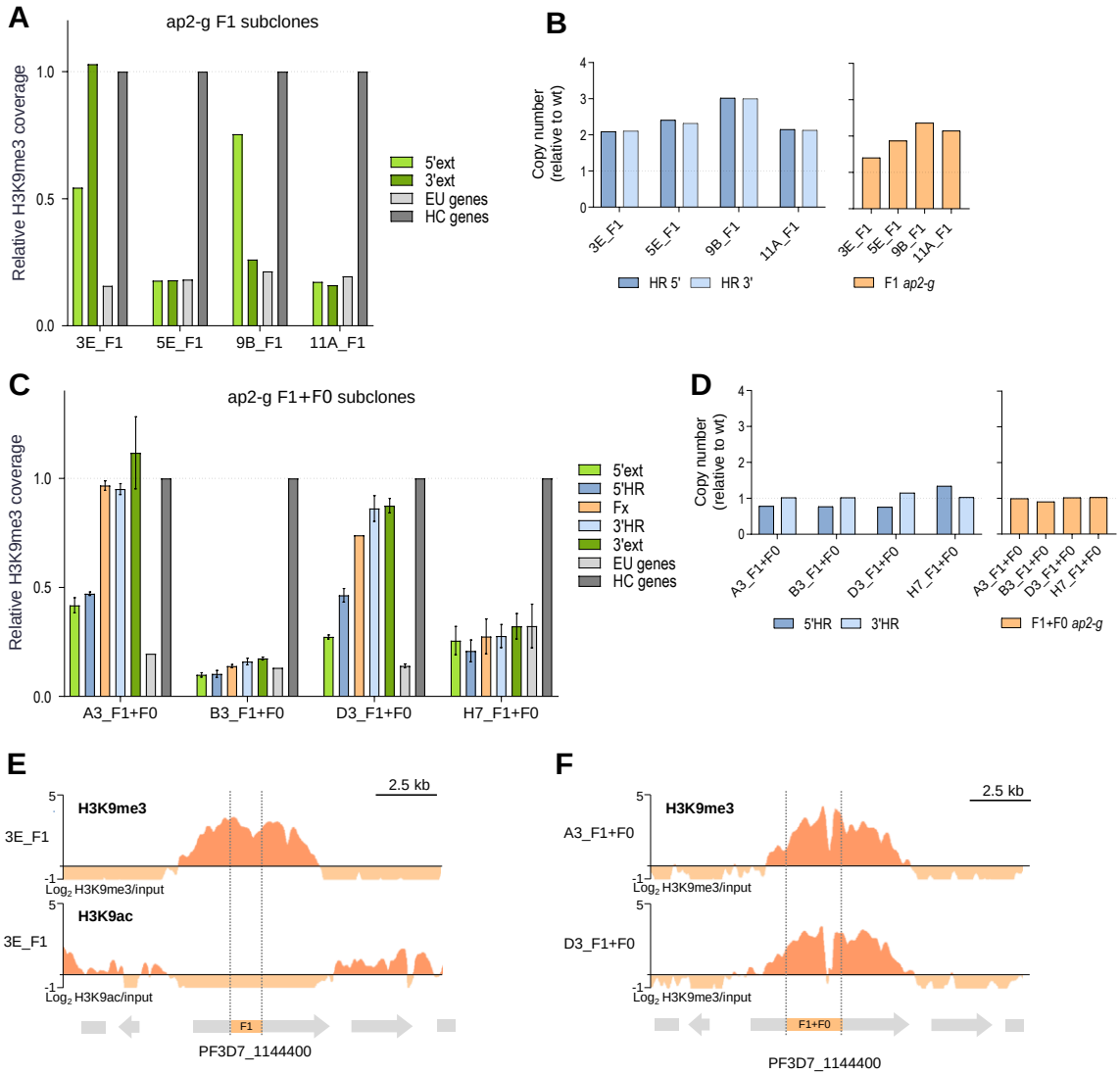

Figure 4. Cell to cell heterogeneity in HC nucleation by *pfap2-g* fragments

(A) H3K9me3 ChIP-qPCR analysis, as in Fig. 2C, of subclones of the transgenic line with the ap2-g F1 fragment. Data is from a single biological replicate.

(B) qPCR analysis of copy number, as in Fig. 2D.

(C) H3K9me3 ChIP-qPCR analysis, as in panel A, of subclones of the transgenic line with the ap2-g F1+F0 fragment. Data are presented as the average and s.e.m. of two biological replicates.

(D) qPCR analysis of copy number, as in panel B.

(E) H3K9me3 (top) and H3K9ac (bottom) ChIP-seq profile at the edited PF3D7_1144400 locus in the 3E subclone of the ap2-g F1 transgenic line. Reads

were aligned against the edited genome. The vertical dashed lines mark the position of the integrated fragment.

(F) H3K9me3 ChIP-seq profile at the edited PF3D7_1144400 locus, as in panel E, in the A3 and D3 subclones of the ap2-g F1+F0 transgenic line.

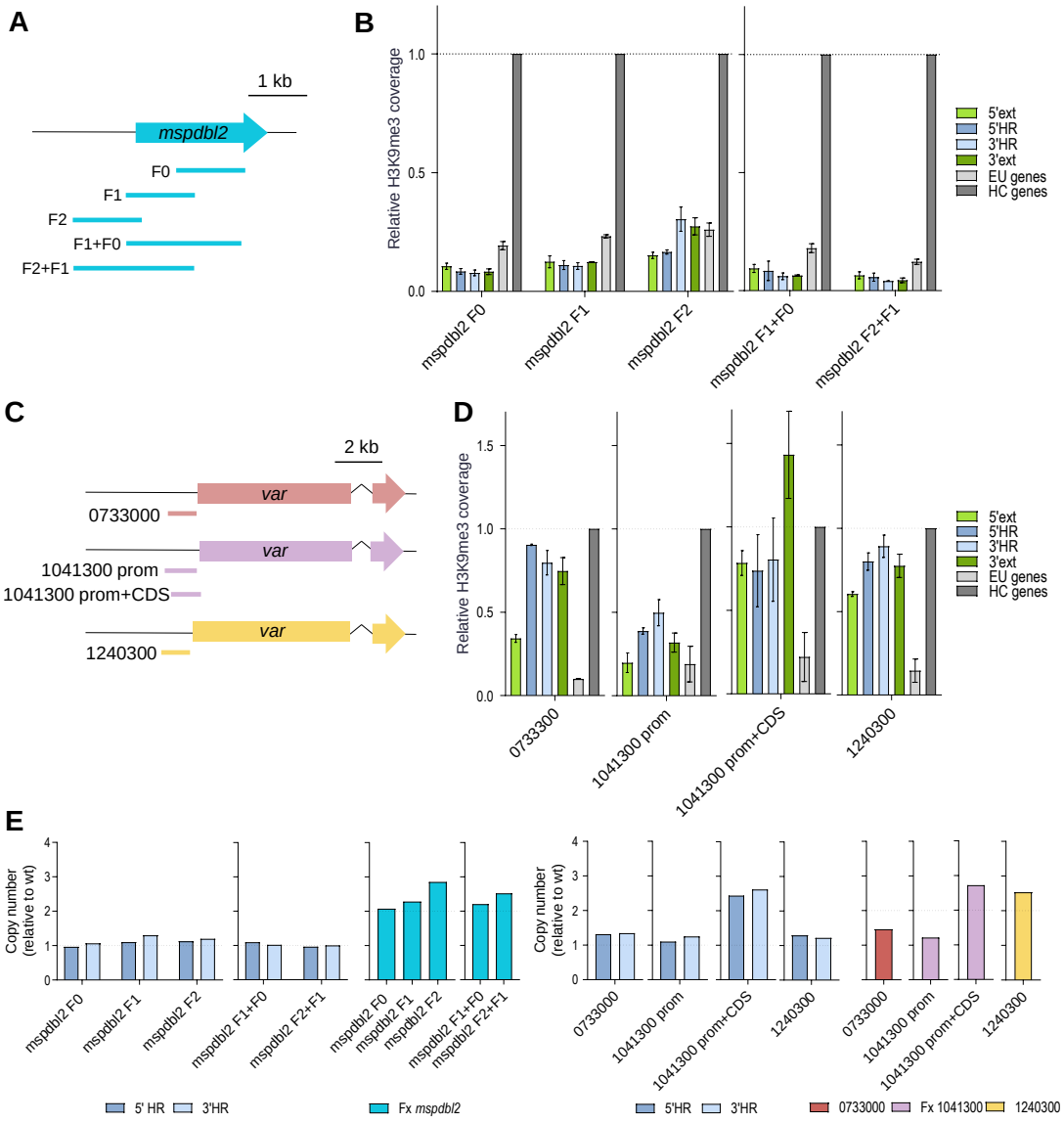

Figure 5. HC nucleation by sequences from other CVGs

(A) Schematic (at scale) of the position of *mspdbl2* fragments tested for HC nucleation capacity.

(B) H3K9me3 ChIP-qPCR analysis, as in Fig. 2C, of transgenic lines carrying *mspdbl2* fragments. Data are presented as the average and s.e.m. of two biological replicates.

(C) Schematic (at scale) of the position of fragments from different *var* genes tested for HC nucleation capacity: PF3D7_ 0733000 (0733000), PF3D7_1041300 (1041300) and PF3D7_1240300 (1240300).

(D) H3K9me3 ChIP-qPCR analysis, as in panel B, of transgenic lines carrying *var* fragments.

(E) qPCR analysis of copy number, as in Fig. 2D.

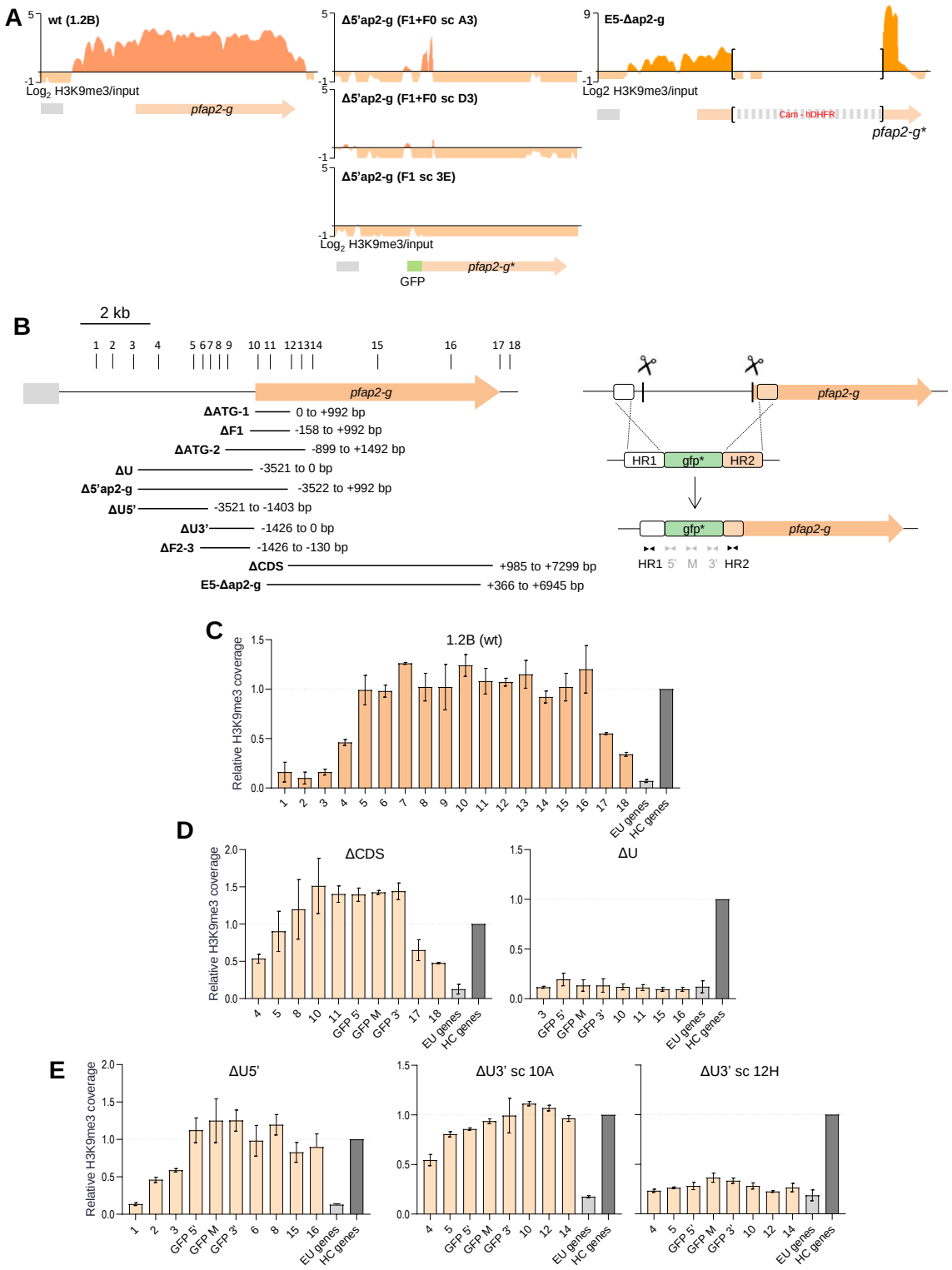

Figure 6. Analysis of the role of different regions of *pfap2-g* in HC maintenance

(A) H3K9me3 ChIP-seq profiles at the *pfap2-g* locus in wt (1.2B) and transgenic parasite lines. In experiments with Δ5'ap2-g-derived lines, reads were aligned

against the edited genome, whereas in the experiment with the E5-Δap2-g line, reads were aligned against a wt genome (the region deleted is indicated by square brackets).

(B) Left: Schematic (at scale) of the *pfap2-g* regions deleted to assess their role in HC maintenance. The numbered vertical lines on top of the scheme indicate the position of the primer pairs used for ChIP-qPCR analysis. Right: strategy for the deletion of *pfap2-g* regions using CRISPR/Cas9. The scissors indicate the positions targeted by the sgRNAs. A non-functional *gfp* gene (gfp*) was used to replace the deleted region. In this example, part of the upstream region and CDS were deleted, but in other cases different regions were eliminated. Black and grey arrowheads represent the primers used to assess the number of integrated copies. Primers for gfp* were also used in the ChIP-qPCR analyses.

(C) H3K9me3 ChIP-qPCR analysis of the wt 1.2B parasite line, from which transgenic lines with deletions at the *pfap2-g* locus were derived. Values are the % input at each position relative to the % input in the positive control heterochromatic genes, as in Fig. 2C. Numbers in the x axis are the primer pairs shown in panel B. The dashed line indicates the level of enrichment in the positive control heterochromatic genes. Data are presented as the average and s.e.m. of two biological replicates.

(D) H3K9me3 ChIP-qPCR analysis, as in panel C, of the ΔCDS and ΔU transgenic lines.

(E) H3K9me3 ChIP-qPCR analysis, as in panel C, of the ΔU5' transgenic line and the 10A and 12H subclones of the ΔU3' transgenic line.

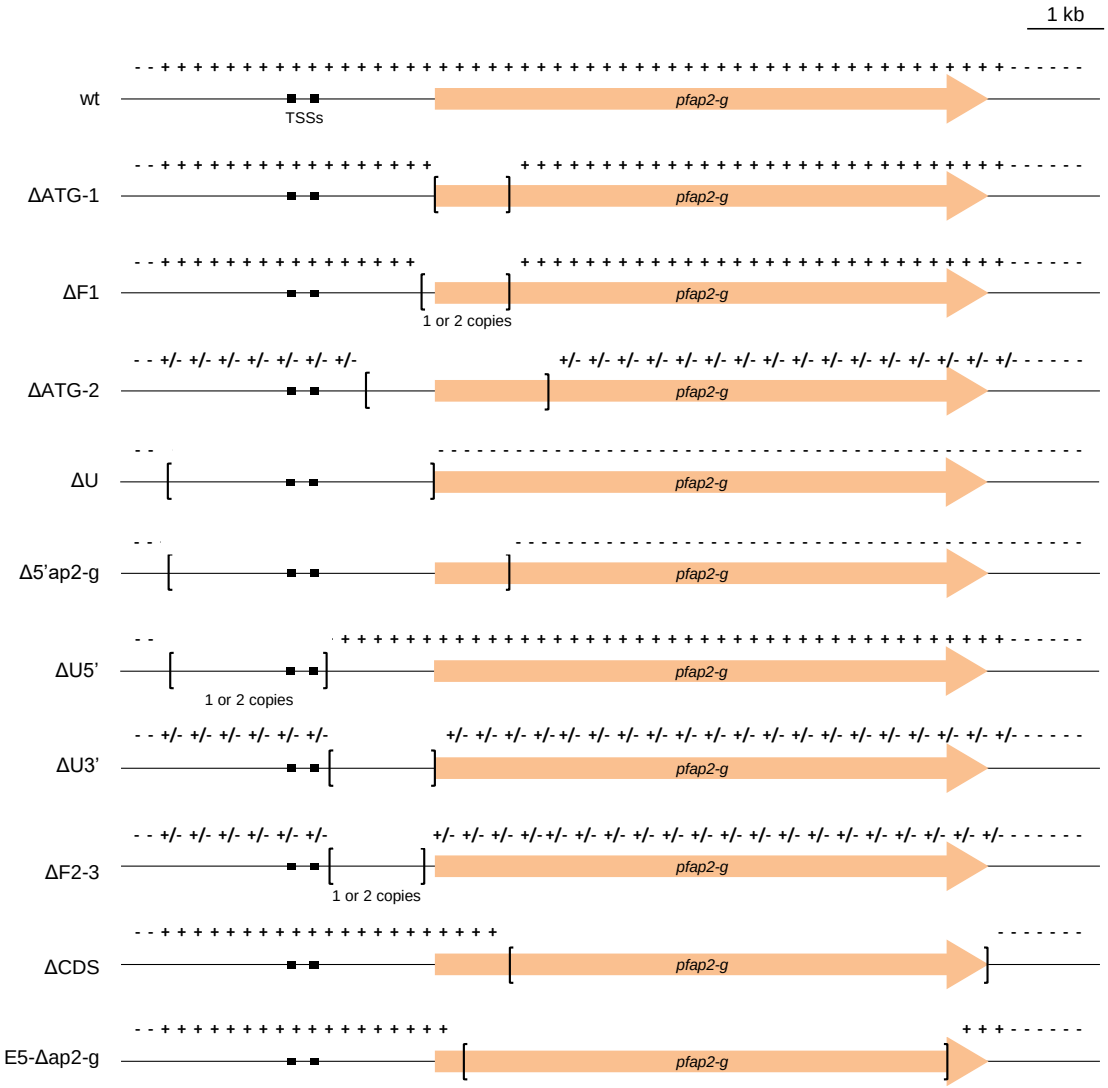

Figure 7. Summary of HC distribution in parasite lines with different deletions at the *pfap2-g* locus

Schematic (at scale) of HC distribution, as determined by ChIP-seq or ChIP-qPCR, at the *pfap2-g* locus after deletion of different regions. The deleted region in each transgenic line is indicated by square brackets. TSSs are represented by black boxes. Symbols + and – indicate presence or absence of HC, respectively, whereas +/– indicates partial HC coverage (compared to the wt 1.2B line) in a bulk parasite population or occurrence of HC in some subclones but not others. “1 or 2 copies” refers to the number of gfp* copies integrated.

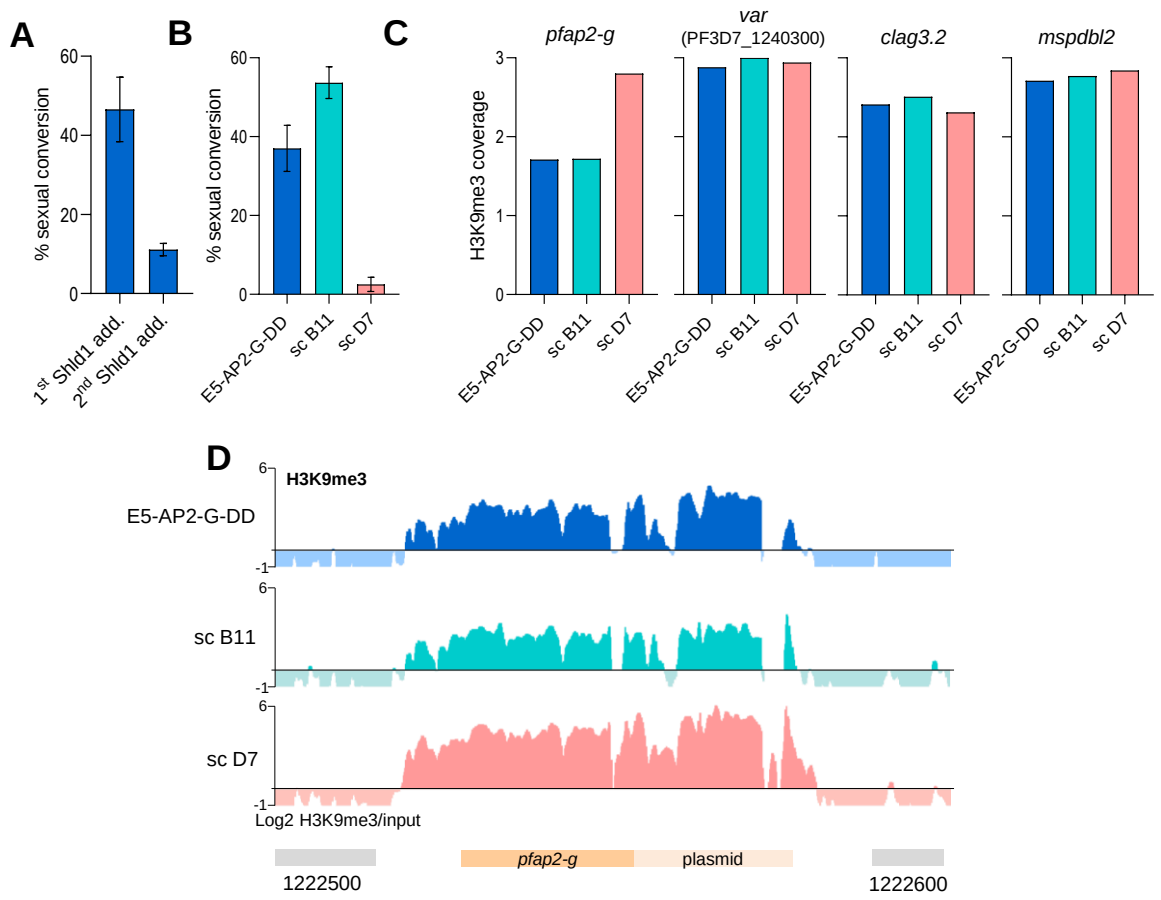

Figure 8. HC dynamics at the *pfap2-g* locus in the E5-AP2-G-DD line

(A) Sexual conversion rate of the E5-AP2-G-DD line after exposing cultures to Shld1 for the first time (1st Shld1 add.) and after re-exposing them to Shld1 following 2 growth cycles without Shld1 (2nd Shld1 add.). Data are presented as the average and s.e.m. of two biological replicates.

(B) Sexual conversion rates after Shld1 addition of the parental E5-AP2-G-DD line and two subclones (sc B11 and sc D7). Data are presented as the average and s.e.m. of three biological replicates.

(C) Quantification of H3K9me3 ChIP-seq coverage at the *pfap2-g, var* gene PF3D7_1240300, *clag3.2* and *mspdbl2* loci in the parental E5-AP2-G-DD line and its subclones B11 and D7 (no Shld1 added). Average coverage was calculated at the region from -1 kb to +0.5 kb from the start codon of each gene.

(D) H3K9me3 ChIP-seq profiles at the *pfap2-g* locus in the E5-AP2-G-DD line and the subclones B11 and D7. Reads were aligned against the edited genome.