A novel DNA-binding protein regulates sexual differentiation in malaria parasites and is essential for transmission.

R. Campelo Morillo,[1] X. Tong, [1] W. Xie ,[2] T. Lenz,[3] G. Batugedara,[3] N. Tabassum,[1] L. M. Orchard,[4] W. Daher,[5] D. J. Patel,[2] W. S. Noble,[6] M. Llinás, [4,7] K. G. Le Roch,[3] B. F.C. Kafsack [1,*]

Affiliations
[1] Department of Microbiology & Immunology, Weill Cornell Medicine, New York, NY 10065, USA.
[2] Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.
[3] Department of Molecular, Cell and Systems Biology, University of California Riverside, Riverside, CA 92521, USA.
[4] Department of Biochemistry & Molecular Biology and Huck Center for Malaria Research, Pennsylvania State University, University Park, PA, 16802.
[5] Dynamique des Interactions Membranaires Normales et Pathologiques, UMR5235 CNRS, INSERM, Université de Montpellier, Montpellier.
[6] Department of Genome Sciences, University of Washington, Seattle, WA, USA.
[7] Department of Chemistry, Pennsylvania State University, University Park, PA, 16802, USA.
* Please send correspondence to bkafsack@med.cornell.edu

ABSTRACT

Transmission of *Plasmodium falciparum* and other malaria parasites requires their differentiation from asexual blood stages into gametocytes, the non-replicative sexual stage necessary for transmission to the mosquito vector. This transition involves changes in gene expression and chromatin reorganization mediating the silencing and activation of stage-specific genes. However, malaria parasites have been noted for their dearth of transcriptional and chromatin regulators and the molecular mediators of these changes remain largely unknown. We identified HomeoDomain Protein 1 (HDP1) as a novel chromatin-associated DNA-binding protein that drives changes in chromatin structure and gene expression during early sexual differentiation. The discovery of a homeodomain-like DNA-binding protein marks a new class of transcriptional regulator in malaria parasites outside of the better-characterized ApiAP2 family. In this study, we demonstrate that HDP1 is required for gametocyte maturation and parasite transmission by driving the necessary upregulation of inner membrane complex components in early gametocytes.

INTRODUCTION

In order to complete its life cycle, *Plasmodium falciparum*, the most widespread and virulent of the human malaria parasite, must repeatedly differentiate into unique cell types that are able to access and exploit specialized niches within their human and mosquito hosts. One of these key developmental transitions occurs during the parasite's blood stage. Asexual blood-stages maintain a persistent infection through continuous lytic replication within erythrocytes but are not infectious to the mosquito vector, and therefore cannot mediate transmission to the next human host. This requires their differentiation into non-replicating male and female gametocytes that are able to infect the mosquito once taken up during a blood meal and involves changes in gene expression and chromatin reorganization that mediate the silencing and activation of stage-specific genes.

All differentiation requires the repression and activation of genes that underlie the specific phenotypes of the origin and destination cell types, respectively. To ensure commitment to one cell type or another, these transitions are often controlled by a bistable switch that controls the activity of a key regulator at the top of the transcriptional cascade that underlies the differentiation program (*1-4*). The broader downstream changes in gene expression are then accomplished by both changing the availability of specific transcription factors and changing their access to cell type-specific promoters via chromatin re-organization.

Recent work has found that this paradigm also applies in malaria parasites, where the initiation of sexual differentiation is controlled by bistable expression of a master regulator, the transcription factor AP2-G (*5*, *6*). During asexual replication the *ap2-g* locus is silenced by heterochromatin (*5*, *7*, *8*) but, due to the presence of AP2-G binding sites within its own promoter region, incomplete repression of *ap2-g* in individual cells can activate a transcriptional feedback loop that drives AP2-G expression to high levels, thereby committing cells to the gene expression program underlying sexual differentiation (*5*, *9*, *10*). Under conditions that impair heterochromatin maintenance, this feedback loop is activated in a larger fraction of cells, thereby increasing the frequency of sexual differentiation (*8*, *11*, *12*).

Commitment to the sexual differentiation program also triggers a substantial re-distribution of heterochromatin during the early stages gametocyte development (*13*, *14*). These changes involve substantial expansion of sub-telomeric heterochromatin domains on several chromosomes to silence asexual blood-stage specific genes. One such example is a cluster of genes on chromosome 2 that are required for trafficking of PfEMP1, a parasite-encoded adhesin on the erythrocyte surface

that is critical for immune evasion in asexually-replicating parasites but not expressed in gametocytes (*15*). Conversely, heterochromatin is lost from several loci that were previously silenced in asexual stages but upregulated during early gametocyte development (*13*, *14*). While AP2-G is critical for the initiation of sexual differentiation, it is expressed only during a small window beginning with sexually committed schizonts and is no longer required after the first 48 hours of gametocyte development (*10*, *16*). This suggests that the changes in gene expression underlying gametocyte development are carried out by a second wave of transcriptional regulators, which is consistent with the observation that AP2-G upregulates a number of putative transcription factors and chromatin remodeling enzymes (*9*, *10*).

While most malaria parasite species form round gametocytes within 2-6 days, sexual differentiation in *P. falciparum* takes substantially longer, 12-14 days, and produces gametocytes with a characteristic falciform morphology that give the parasite its name. Unsurprisingly, this prolonged gametocyte maturation is accompanied by a wide array of gene expression changes (*17-20*). However, relatively little is known about the transcriptional regulators downstream of AP2-G that mediate these changes. Compared to other single-celled eukaryotes, the genomes of malaria parasites contain a small fraction of DNA binding proteins. Like AP2-G, most belong to the ApiAP2 family (*21*), but only a small number have been shown to function specifically during gametocyte development (*22*). In the rodent malaria parasite *P. berghei* the DNA-binding protein PbAP2-G2 functions as a transcriptional repressor of asexual-specific gene expression (*6*, *23*), while PbAP2-FG (PyAP2-G3 in *P. yoelii*) was shown to mediate upregulation of female-specific transcripts (*24*, *25*). Whether these functions are conserved in *P. falciparum* remains to be determined.

In this study, we identify HDP1, a previously uncharacterized DNA-binding protein that is upregulated during sexual differentiation. Parasites lacking HDP1 and are unable to upregulate genes required for the expansion of the inner membrane complex and fail to progress beyond the earliest stages of gametocyte development. Additionally, loss of HDP1 expression results in an increase in chromatin accessibility, reduced telomere clustering, and leaky transcription of heterochromatin-associated genes in early gametocytes. Our findings indicate that HDP1 is an essential regulator of gene expression and chromatin during early gametocytogenesis.

RESULTS

hdp1 is essential for gametocyte development.

Searching for possible regulators of gene expression during *P. falciparum* sexual differentiation, we identified Homeodomain-like protein 1 *(hdp1, PF3D7_1466200)* encoding a 3078 amino acid protein with a C-terminal homeodomain-like domain, a helix-turn-helix structural motif commonly involved in DNA binding (fig. S1A) (*26*). Syntenic orthologs of *hdp1* could readily be identified in other malaria parasites with homology to the homeodomain-like domain extending to the coccidia but apparently absent from other apicomplexan clades (fig. S1B). Analysis of *hdp1* expression in *P. falciparum* blood stages by qRT-PCR detected only minimal expression in cultures of asexual blood stages (which always contain low number of gametocytes) with substantial upregulation during the early stages of gametocytogenesis (fig. S1C). *Pf*AP2-G, the transcriptional master switch that controls the initiation of the sexual differentiation gene expression program (*5*, *9*, *10*, *27*), binds at two sites located upstream of *hdp1* early in early gametocytes (*10*), consistent with our hypothesis that AP2-G activates additional regulators of gene expression during early gametocytogenesis.

In an effort to determine HDP1's subcellular localization, we inserted an in-frame N-terminal HaloTag at the endogenous locus (fig. S2A) in order to avoid possible interference with the putative DNA-binding activity of the C-terminal homeodomain-like domain. As expected based on transcript abundance, no Halo-tagged protein was detected in asexual stages. However, when we attempted to determine Halo-HDP1 levels in the sexual stages, we found that *halo-hdp1* cultures were unable to produce the characteristic crescent shapes of maturing *P. falciparum* gametocytes (Fig. 1A-B). Subsequent tagging the HDP1 C-terminus with either GFP or a triple Ty1 epitope tag (*hdp1-gfp* and *hdp1-Ty1*, fig. S2B-C), yielded parasite lines that produce gametocytes indistinguishable in numbers and morphology from the wildtype parent (Fig. 2A-B), despite the proximity to the putative DNA-binding domain. To test whether insertion of the N-terminal tagging resulted in a loss of HDP1 function, we generated a Δ*hdp1* knockout line for comparison. We used genome editing to replace 1.4 kb at the 5' end of the *hdp1* locus with a selectable marker cassette (fig. S2D). The resulting Δ*hdp1* parasites exhibited no discernible change in phenotype in asexual blood-stages, but like the *halo-hdp1* parasites, were unable to form viable mature gametocytes (Fig. 1D-F). More detailed analyses using synchronous induction of gametocytogenesis found that both *halo-hdp1* and Δ*hdp1* have sexual commitment rates comparable to the NF54 parent (fig. S3) but are unable to complete gametocyte development (Fig. 1C).

As the substantial length of the *hdp1* coding sequence made genetic complementation infeasible, we generated an inducible *hdp1-glmS* knockdown parasites by inserting a triple Ty1 epitope tag followed by the autocatalytic *glmS* ribozyme at the 3' end of the endogenous *hdp1* coding sequence through genome editing (fig. S2E) (*28*). In the absence of glucosamine, the resulting *hdp1-glmS* gametocytes expressed HDP1 protein at levels comparable to the *hdp1-Ty1* parasites without the ribozyme and produced normal gametocytes in terms of number and morphology (fig. S4). Supplementation of the culture medium with 5 mM glucosamine during the first 5 days of gametocyte development had no discernible effect on *hdp1-Ty1* parasites but diminished HDP1 expression by 70% and reduced falciform gametocytes by 80% in *hdp1-glmS* parasites, recapitulating the phenotype of the *halo-hdp1* and Δ*hdp1* lines (Fig. 1G-H). Since *hdp1* transcript levels remain relatively constant during gametocyte development (fig. S1C), we wanted to test whether continuous expression of HDP1 was required for gametocyte maturation. Knockdown of *hdp1* levels with glucosamine reduced gametocyte maturation prior to day 5 but not thereafter, indicating that sufficient HDP1 protein had been produced by that time to support gametocyte maturation (fig. S4).

HDP1 is chromatin-associated nuclear protein in gametocytes that recognizes a GC-rich DNA motif as a dimer. An earlier study using antibodies raised against a low-complexity region of HDP1 (fig. S1A) had reported a variety of subcellular localizations, including export to the erythrocyte membrane of early gametocytes (*29*). This was surprising given the absence of a signal peptide, the presence of two predicted nuclear localization signals (fig. S1A), and localization of the *Toxoplasma gondii* ortholog to the nucleus of tachyzoites (fig. S5). To resolve this apparent disagreement, we carried out live-cell and immunofluorescence microscopy *hdp1-gfp* and *hdp1-Ty1* gametocytes, respectively. In both lines, HDP1 localized exclusively to the gametocyte nucleus (Fig. 2A-B) and we were unable to replicate the localization(s) described in the earlier study. Unfortunately, the antisera used in that study were no longer available according to the corresponding author, precluding a direct comparison. HDP1 expression was not detectable by microscopy or western blotting in asexual blood stages (fig. S6, Fig. 2C) but showed increasing expression from day 2 of gametocytogenesis (Stage I-II) onward, reaching maximal levels by day 5 (Stage III) before dropping again by day 8 (Stage IV) (Fig. 2C). Analysis of subcellular fractions from day 5 *hdp1-Ty1* gametocytes showed HDP1 almost exclusively in the parasite nucleus, with about 70% resistant to solubilization with up to 600 mM NaCl, indicating a tight association with chromatin (Fig. 2D) and validating HDP1 as a nuclear protein.

68 Since HDP1 is an integral component of chromatin and homeo-like domains typically mediate
69 protein-DNA interactions, we tested the HDP1's ability to bind DNA using a protein-binding
70 microarray (PBM). Recombinant HDP1 DNA-binding domain (DBD) binding was highly enriched
71 on probes containing the palindromic hexamer GTGCAC (Fig. 2E, fig. S7A-C). Since homeo-
72 domains generally recognize DNA as dimers (*26*, *30*), we carried out isothermal titration
73 calorimetry to measure the interaction of HDP1 with double-stranded DNA containing a tandem
74 motif with a 5 bp spacer that places the motifs one helical turn apart. The results indicated optimal
75 binding at 2:1 protein to DNA molar ratio with a dissociation constant of 2.8 μM supporting DNA
76 recognition as a dimer (Fig. 2F). We were able to confirm this preference of binding as a dimer
77 using DNA gel-shift assays with double-stranded probes containing either no match, a single
78 binding-motif, or a tandem motif. When compared to the tandem-motif probe, the gel-shift of the
79 single motif probe was substantially weaker but identical in size (Fig. 2G), again consistent with
80 DNA recognition as a dimer even when only a single motif is present. Intriguingly, a scan of the
81 3D7 reference genome for instances of GTGCAC tandem motifs with a 5-6 bp spacer found them
82 exclusively located at each chromosome end with the first tandem located within either 7 or 28 bp
83 of the ultimate *rep20* sub-telomeric repeat, with the exception of those ends where this region was
84 deleted as the result of telomere healing events (*31*) (fig. S7D, Data Set 1A). Earlier work had
85 predicted these conserved sequences to be recognized by the ApiAP2-domain protein SIP2 in
86 asexual blood stages based on its ability to bind them *in vitro* (*32*), however, their unique spacing
87 and palindromic nature clearly distinguishes these tandem motifs from the arrays of SIP2 binding
88 sites just downstream. While the functional role of these motifs in asexual and sexual blood stages
89 remains unclear, their exclusive localization at chromosome ends is highly conserved across *P.*
90 *falciparum* isolates of diverse geographic origins (Data Set 1B).
91
92 **Loss of HDP1 leads to dysregulation of gene expression in early gametocytes.**
93 Based on HDP1's ability to bind DNA, we wanted to test whether it plays a role in the regulation
94 of gene expression during early gametocytogenesis. In order to find changes in expression that may
95 account for the aberrant development of Δ*hdp1* gametocytes and to avoid looking at the
96 consequences of the loss in gametocyte viability that occurs by Day 5 of development, we decided
97 to analyze the transcriptome of Δ*hdp1* and parental NF54 on Day 2, when HDP1 is initially
98 expressed, but prior to any change in viability based on flow cytometric measurements of
99 mitochondrial membrane potential or morphology in HDP1 deficient parasites (Fig. 1D, fig. S3B).
100 Global comparison of transcript abundances found that the majority of genes were expressed at
101 similar levels, including canonical markers of very early gametocytes such as *pfs16* and *gexp5* (Fig.

202 3A). This confirms our earlier observation that HDP1-deficient parasites initiate gametocyte
203 development at similar rates and that changes in gene expression are not due to change in viability
204 at that point. However, we observed reduced transcript levels for 188 genes and increased levels
205 for 150 genes in Δ*hdp1* Stage I gametocytes compared to those of the NF54 parent line. Gene set
206 enrichment analysis (GSEA) found that transcripts encoding components of the inner membrane
207 complex were significantly down-regulated (Fig. 3B) while transcripts from heterochromatin-
208 associated multi-copy gene families were significantly over-represented among up-regulated genes
209 (Fig. 3C), including members of the *var*, *rifin*, *stevor*, and *PHISTa/b/c* gene families (Fig. 3D, fig.
210 S8A-B). While fold-changes for these genes could be substantial, the absolute increase in
211 transcripts for these genes was often relatively small but significantly above levels in WT cells
212 (Data Set 2), possibly indicating impaired heterochromatin-mediated silencing allowing for leaky
213 background transcription rather than full transcriptional activation. Comparison of transcript levels
214 of *var* and *rifin* family genes found similar levels of upregulation regardless of whether these were
215 located in subtelomeric or non-subtelomeric regions of heterochromatin (fig. S8C). Analysis of *var*
216 gene expression in asexual Δ*hdp1* ring-stages found expression of a single major *var* gene as
217 expected for a recently cloned parasite line and indicating that mutually exclusive *var* gene
218 expression remained unaffected (fig. S8D).
219
220 **Loss of HDP1 increases chromatin accessibility and decreases telomeric clustering.**
221 Since our experiments clearly showed that HDP1 can bind DNA and is tightly associated with
222 nuclear chromatin (Fig. 2), we set out to determine where in the genome HDP1 is bound by
223 chromatin immunoprecipitations sequencing. However, numerous attempts at HDP1 ChIP-seq with
224 both the GFP- and Ty1-tagged lines were unsuccessful in multiple labs (see Materials and Methods
225 for details). As an alternative approach, we used high-throughput chromatin-conformation capture
226 (Hi-C) to look for differences in chromatin organization of day 2 gametocytes between the NF54
227 and Δ*hdp1* lines. While correlation analysis of Hi-C replicates indicated a clear difference between
228 these lines (Fig. 4A), overall intra-chromosomal interactions appeared largely unchanged (fig. S9).
229 However, further statistical analysis revealed a genome-wide reduction in long distance interactions
230 in Δ*hdp1* (Fig. 4B). Genome-wide mapping of interaction changes revealed a substantial reduction
231 in interaction frequency between virtually all chromosomes ends (Fig. 4C), consistent with reduced
232 but not abrogated telomere clustering. Based on the small but significant upregulation of
233 heterochromatin associated genes and a reduction in telomeric clustering, we hypothesized that
234 chromatin compaction might be impaired in HDP1-deficient parasites resulting in increased
235 chromatin accessibility. Indeed, brief treatment with micrococcal nuclease (MNase) revealed

accelerated digestion of chromatin from Δ*hdp1* gametocytes (Fig. 5D), consistent with an increase in chromatin accessibility.

HDP1 is required for extension of the inner membrane complex in early gametocytes.

Among the genes with reduced transcript levels in HDP1 knockout parasites, those associated with the inner membrane complex were substantially enriched (Fig. 3A-B, Fig. 5A). This was notable since Δ*hdp1* gametocytes fail to elongate from spherical stage I gametocytes into the oblong Stage II, a process that depends on the extension of the inner membrane complex. Six separate inner membrane complex genes were confirmed by qRT-PCR to have greater than 2-fold lower transcript levels in HDP1 knockout gametocytes (Fig. 5B). These included PhIL1 and the PhIL1-interacting protein 1 (PIP1), both of which are highly upregulated in early *P. falciparum* gametocytes and essential for inner membrane complex assembly and gametocyte maturation (*33*). Strikingly, the spherical morphology of *hdp1*-deficient gametocytes perfectly resembles the phenotype of gametocytes when PhIL1 is knocked down. Comparison of PhIL1 protein levels showed a 75% percent reduction in Δ*hdp1* gametocytes compared to wild-type parent (Fig. 5C). The concomitant reduction in PhIL1 expression (Fig. 5D) and inner membrane complex extension (Fig. 5E) when HDP1 is knocked down with glucosamine in *hdp1-glmS* gametocytes confirmed the dependence of PhIL1 expression on HDP1.

DISCUSSION

Our understanding of the regulatory mechanisms that control sexual differentiation in *P. falciparum* has improved substantially in recent years. Much of this work has focused on the regulation of AP2-G, the master regulator of this developmental decision. However, it is becoming clear that the function of AP2-G is constrained to the initiation of the transcriptional program that underlies the nearly two week-long process of forming gametocytes (*16*). This suggests that a second wave of hitherto unknown transcriptional regulators is required to drive gametocyte-specific gene expression during early gametocyte development.

In this study, we demonstrate that HDP1 functions as an essential regulator during gametocytogenesis. A nuclear DNA-binding protein, HDP1 is specifically up-regulated during early gametocyte development and essential for their maturation. In particular, HDP1 regulates the expression of genes required for the expansion of the inner membrane complex that gives *P. falciparum* gametocytes their characteristic crescent shape. While its expression is gametocyte-specific in *P. falciparum,* the *P. berghei* ortholog (PBANKA_1329600) may also have a role during

asexual blood-stage replication as its disruption leads to significant reduction in growth (*34*). This suggests that the gametocyte-specific function of HDP1 may have evolved more recently in the Laveranian clade of malaria parasites which uniquely produce falciform gametocytes. While essential during early gametocyte development, whether HDP1 is required in other parasite stages remains to be determined. Since the knockdown system used in this study regulates expression at the transcript level, we can infer that additional translation of HDP1 is not required during the later stages of gametocyte maturation but HDP1 protein expressed during the earlier stages of gametocytogenesis may still be required during the later stages. Previous transcriptomic studies indicate that HDP1 is also expressed in ookinetes, suggesting a role even after mating in the mosquito midgut. Intriguingly, homeodomain-like proteins have been implicated in mating processes of other haploid protozoa, such as Dictyostelium (*35*).

Our experiments clearly show that HDP1 has the ability to bind DNA and is tightly associated with nuclear chromatin (Fig. 2) but chromatin immuno-precipitation remained unsuccessful, despite exhaustive attempts (see Materials and Methods). One likely explanation is that the C-terminal tags are not accessible to antibodies when HDP1 is component of chromatin. This explains the to carry out ChIP as well as the diffuse nuclear localization we observed (Fig. 2A-B). Unfortunately, N-terminal tagging of HDP1 is not a viable alternative either as it results in loss of function (Fig. 1). Another possibility is that HDP1 resides in especially difficult to solubilize fraction of chromatin which has been shown to lead to incomplete recovery of chromatin under standard ChIP conditions (*36*).

We also found that HDP1 is required for efficient heterochromatin-mediated gene silencing during sexual differentiation. In gametocytes lacking HDP1, we observed low but significant up-regulation across a wide array of heterochromatin-associated gene families. Chromosome conformation capture analysis showed broad alterations to chromatin organization in HDP1-deficient parasites with an overall reduction in long-distance interactions that was most pronounced at the end of chromosomes, consistent with reduced, but not ablated, clustering of the sub-telomeric heterochromatin regions in the nuclear periphery. While its cognate DNA motifs are exclusively found at the boundary of the subtelomeric repeats, localization of HDP1 appeared evenly distributed throughout the gametocyte nucleus rather than in foci at the nuclear periphery, suggesting that the HDP1 does not localize to these sites exclusively. One explanation is that HDP1 promotes chromatin packaging, leading to an overall increase in chromatin accessibility observed in HDP1 deficient cells. This would also explain the leaky expression of heterochromatin-associated genes

as efficient silencing depends on limiting promoter accessibility while the promoters of euchromatin-associated genes are already generally accessible and a reduction in packaging would therefore be predicted to have little impact on transcript levels.

Additional studies will focus on identifying interaction partners of HDP1 and elucidating its structure-function relationship. Only 2% of its protein sequence is comprised of the homeodomain-like DNA-binding domain while the remainder contains no other identifiable domains and conservation is generally weak outside this region. The fact that insertion of a large tag at the N-terminus results in loss of function indicates that critical interactions occur in this region, but which other regions are essential for HDP1's function remains unclear. Similarly, identifying interactions with other nuclear proteins and genomic locations will offer important insights into its function. Of particular interest will be whether HDP1 regulates gene expression by altering chromatin structure or is more directly involved in the recruitment of RNA polymerase II.

MATERIALS & METHODS

Parasite culture

Unless otherwise noted, parasite strains were grown in 0.5% AlbuMAX II supplemented malaria complete medium using stablished cell culture techniques (*37*) at 3% hematocrit and below 3% parasitemia. Strains expressing selectable markers were maintained under constant drug-selection. Toxoplasma tachyzoites were cultured as described in (*38*).

Gametocyte induction

Gametocytes were induced synchronously as previously described in (*9*). Gametocyte maturation was monitored by Giemsa-stained thin blood smears and gametocytemia was counted on the fifth day of development. The gametocyte commitment rate was calculated by dividing the day 5 gametocytemia by the day 1 parasitemia, counted before addition of N-acetyl-D-glucosamine. Gametocytes were purified from culture at the required development stage using magnetic columns (LS columns, Miltenyi Biotec). For knockdown experiments in the HDP1-glmS line, at the gametoring stage either 5 mM glucosamine or solvent control was added.

Generation of transgenic strains

Transfection of ring-stage parasites were performed as previously described (*39*) and selected on 4 nM WR99210 and 1.5 µM DSM1 for two weeks. To obtain Δ*hdp1* parasites, selection with

WR99210 was kept all the time. Genome editing was performed by CRISR/Cas9 technology using the system described by (*40*). The pL6-DX2 plasmid, containing the sgRNA-expression cassette and the positive selectable marker cassette (*cam* promoter-*hdhfr*-*hrp2*/ts 3'*utr*) was used as backbone. Homology boxes were PCR amplified from NF54 genomic DNA, using specific primers flanked by appropriate overlapping region from pL6-DX2 plasmid (list of primers, see Supplementary Table) that allowed the cloning by Gibson assembly. All PCR reactions were set using Advantage Genomic LA polymerase (Takara, Cat. Nº 639152). *Plasmodium* codon optimized sequences for HALO-tag and triple Ty1 epitope tag were synthetized as gene-Blocks (Genewiz). PCR products were purified from agarose gel using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat. Nº D4008). Homology boxes were cloned through *AflII* and *SpeI* cutting sites, while annealed oligonucleotides paired encoding sgRNA with targeting sequence (Table) were inserted through the *XhoI* site, placed upstream of the sgRNA scaffold sequence. Sanger sequencing confirmed the absence of undesired mutations in the homology boxes and the sgRNA. Genomic DNA from transfectant parasites was isolated with QIAamp DNA blood Kit (Qiagen, Cat. Nº 51106) and diagnostic PCR's were set using Taq phusion DNA polymerase (Invitrogen). The TGME49_233160-HA parasite line was generated as part of an earlier study by tagging of the endogenous locus in the *T. gondii* RH-ku80ko strain as described in (*38*).

Flow cytometric analysis of gametocyte viability

Gametocytes were stained with 16 µM Hoechst33342 and 50 nM DilC1 for 30 minutes at 37˚C. Using a Cytek DxP12 flow cytometer, gametocytemia was determined by gating for DNA-positive cells and gametocyte viability was inferred based on mitochondrial membrane potential dependent accumulation of DilC1(5) for 1000 gametocytes (*41*). Analysis was carried out using FlowJo 10. The gating strategy is shown in fig. S10.

Nuclear extract preparation and chromatin high salt fractionation

Nuclear enriched fraction was prepared following the protocol previously described in (*42*), with some modifications. Briefly, parasites released from RBC's by saponin treatment (0.01%) were lysed with ice-chilled CLB (20 mM HEPES, pH 7.9; 10 mM KCl; 1mM EDTA, pH 8.0; 1 mM EGTA, pH 8.0; 0.65% NP-40; 1 mM DTT, 1x Roche Complete protease inhibitors cocktail). Nuclei were pelleted at 3,000 x *g* for 20min at 4ºC and cytoplasmic fraction was removed. Nuclei were resuspended in digestion buffer (20 mM Tris- HCl, pH 7.5, 15 mM NaCl, 60 mM KCl, 1 mM $CaCl_2$, 5 mM $MgCl_2$, 300 mM sucrose, 0.4% NP-40, 1 mM DTT, 1x Roche Complete protease inhibitors cocktail EDTA-free) and treated with 5U of micrococcal nuclease (ThermoFisher, Cat. Nº 88216)

for 30 min in a water bath at 37ºC. Soluble and insoluble nuclear fractions were recovered by centrifugation at 3,000 x *g* for 10 min at 4º C. Insoluble nuclear fraction were treated with salt fractionation buffer (10 mM Tris-HCl, pH 7.4; 2 mM $MgCl_2$; 2 mM EGTA pH 8.0; 0.1% Triton X-100; 0.1 mM PMSF; 1x Roche Complete protease inhibitors cocktail) supplemented with increasing NaCl concentrations (80-600 mM) while rotating at 4ºC for 30 min. All supernatants were recovered by centrifugation at 700 x *g* for 4 min at 4ºC and last remaining pellet was resuspended in 1x PBS supplemented with protease inhibitors cocktail. 5% of each fraction was prepared for Western blotting in order to check quality of fractionation.

Microccocal nuclease digestion

Gametocytes on day 2 of development were isolated and crosslinked with 1% formaldehyde then lysed with hypotonic lysis buffer (10 mM Tris-HCl, pH 8.0; 3 mM $MgCl_2$; 0.2% NP-40; 1 mM DTT and 1x Roche Complete protease inhibitors cocktail). Mechanical lysis was enhanced by passing the suspension ten times through a 26 G 1/2 needle. Nuclei were pelleted at 3,000 x *g* for 20 min at 4ºC and cytoplasmic fraction was removed. Nuclei were resuspended in MNase buffer (50 mM Tris-HCl, pH 8.0; 5 mM $CaCl_2$; 0.4% NP-40) and 1U of MNase (ThermoFisher, Cat. Nº 88216) was added. Digestion was carried for 5 min at 37ºC and was stopped by adding 20mM EGTA pH 8.0. Soluble and insoluble fractions were recovered by centrifugation at 3,000 x *g* for 10 min at 4ºC. Samples were decrosslinked and deproteinized by adding 0.1% SDS and 100ug/ml of Proteinase K (Roche, Cat. Nº 03115887001) and incubated for 1 hour at 50ºC. DNA was isolated through standard phenol-chloroform-isoamyl alcohol extraction.

Immunoblotting

For SDS-PAGE, total protein lysates were prepared using saponin-lysed parasites resuspended with 1x Laemmli loading buffer diluted in 1x PBS supplemented with 1x Roche Complete protease inhibitors cocktail. Protein samples were separated in 4-15% polyacrylamide gels and transferred to 0.2 µm Immobilion-P^{SQ} transfer membrane (Millipore, Cat. No ISEQ00010) using a Bio-Rad transfer system. Membranes were blocked in 5% skim milk/1x TBS-Tween20 for 1 hour at RT. Primary and secondary antibodies were prepared in 3% skim milk/1x TBS-Tween20 and incubated for 1 hour at RT. Membranes were washed four times with 1x TBS-Tween20 for 10 min, after primary and secondary antibody incubations. The following primary antibodies were used in this study: Anti-Ty1 BB2 mouse (1:2,500; Invitrogen Cat. Nº MA5-23513), anti-PhIL1 rabbit (1:5,000 (*43*)), anti-PfHsp70 rabbit (1:5,000; StreesMarq Biosciences Cat. Nº SPC-186D), anti-Histone 4 rabbit (1:2,000; Diagenode Cat. Nº C15410156-50). HRP-conjugated anti-mouse and anti-rabbit

antibodies were used (1:5,000, Millipore). Immunoblots were incubated with the chemiluminescent substrate SuperSignal West Pico PLUS (ThermoFisher, Cat. Nº 34578) following manufacturer directions. Images were obtained using Azure c300 digital imaging system (Azure Biosystems).

Live-cell and Immunofluorescence microscopy

For live-cell microscopy of *hdp1-gfp* and NF54 gametocytes, infected red blood cells were stained with 16 µM Hoechst33342 in incomplete media for 15 min at 37º C and imaged at 1000× magnification using a Leica DMI6000 microscope with differential interference contrast bright field optics, DAPI, and GFP filter cubes with identical exposure times.

For immunofluorescence microscopy of *hdp1-Ty1* and *hdp1-glmS* gametocytes, cells were seeded onto slides with Concanavalin A (5 mg/ml; Sigma) as described in (*44*), then fixed with a solution of 4% paraformaldehyde/0.0075% glutaraldehyde for 20 min at 37º C. Parasites were permeabilized with 0.1% Triton X-100 for 15 min at RT followed by blocking with 3% BSA. Primary antibodies (anti-Ty1 BB2 mouse 1:1,000; anti-PhiL1 rabbit 1:400) were allowed to bind for 1 hour in 3% BSA/PBS followed by three washes with blocking buffer for 5 min. Secondary antibodies diluted 1:500 (anti-mouse-Alexa546 and anti-rabbit-Alexa488, Invitrogen) plus 16 µM Hoechst33342 were added in fresh blocking buffer and incubated for 1 hour. Z-stacks of stained specimens were collected at 1000× magnification using a Leica DMI6000 microscope with differential interference contrast bright field optics, DAPI, and RFP filter cubes with identical exposure times. Fluorescent channel z-stacks were deconvolved using the ImageJ DeconvolutionLab2 plugin (NLLS algorithm) followed by maximum intensity z-projection and background adjustment. Immunofluorescence microscopy of Toxoplasma tachyzoites was carries out as previously described (*38*).

Protein Expression and Purification

Expression of recombinant HDP1-DBD motif was done using the Glutathione S-transferase (GST) gene fusion system (GE, Healthcare). The pGEX-4T-1 (Addgene Cat. Nº 27458001) plasmid was used as backbone for cloning the codon optimized sequence comprising the last 87aa of HDP1. Plasmid pGEX-GST-HDP1-DBD was transformed into BL21 (DE3) competent *E. coli* strain (NEB, Cat. Nº C2527) and protein expression was done following the manufacturer directions with some modifications. Briefly, an overnight culture was inoculated with one bacterial colony in 2x YT media supplemented with the corresponding antibiotic. Next day, culture was diluted 1:100 with fresh media and kept at 30º C with vigorous agitation. Bacterial growth was monitored until culture reach exponential phase. At this point, IPTG (1 mM final concentration) was added and the culture was kept for another 2 hours at 30º C with vigorous agitation. Cells were harvested and

resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5; 100 mM NaCl; 1 mM DTT; 5% Glycerol; 1 mM PMSF; 1 mM EDTA; 1x protease inhibitors cocktail) supplemented with Lysozyme (1mg/ml, Sigma). In order to remove bacterial DNA from our putative DNA binding protein, lysate was treated with polyethyleneimine (PEI) (0.1% v/v) (*45*). Lysate was sonicated, cleared by centrifugation at 14,000 x *g* for 30 min at 4º C. Protein extract was recovered and GST-HDP1-DBD protein was purified using Pierce GST Spin purification kit (Cat. Nº 16106) following manufacturer directions. Protein of interest was dialyzed using Slide-A-Lyzer Dialysis Cassette 10,000 MWCO (ThermoScientific, Cat. Nº 66810) and concentrated using Amicon Ultra Centrifugal filters 10,000 MWCO (Millipore, Cat. Nº UFC901024). Purity was verified by Coomassie staining after SDS-PAGE and concentration was measured by Bradford assay.

Sequence encoding HDP1 aa2991-3078 were cloned into the MCS1 of the pRSFDuet-1 vector (Novagen) engineered with an N-terminal His-SUMO tag. The proteins were expressed in *E. coli* strain BL21 CodonPlus (DE3)-RIL (Stratagene). Bacteria were grown in Luria-Bertani medium at 37°C to OD600=0.8 and induced with 0.4 mM IPTG at 18°C overnight. Cells were collected via centrifugation at 5000×g and lysed via sonication in Lysis Buffer (20 mM Tris-HCl, pH 8.0; 500 mM NaCl; 20 mM imidazole, and 5% Glycerol) supplemented with 1 mM phenylmethylsulfonyl fluoride and 0.5% Triton X-100. Cellular debris was removed by centrifugation at 20,000×g, and the supernatant was loaded onto 5 ml HisTrap FF column (GE Healthcare) and eluted using the lysis buffer supplemented with 500 mM imidazole. The elution was dialyzed at 4°C overnight against the buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 5 mM β-mercaptoethanol) with ULP1 protease added (lab stock). The sample was reloaded on the HisTrap FF column to remove the His-SUMO tag. The flow-through was loaded on the Heparin column (GE Healthcare) and eluted with a gradient of NaCl from 300 mM to 1 M. The target protein was further purified by size exclusion chromatography (Superdex 200 [16/60], GE Healthcare) in the buffer (20 mM Tris-HCl, pH 7.5; 200 mM NaCl; 1 mM $MgCl_2$; and 1 mM DTT). The high purity eluting fractions were detected by SDS-PAGE and concentrated to around 20 mg/ml. The protein was flash-frozen in liquid nitrogen and stored at -80°C.

Protein Binding Microarray

All GST-HDP1-DBD binding was analyzed twice on Protein Binding Microarrays (PBMs) as previously described (*46*, *47*). In this study two different universal arrays: Agilent AMADID 016060 (v9) and AMADID 015681(v11) were used covering all contiguous 8-mers, as well as, gapped 8-mers spanning up to 10 positions. Purified GST-HDP1-DBD fusion proteins were

visualized on the PBMs using Alexa-488 conjugated anti-GST antibody. Data analysis was carried out using the PBM analysis software suite downloaded from http://thebrain.bwh.harvard.edu/PBMAnalysisSuite/index.html. Following normalization, enrichment scores were calculated and the "Seed-and-Wobble" algorithm was applied to the combined data to generate position weight matrices (PWMs). An enrichment score cut-off of 0.45 was used to separate high affinity binding from non-specific and low affinity binding. Secondary motifs were identified by running the "rerank" program until E-scores below 0.45 were obtained.

Isothermal Titration Calorimetry

All the binding experiments were performed on a Microcal ITC 200 calorimeter. Purified HDP1-DBD proteins were dialyzed overnight against the buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT) at 4°C. DNA oligos were synthesized by Integrated DNA Technologies (IDT) and dissolved in the same buffer. The assays perform with 1 mM DNA duplexes containing the tandem motif (TAGTGCACCTATGGTGCACTT) with 0.1 mM HDP1-DBD proteins. Each reaction's exothermic heat was measured by sequential injection of the 2 µL DNA duplexes into proteins solution, spaced at intervals of 180 seconds. The titration was according to standard protocol at 20°C and the data were fitted using the program Origin 7.0.

Gel-shift Assays

Electrophoretic mobility shift assays were performed using Light Shift EMSA kits (Thermo Scientific) using 24 pg of protein and 40 fmol of probe, as previously described (*48*). Biotinylated double-stranded probes were generated with the indicated sequence.

RNA Extraction, cDNA synthesis, and quantitative RT-PCR

Total RNA from saponin-lysed parasites was extracted using Trizol (Invitgrogen) and Direct-Zol RNA MiniPrep Plus kit (Zymo Research). The cDNA was prepared from 100-500ng total RNA (pre-treated with 2U DNase I, amplification grade) using SuperScript III Reverse Transcriptase kit (Invitrogen) and random hexamers. Quantitative PCR was performed on the Quant Studio 6 Flex (Thermo Fisher) using iTaq Sybr Green (Bio-Rad) with specific primers for selected target genes (Table). Quantities were normalized to seryl-tRNA synthetase (PF3D7_0717700). Analysis of expression of the *var* gene family was performed by using the primer set described in Salanti et al. 2003 (*49*).

RNA sequencing

Following gametocyte induction, highly synchronous cultures containing committed schizonts were added to fresh RBCs and allowed to reinvade for 12 hours prior to the addition of 50 mM N-acetyl glucosamine to block the development of hemozoin-containing asexual trophozoites. On day 2 of gametocyte development, stage I gametocytes were purified magnetically, and total RNA was extracted as described above. Following RNA isolation, total RNA integrity was checked using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA concentrations were measured using the NanoDrop system (Thermo Fisher Scientific, Inc., Waltham, MA). Preparation of RNA sample library and RNA-seq were performed by the Genomics Core Laboratory at Weill Cornell Medicine. rRNA was removed from Total RNA using Illumina Ribo Zero Gold for human/mouse/rat kit. Using Illumina TruSeq RNA Sample Library Preparation v2 kit (Illumina, San Diego, CA), Messenger RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. Second strand cDNA synthesis followed, using DNA Polymerase I and RNase H. The cDNA fragments then went through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were then purified and enriched with PCR to create the final cDNA library. Libraries were pooled and sequenced on Illumina HiSeq4000 sequencer with single-end 50 cycles. Read files were checked for quality by using FASTQC v0.11.5 (https://github.com/s-andrews/FastQC). Reads were trimmed to remove low-quality positions and adapter sequences using cutadapt (version 1.16) (*50*). Reads were mapped against the *P. falciparum* 3D7 reference genome v40 (*51*) using STAR aligner (version 2.61) (*52*) and nuclear-encoded genes were analyzed for differential gene expression using cufflinks (version 2.2.1) (*53*). Genes with a false discovery rate of <=0.05 with a mean FPKM >5 in at least one strain were called significant. For genes with FPKM > 5 in one strain and no detectable expression in the other, FKPM values were set to 0.1 for purposes of fold-change calculation. Gene Set Enrichment Analysis was carried out with the FGSEA v1.16.0 Bioconductor v 3.12 package (*54*) with an FDR cutoff of <=0.05.

Hi-C Sequencing

Parasites were crosslinked in 1.25% formaldehyde in warm PBS for 25 min on a rocking platform. Glycine was added to a final concentration of 150 mM, followed by 15 min of incubation at 37°C and 15 min of incubation at 4°C. The parasites were centrifuged at 660 x *g* for 20 min at 4°C, resuspended in 5 volumes of ice-cold PBS, and incubated for 10 min at 4°C. Parasites were centrifuged at 660 x *g* for 15 min at 4°C, washed once in ice-cold PBS, and stored as a pellet at

−80°C. Parasite pellets were then thawed on ice in Hi-C lysis buffer (10 mM Tris-HCl, pH8.0; 10 mM NaCl, 2 mM AEBSF, Roche Complete Mini EDTA-free protease inhibitor cocktail, 0.25% Igepal CA-630). Parasite membranes were disrupted by passing the lysate through a 26.5- gauge needle 15 times with a syringe. Samples were spun and pellets were washed once with ice-cold Hi-C lysis buffer, resuspended in 0.5% SDS and incubated at 62°C for 10 min to solubilize the chromatin. The SDS was neutralized by the addition of nuclease-free water and 1% Triton X-100, for 15-minutes at 37°C. DNA was then digested using *MboI* restriction overnight at 37°C. After digestion, the enzyme was heat inactivated at 62°C for 20 min, and then cooled to RT. 5' overhangs were filled in by Biotin-14-dCTP (Invitrogen) using DNA Polymerase I and Large (Klenow) Fragment (NEB) for 45 min at 37°C. Blunt-end ligation was performed using T4 DNA ligase at 20°C for 4 hours. The nuclei were spun down to remove random ligation products and to reduce the overall reaction volume and resuspend in decrosslinking buffer (50 mM Tris-HCl, pH 8.0, 1% SDS, 1 mM EDTA, 500 mM NaCl) before adding RNaseA (20 mg/ml) for 45 min at 37°C and proteinase K (NEB) treatment overnight at 45°C. DNA was extracted using Agencourt AMPure XP beads (Beckman Coulter) and eluted in 10 mM Tris-HCl at pH 8.0. The purified DNA was sheared on a Covaris S220 and biotinylated DNA fragments were pull-down using MyOne Streptavidin T1 magnetic beads (Invitrogen) at 55°C for 2 min, washed and resuspended in T4 DNA Ligase Buffer to perform end-repair on DNA fragments bound to the beads containing DNA Polymerase I, Large (Klenow) Fragment (NEB), and T4 DNA Polymerase (NEB) and incubated for 30 min at 20°C. The beads were washed twice before performing A-tailing using Klenow Fragment (3′→5′ exo–nuclease) (NEB)) for 30 min at 37°C. The beads were then washed twice and resuspended in T4 DNA Ligase Buffer to perform adapter ligation using the NEBNext Illumina Adapter for 15 min at 20°C in a thermomixer. The beads were then resuspended in 100 µl of 10 mM Tris-HCl at pH 8.0 and transferred to a new tube to amplify the library using the HiFi HotStart ReadyMix (KAPA Biosystems) as well as the universal forward primer and barcoded reverse primer before being incubated with the following PCR program: 45 sec at 98°C, 12 cycles of 15 sec at 98°C, 30 sec at 55°C, 30 sec at 62°C and a final extension of 5 min at 62°C. The library was then purified using double-SPRI size selection, with 0.5Å~ right-side selection (25 µl AMPure XP beads) and 1.0Å~ left-side selection (25 µl AMPure XP beads). Libraries were quantified by NanoDrop (Thermo Scientific) and Bioanalyzer (Agilent), prior to multiplexing and sequencing in a 75-bp paired-end run on a NextSeq500 (Illumina).

575 **Analysis of Hi-C Data**

576 Bed files of virtual MboI (GATC) digest of the *P. falciparum* genome were generated with
577 digest_genome.py from the HiC-Pro suite (*55*). The paired-end reads were processed (i.e. mapped,
578 paired, and de-duplicated), binned at 10kb resolution, and normalized using the ICE algorithm using
579 HiC-Pro v2.11.4 (*56*). Reproducibility and strain-dependent differences (Fig. 4A) between Hi-C
580 samples was evaluated by correlation analysis using HiCSpector v1.0.1 (*57*) and GenomeDISCO
581 (same correlation pattern as HiCSpector, not shown) (*58*). Interaction maps (Fig. S9) were
582 generated by plotting intra-chromosomal ICE-normalized interactions. Summary statistics (Fig.
583 4B) of intra-chromosomal interaction distances were calculated for each strain and chromosome
584 from normalized ICE interaction counts using R v4.0.0. Differential interaction analysis was carried
585 out at 10 kb resolution for each replicate using selfish (*59*) based on ICE-normalized interaction
586 counts from the Hi-C Pro all-valid-pairs matrix and reformatted into the ".hic" file-format using
587 Juicebox v1.9.9 (*60*). All genome-wide bin pairs with significant changes in interaction frequency
588 between the *hdp1*-ty1 and Δ*hdp1* strains with an FDR <=0.1 in both replicate pairs are shown in
589 the top left of Fig. 4C. For the analysis of chromosome-end interactions, interactions fold-change
590 were calculated for all interactions in bins between the chromosome end and the first protein coding
591 gene. As the final bin on each chromosome varies in size, an additional bin was added for ends
592 truncated by telomere healing (5R, 11R, and 14R) to ensure interaction counts were calculated for
593 at least 10kb.

594

595 **Description of Chromatin Immunoprecipitation Sequencing attempts.**

596 Since our experiments clearly that HDP1 has the ability to bind DNA and is tightly associated with
597 nuclear chromatin (Fig. 2), we have spent the last 2.5 years attempting determine where in the
598 genome HDP is bound by chromatin immunoprecipitation sequencing (ChIP-seq). Repeated
599 independent efforts at HDP1 ChIP-seq using crosslinked chromatin from the *hdp1-Ty1* gametocytes
600 were unsuccessful in the Kafsack (Weill Cornell), and Bartfai (Radboud University) labs while
601 control ChIP-seq for known chromatin-associated proteins worked. We then generated the *hdp1-*
602 *gfp* line in case the Ty1 epitope tags were no accessible. But independent ChIP-seq attempts with
603 crosslinked chromatin in the Kafsack and Llinás labs with hdp1-gfp gametocytes were also
604 unsuccessfully while controls worked. Finally, we adapted the Cut & Run (a method for native, *in*
605 *situ* ChIP) (*61*, *62*) for use in malaria parasites (manuscript forthcoming). Again, Cut & Run worked
606 great for other chromatin-associated proteins but failed repeatedly for HDP1.

607

REFERENCES AND NOTES

1. B. O. Park, R. Ahrends, M. N. Teruel, Consecutive Positive Feedback Loops Create a Bistable Switch that Controls Preadipocyte-to-Adipocyte Conversion. *Cell Reports*. **2**, 976–990 (2012).

2. S. Bhattacharya *et al.*, A Bistable Switch Underlying B-Cell Differentiation and Its Disruption by the Environmental Contaminant 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicol Sci*. **115**, 51–65 (2010).

3. T. M. Norman, N. D. Lord, J. Paulsson, R. Losick, Stochastic Switching of Cell Fate in Microbes. *Annu Rev Microbiol*. **69**, 381–403 (2015).

4. D. Satory, A. J. Gordon, J. A. Halliday, C. Herman, Epigenetic switches: can infidelity govern fate in microbes? *Curr Opin Microbiol*. **14**, 212–217 (2011).

5. B. F. C. Kafsack *et al.*, A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*. **507**, 248–252 (2014).

6. A. Sinha *et al.*, A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature*. **507**, 253–257 (2014).

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

8. N. M. B. Brancucci *et al.*, Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe*. **16**, 165–176 (2014).

9. A. Poran *et al.*, Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature*. **551**, 95–99 (2017).

10. G. A. Josling *et al.*, Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nature Communications*. **11**, 1–13 (2020).

11. B. I. Coleman *et al.*, A Plasmodium falciparum Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. *Cell Host Microbe*. **16**, 177–186 (2014).

12. M. Filarsky *et al.*, GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science*. **359**, 1259–1263 (2018).

13. E. M. Bunnik *et al.*, Changes in genome organization of parasite-specific gene families during the Plasmodium transmission stages. *Nature Communications*. **9**, 1910 (2018).

14. S. A. Fraschka *et al.*, Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites. *Cell Host Microbe*. **23**, 407–420.e8 (2018).

15. M. Tibúrcio *et al.*, Early gametocytes of the malaria parasite Plasmodium falciparum specifically remodel the adhesive properties of infected erythrocyte surface. *Cell Microbiol*. **15**, 647–659 (2013).

16. C. Bancells *et al.*, Revisiting the initial steps of sexual development in the malaria parasite Plasmodium falciparum. *Nat. Microbiol*. **4**, 144–154 (2019).

17. J. A. Young *et al.*, The Plasmodium falciparum sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol Biochem Parasitol*. **143**, 67–79 (2005).

18. H. J. Painter, M. Carrasquilla, M. Llinás, Capturing in vivo RNA transcriptional dynamics from the malaria parasite Plasmodium falciparum. *Genome Res*. **27**, 1074–1086 (2017).

19. R. van Biljon *et al.*, Hierarchical transcriptional control regulates Plasmodium falciparum sexual differentiation. *BMC Genomics*. **20**, 920 (2019).

20. R. S. Kent *et al.*, Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite Plasmodium berghei. *Nat. Microbiol*. **3**, 1206–1213 (2018).

21. S. Balaji, M. M. Babu, L. M. Iyer, L. Aravind, Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res*. **33**, 3994–4006 (2005).

22. K. Modrzynska *et al.*, A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the Plasmodium Life Cycle. *Cell Host Microbe*. **21**, 11–22 (2017).

23. M. Yuda, S. Iwanaga, I. Kaneko, T. Kato, Global transcriptional repression: An initial and essential step for Plasmodium sexual development. *Proceedings of the National Academy of Sciences* (2015), doi:10.1073/pnas.1504389112.

24. C. Zhang *et al.*, Systematic CRISPR-Cas9-Mediated Modifications of Plasmodium yoelii ApiAP2 Genes Reveal Functional Insights into Parasite Development. *mBio*. **8**, 610 (2017).

25. M. Yuda, I. Kaneko, S. Iwanaga, Y. Mura, T. Kato, Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. *Mol Microbiol*. **6**, 1–51 (2019).

26. T. R. Bürglin, M. Affolter, Homeodomain proteins: an update. *Chromosoma*. **125**, 497–521 (2016).

27. O. Llorà-Batlle *et al.*, Conditional expression of PfAP2-G for controlled massive sexual conversion in Plasmodium falciparum. *Science advances*. **6**, eaaz5057 (2020).

28. P. Prommana *et al.*, Inducible Knockdown of Plasmodium Gene Expression Using the glmS Ribozyme. *PLoS ONE*. **8**, e73783 (2013).

29. C. P. Nixon *et al.*, Antibodies to PfsEGXP, an Early Gametocyte-Enriched Phosphoprotein, Predict Decreased Plasmodium falciparum Gametocyte Density in Humans. *J INFECT DIS*. **218**, 1792–1801 (2018).

30. L. Pradhan *et al.*, Crystal Structure of the Human NKX2.5 Homeodomain in Complex with DNA Target. *Biochemistry*. **51**, 6312–6319 (2012).

31. S. F. Calhoun *et al.*, Chromosome End Repair and Genome Stability in Plasmodium falciparum. *mBio*. **8**, e00547–17 (2017).

32. C. Flueck *et al.*, A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog*. **6**, e1000784 (2010).

33. M. Parkyn Schneider *et al.*, Disrupting assembly of the inner membrane complex blocks Plasmodium falciparum sexual stage development. *PLoS Pathog*. **13**, e1006659 (2017).

34. E. Bushell *et al.*, Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell*. **170**, 260–272.e8 (2017).

35. K. Hedgethorne *et al.*, Homeodomain-like DNA binding proteins control the haploid-to-diploid transition in Dictyostelium. *Science advances*. **3**, e1602937 (2017).

36. S. Henikoff, J. G. Henikoff, A. Sakai, G. B. Loeb, K. Ahmad, Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res*. **19**, 460–469 (2009).

37. K. Moll, I. Ljungström, H. Perlmann, A. Scherf, Methods in malaria research. *Manassas* (2008).

38. J. Morlon-Guyot *et al.*, A proteomic analysis unravels novel CORVET and HOPS proteins involved in Toxoplasma gondii secretory organelles biogenesis. *Cell Microbiol*. **20**, e12870 (2018).

39. M. Rug, A. G. Maier, in *Malaria* (Humana Press, Totowa, NJ, Totowa, NJ, 2012), vol. 923 of *Methods in Molecular Biology*, pp. 75–98.

40. M. Ghorbal *et al.*, Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat Biotechnol*. **32**, 819–821 (2014).

41. T. Q. Tanaka, K. C. Williamson, A malaria gametocytocidal assay using oxidoreduction indicator, alamarBlue. *Mol Biochem Parasitol*. **177**, 160–163 (2011).

42. C. Flueck *et al.*, Plasmodium falciparum Heterochromatin Protein 1 Marks Genomic Loci Linked to Phenotypic Variation of Exported Virulence Factors. *PLoS Pathog*. **5**, e1000569 (2009).

43. E. Saini *et al.*, Photosensitized I NA- L abelled protein 1 (PhIL1) is novel component of the inner membrane complex and is required for Plasmodium parasite development. *Sci. Rep.* **7**, 15577 (2017).

44. A.-K. Mehnert, C. S. Simon, J. Guizetti, Immunofluorescence staining protocol for STED nanoscopy of Plasmodium-infected red blood cells. *Mol Biochem Parasitol*. **229**, 47–52 (2019).

45. R.R. Burgess, Use of polyethyleneimine in purification of DNA-binding proteins. *Meth. Enzymol.* **208**, 3–10 (1991).

46. M. F. Berger *et al.*, Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol*. **24**, 1429–1435 (2006).

47. M. F. Berger, M. L. Bulyk, Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols*. **4**, 393–411 (2009).

48. T. L. Campbell, E. K. De Silva, K. L. Olszewski, O. Elemento, M. Llinás, Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog*. **6**, e1001165 (2010).

49. A. Salanti *et al.*, Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria. *Mol Microbiol*. **49**, 179–191 (2003).

50. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10–12 (2011).

51. S. Warrenfeltz *et al.*, EuPathDB: The Eukaryotic Pathogen Genomics Database Resource. *Methods Mol Biol*. **1757**, 69–113 (2018).

52. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2012).

53. C. Trapnell *et al.*, Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. **31**, 46–53 (2013).

54. G. Korotkevich, V. Sukhov, A. Sergushichev, Fast gene set enrichment analysis. *bioRxiv*. **10**, 060012 (2019).

55. N. Servant *et al.*, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. **16**, 1–11 (2015).

56. M. Imakaev *et al.*, Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Meth*. **9**, 999–1003 (2012).

57. K.-K. Yan, G. G. Yardımcı, C. Yan, W. S. Noble, M. Gerstein, HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. *Bioinformatics*. **33**, 2199–2201 (2017).

58. O. Ursu *et al.*, GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics*. **34**, 2701–2707 (2018).

59. A. R. Ardakany, F. Ay, S. Lonardi, Selfish: discovery of differential chromatin interactions via a self-similarity measure. *Bioinformatics*. **35**, i145–i153 (2019).

60. J. T. Robinson *et al.*, Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *cels*. **6**, 256–258.e1 (2018).

61. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols*. **13**, 1006–1019 (2018).

62. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *elife*. **8**, 245 (2019).

ACKNOWLEDGMENTS

General: We wish to thank Dr. Pawan Malhotra for the generous gift of anti-PhIL1 antibodies, V. Carruthers for the generous gift of anti-Ty1 antibodies, the Weill Cornell Medicine genomics core for technical support, M. Llinás and R. Bartfai for separate HDP1 ChIP-seq attempts with *hdp1*-gfp and *hdp1*-ty1 gametocytes, respectively, as well as K. Deitsch and J. King for valuable feedback on the manuscript.

Funding: This work was supported by startup funds from Weill Cornell Medicine (BK), 1R01AI141965 (BK), 1R01AI138499 (BK), 1R01 AI125565 (ML), 1R01 AI136511 (KLR), R21 AI142506-01 (KLR), the University of California, Riverside (NIFA-Hatch-225935-KLR), and support from the Mathers Foundation (DJP).

Author contributions: Conceptualization: B.F.C.K.; Methodology: B.F.C.K., R.C.M., W.X., W.N., K.G.L.; Investigation: R.C.M., X.T., W.X., T.L., G.B., L.O., W.D.; Software, Formal Analysis, Data Curation: B.F.C.K., T.L.; Writing – Original Draft: R.C.M.; Writing – Review & Editing: B.F.C.K., R.C.M.; Visualization: R.C.M., B.F.C.K.; Supervision: B.F.C.K., K.G.L., M.L., W.N, D.P., Project Administration: B.F.C.K; Funding Acquisition: B.F.C.K.;

Competing interests: The authors declare no competing interests.

Data and materials availability:
Raw high throughput sequencing data have been deposited in the NCBI Sequence Read Archive under accession number SRPXXXX.
Analysis code is available at https://github.com/KafsackLab/HDP1.
All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

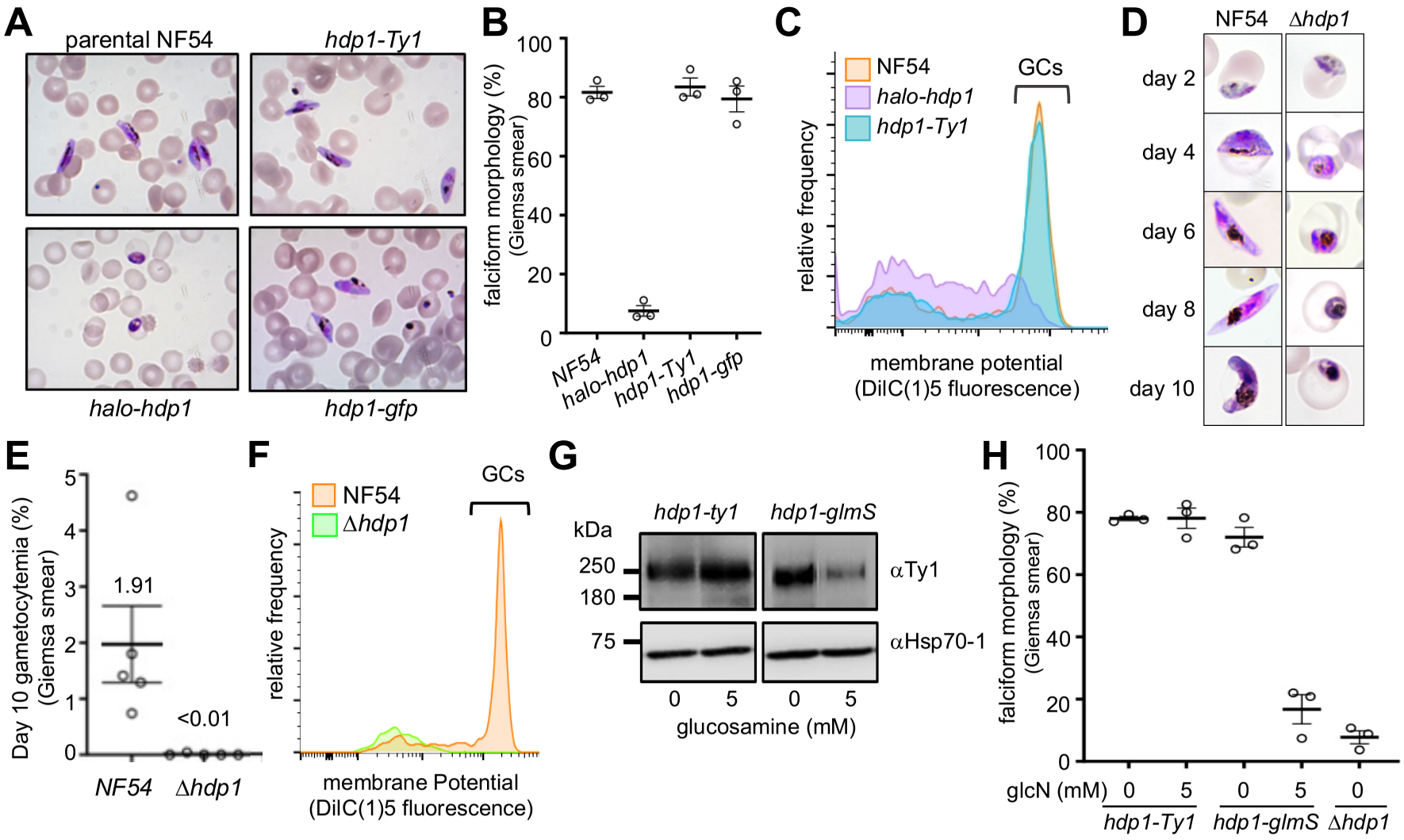

Fig. 1: Loss of HDP1 function disrupts gametocyte maturation. (A-C) N-terminal tagging of the endogenously encoded HDP1 (*halo-hdp1*) blocked maturation of gametocytes (GCs) while C-terminal tagging of the endogenous locus with either GFP (*hdp1-gfp*) or a triple Ty1 (*hdp1-Ty1*) epitope had no effect on gametocyte morphology or viability, as determined by membrane potential staining with DilC(1)-5 on day 5 of gametocyte maturation. **(D-F)** Targeted disruption of the *hdp1* locus (Δ*hdp1*) blocked formation of late gametocytes. Graphs show mean ± s.e.m of n=3-5 **(G)** Glucosamine-inducible knockdown of HDP1 in *hdp1-Ty1* and *hdp1-glmS* day 5 gametocytes. Representative of n=3. **(H)** Percentage of falciform day 5 gametocytes in response to 5 mM glucosamine (GlcN). All Images and flow cytometry plots are representative of n=3.

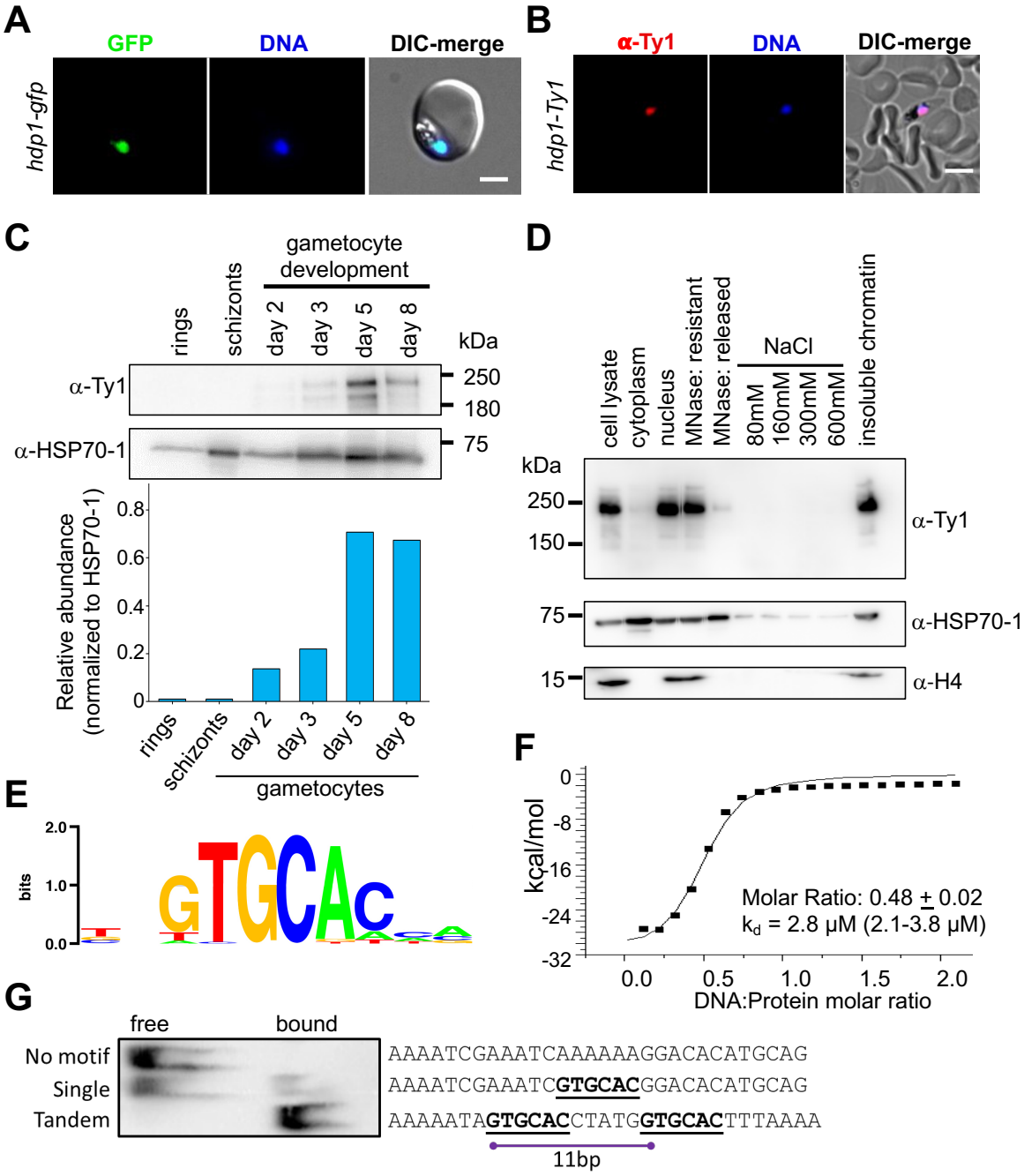

Fig. 2: HDP1 is chromatin-associated nuclear protein in gametocytes that recognizes a GC-rich DNA motif as a dimer. (A) Live-cell fluorescence microscopy of hdp1-gfp on day 5 of maturation stained with the DNA dye Hoechst33342 (blue). Scale Bar: 2mm. Representative of n=2. **(B)** Immunofluorescence microscopy of hdp1-Ty1 and parental NF54 gametocytes on day 5 of maturation co-stained with anti-Ty1 antibodies (red) and Hoechst33342 (blue). Scale Bar: 5 µm. Representative of n=2. **(C)** Western blotting of parasite lysates in asexual stages and during gametocyte maturation shows HDP1 is expressed during the stages of gametocytogenesis. Representative of n=3. **(D)** Western blot of cytoplasmic and nuclear extracts of hdp1-Ty1 gametocytes on day 5 of maturation stained with antibodies against the Ty1 epitope tag, the histone H4, and HSP70-1. Representative of n=3. **(E)** Maximum enrichment DNA motif for the GST-HDP1 DBD from a PBM. **(F)** Isothermal calorimetry indicates the HDP1-DBD domain recognizes DNA as a dimer. n=2. **(G)** Optimal gel-shift was observed for probes containing a tandem motif with a 5bp spacer compared to probes with either a single or no motif. Representative of n=3.

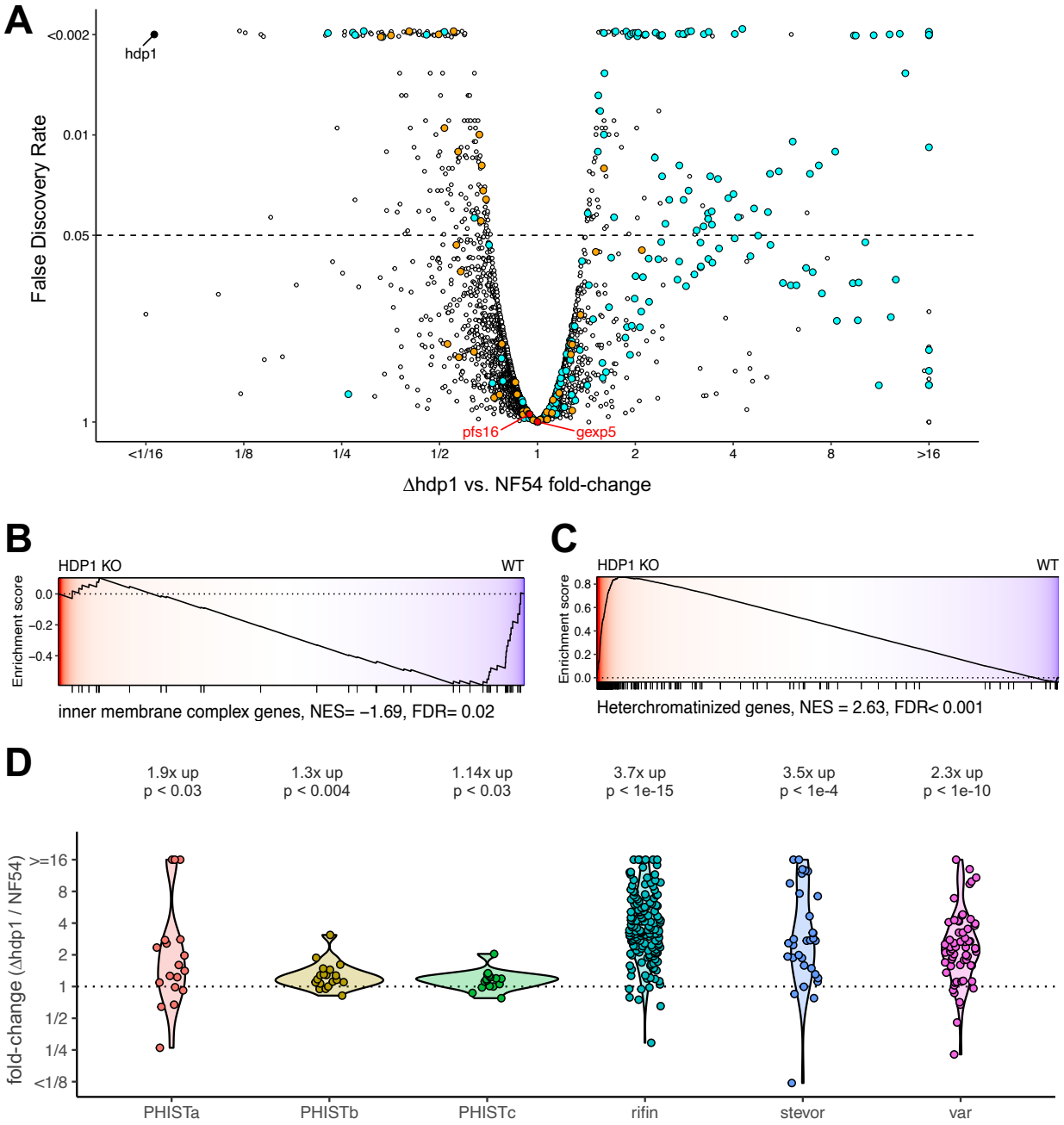

Fig. 3: Disruption of HDP1 results in leaky expression of heterochromatin-associated genes and reduced expression of inner membrane complex genes in early gametocytes. (A) Genome-wide comparison of differential gene expression in Δ*hdp1* and parental NF54 gametocytes on day 2 of gametocytogenesis (stage I, n=2). *hdp1* (solid black), heterochromatin-associated genes (cyan), IMC genes (orange) and the two canonical early gametocyte markers, *pfs16* and *gexp5 (red)* are highlighted. Gene set enrichment analysis (GSEA) indicated significant downregulation of IMC genes **(B)** and global upregulation of heterochromatin associated genes **(C)**. **(D)** Heterochromatin associated gene families with significant up regulation in Δ*hdp1*. Geometric mean fold-changes and p-values (two-sided, one sample t-test) are indicated.

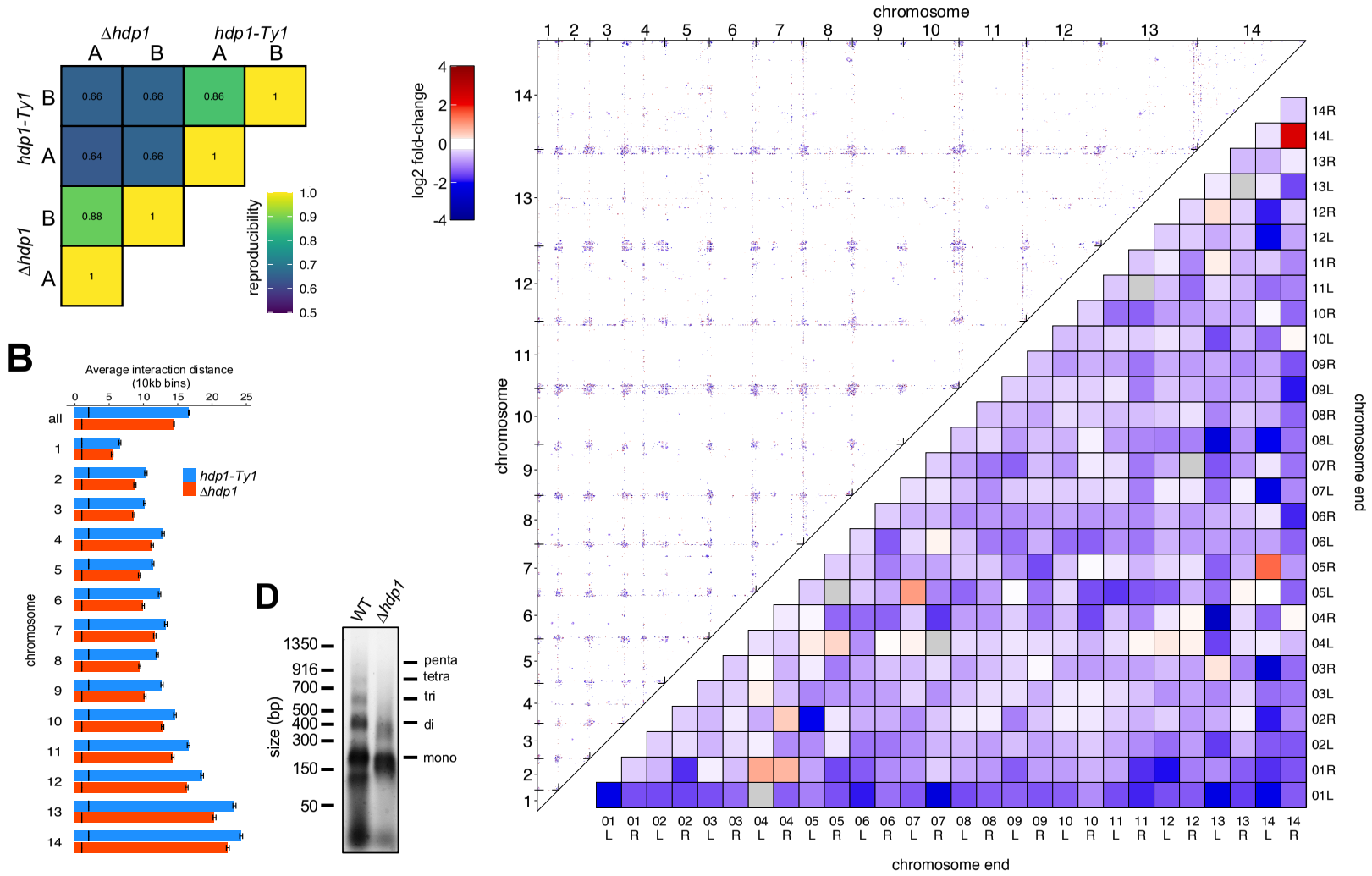

Fig. 4: Loss of HDP1 results in reduced telomere clustering. (A) Analysis of Hi-C interaction maps for each sample by HiCSpector highlights reproducibility of replicates and illustrate systematic differences between the strains. Scale ranges from 0-1.0 but only 0.6-1.0 is shown for improved visibility. **(B)** Mean HiC interaction distance in *hdp1-ty1* (blue) and Δ*hdp1* (red) stage I gametocytes. Error-bars indicate the 95% confidence interval and the black line indicates the median interaction distance. **(C)** Top triangle shows significant genome-wide changes in the Δhdp1 compared to hdp1-ty1 with inside ticks indicating chromosome boundaries. FDR < 0.1. The average interaction fold-changes for all pair-wise interactions of chromosome ends are shown in the bottom triangle. **(D)** Accelerated MNase digestion of Δ*hdp1* chromatin from stage I gametocytes indicates reduced chromatin compaction.

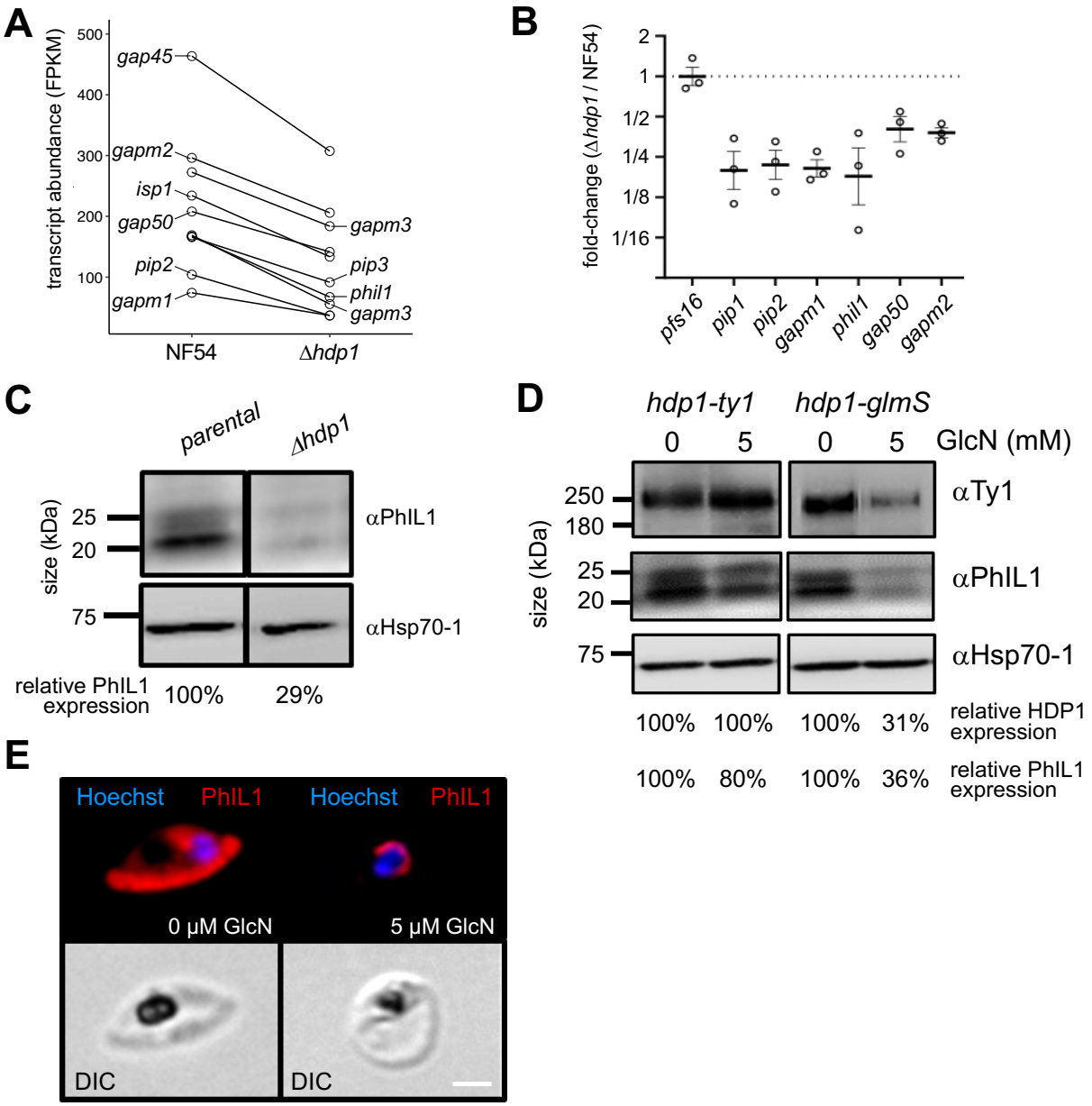

Fig. 5: HDP1 regulates PhIL1-dependent expansion of the inner membrane complex in early gametocytes. (A) Transcripts encoding inner membrane complex proteins are down-regulated in Δ*hdp1* day 2 gametocytes. Normalized transcript levels (FPKM) in NF54 and Δ*hdp1* day 2 gametocytes. **(B)** Validation of down-regulation by qRT-PCR. Levels of the canonical early gametocyte marker *pfs16* remained unchanged (n=3). **(C)** PhiL1 protein levels in parental and Δ*hdp1* day 5 gametocytes. Hsp70-1 abundance shown as the loading control. Representative result of n=2. **(D)** Day 5 morphology of *hdp1-ty1* and *hdp1-glmS* gametocytes under 0 and 5 mM glucosamine. HDP1 and PhiL1 protein levels in *hdp1-ty1* and Δ*hdp1* day 5 gametocytes. Hsp70-1 abundance shown as the loading control. Representative result of n=3. **(E)** Immunofluorescence microscopy of PhiL1 distribution in day 5 gametocytes of *hdp1-glmS* under 0 and 5 mM glucosamine. Scale Bar: 3 µm

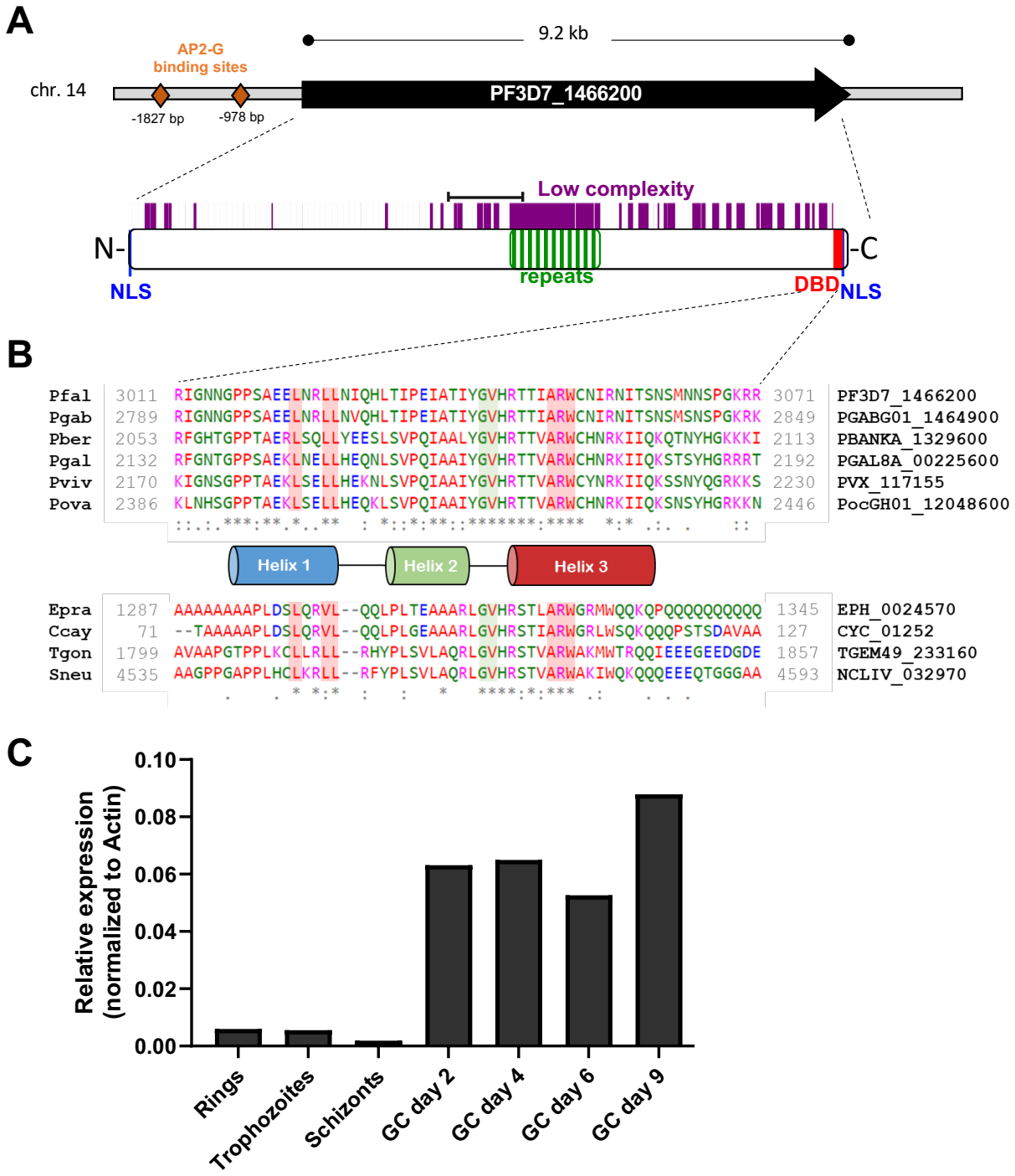

fig. S1: The predicted DNA-binding protein HDP1 is expressed in gametocytes. (A) The single exon locus Pf3D7_1466200 encodes a large 3078aa protein with a predicted C-terminal Helix-Turn-Helix DNA-binding domain (DBD) and two nuclear localization signals (NLS). Multiple AP2-G binding sites are located in the 2kb promoter region. Black bracket indicates the antigen used for generating antisera used in (29). **(B)** Alignment of the helix-turn-helix domain for homologs from other apicomplexan parasites. **(C)** Quantitative RT-PCR of *hdp1* transcripts found minimal expression in asexual blood stages with upregulation during gametocyte development. (mean of n=2).

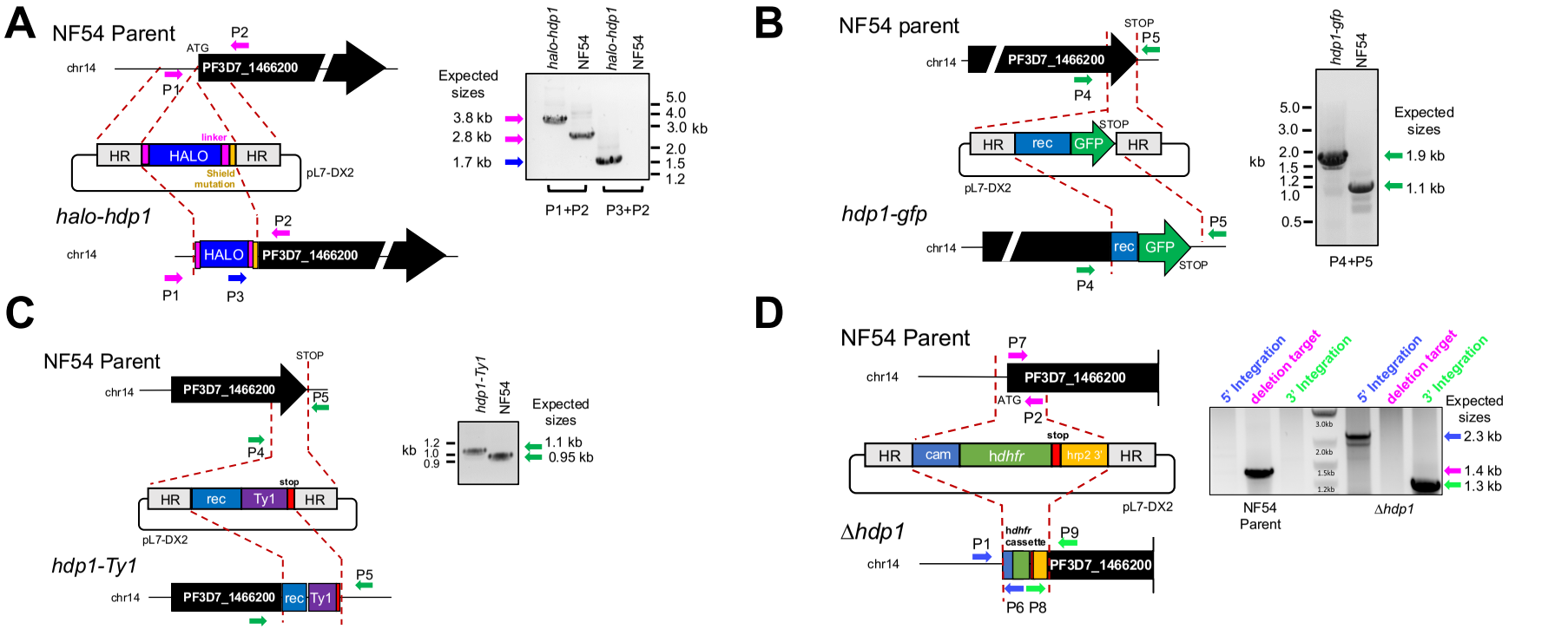

fig. S2: Validation of engineered parasite lines. (A) Generation of h*alo-hdp1* parasites by CAS9 genome editing. Insertion of the N-terminal HALO tag at the 5' end the *hdp1* coding sequence was confirmed by PCR and checked for mutations by Sanger sequencing of the 3.8kb PCR product (not shown). **(B)** Generation of *hdp1-gfp* parasites by CAS9 genome editing. Insertion of the C-terminal GFP tag at the 3' end the *hdp1* coding sequence was confirmed by PCR and checked for mutations by Sanger sequencing of the 1.9kb PCR product (not shown). **(C)** Generation of *hdp1-Ty1* parasites by CAS9 genome editing. Insertion of the C-terminal triple Ty1 epitope tag at the 3' end the *hdp1* coding sequence was confirmed by PCR and checked for mutations by Sanger sequencing of the 1.1kb PCR product (not shown). **(D)** Generation of Δ*hdp1* parasites by CAS9 genome editing. Replacement of 1.4 kb flanking the *hdp1* start codon by a hDHFR selectable marker cassette was confirmed by PCR. **(E)** Generation of *hdp1-glmS* parasites by CAS9 genome editing. Insertion of the C-terminal triple Ty1 epitope tag and the glmS ribozyme at the 3' end the *hdp1* coding sequence was confirmed by PCR and checked for mutations by Sanger sequencing of the 1.4 kb PCR product (not shown).

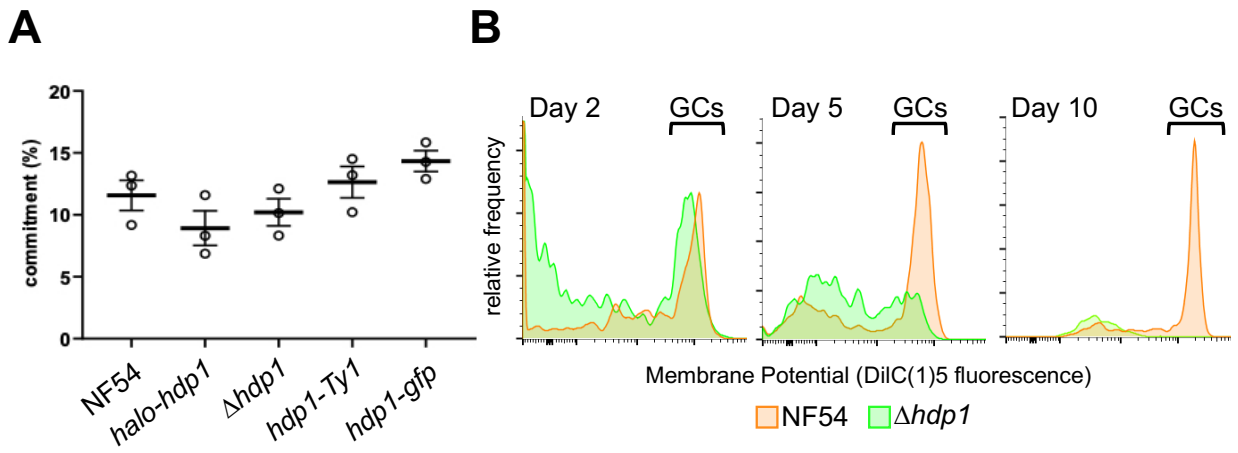

Fig. S3: Loss of HDP1 does not alter the sexual commitment frequency or Stage I gametocyte viability. (A) The sexual commitment frequency (day 5 gametocytes per day 1 ring stages) is not significantly affected in *halo-hdp1* and Δ*hdp1* parasites. n=3 **(B)** Mitochondrial membrane potential of gametocytes (GCs), as measured by DilC(1)5 staining, indicates similar viability on day 2, but not days 5 or 10, for NF54 (orange) and Δ*hdp1* (green) gametocytes. Representative of n=2.

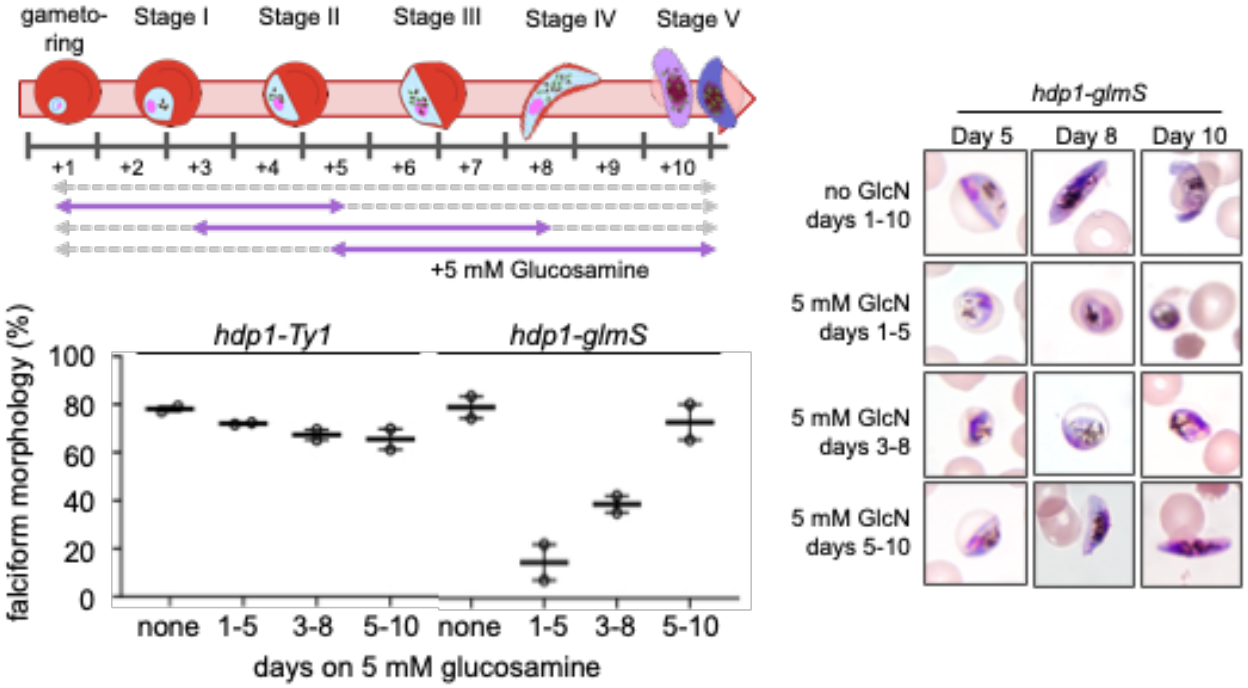

fig. S4: Inducible knockdown of HDP1 reduces gametocyte maturation in early but not late gametocytes. Representative morphology (right) of *hdp1-glmS* gametocytes in response to 5 mM glucosamine on days 1-10, 3-8, 5-10, or in the absence of glucosamine. Percentage of falciform gametocytes on Day 10 in response to 5 mM glucosamine on days 1-10, 3-8, 5-10, or in the absence of glucosamine for *hdp1-Ty1* or *hdp1-glmS* parasites (bottom). mean±s.e.m of n=2.

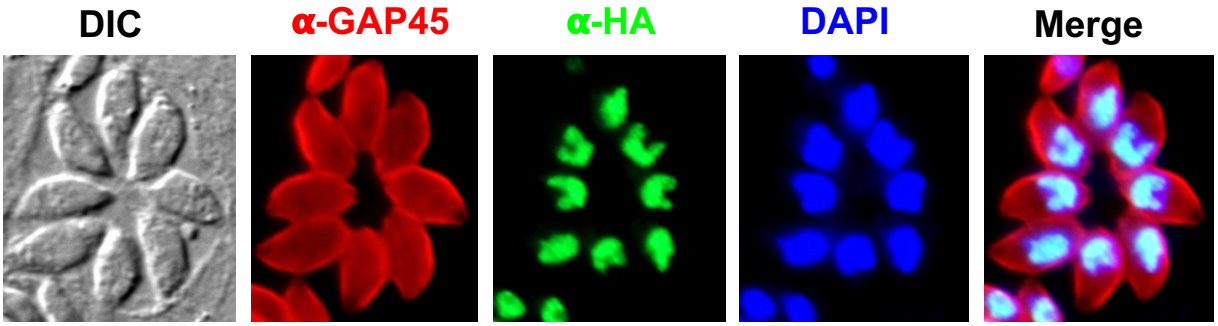

fig. S5: Immunofluorescence microscopy localizes the HA-tagged ortholog TGME49_233160 to the nucleus of *Toxoplasma gondii* tachyzoites.

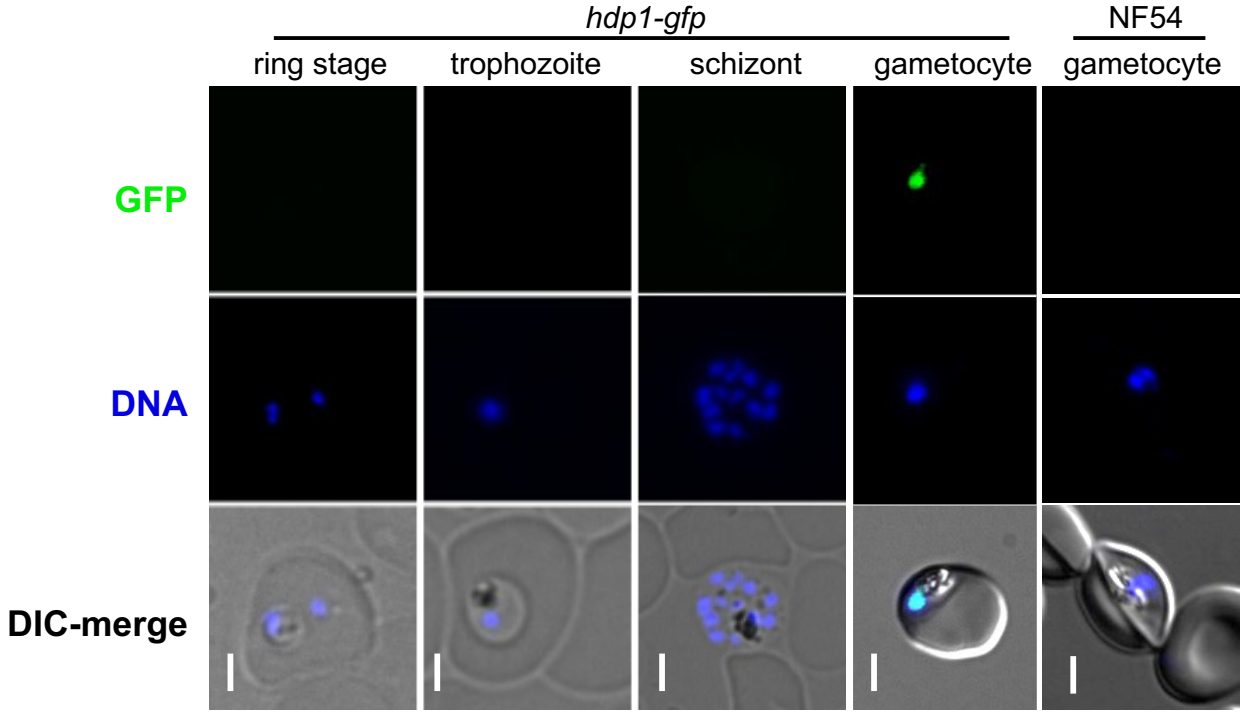

fig. S6: HDP1-GFP localized to the nucleus of hdp1-gfp gametocytes. No signal was observed in *hdp1-gfp* asexual blood stages or gametocytes of the untagged NF54 parent line. Scale bar is 3 microns. Exposure and brightness/contrast settings are uniform across the images shown. Representative of n=3.

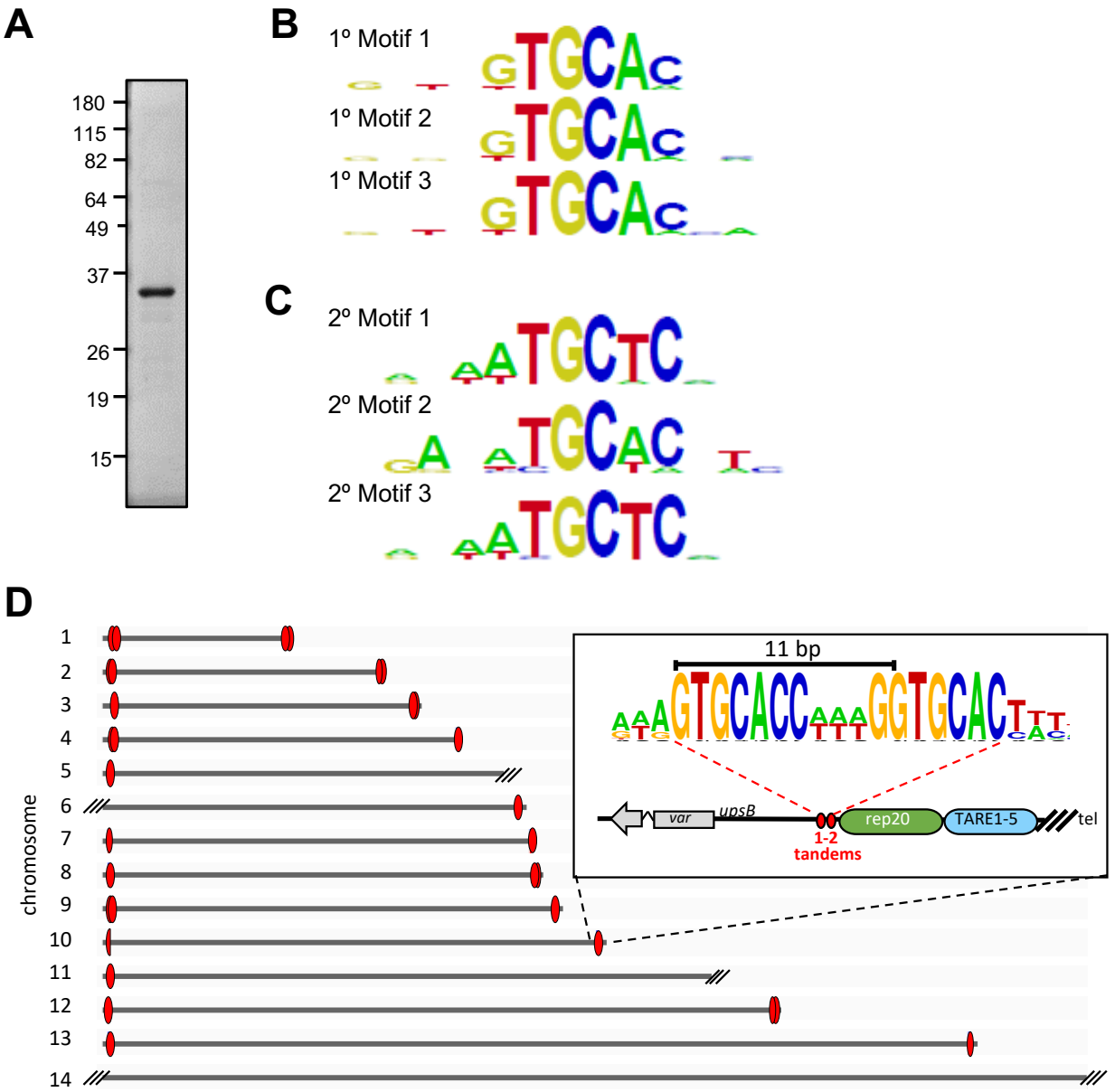

fig. S7: The HDP1 DNA-binding domain dimers recognize a tandem GC-rich motif found at chromosome ends. (A) Coomassie stain of GST-HDP1-DBD used for PBM analysis. **(B)** The three most highly enriched DNA motifs for the GST-HDP1 DBD domain on the protein-binding microarray. **(C)** Top three secondary motif hits after removal of primary hits. **(D)** Tandem HDP1 motifs with a 5-6bp spacer occur exclusively in arrays of 1-2 copies within either 7 bp or 28 bp of the last *rep20* sub-telomeric repeats at all full-length chromosome ends. The motif on the left end of chromosome 10 is degenerate but its positioning is conserved. Truncation by telomere healing resulted in the loss of rep20 adjacent region at five other chromosome ends.

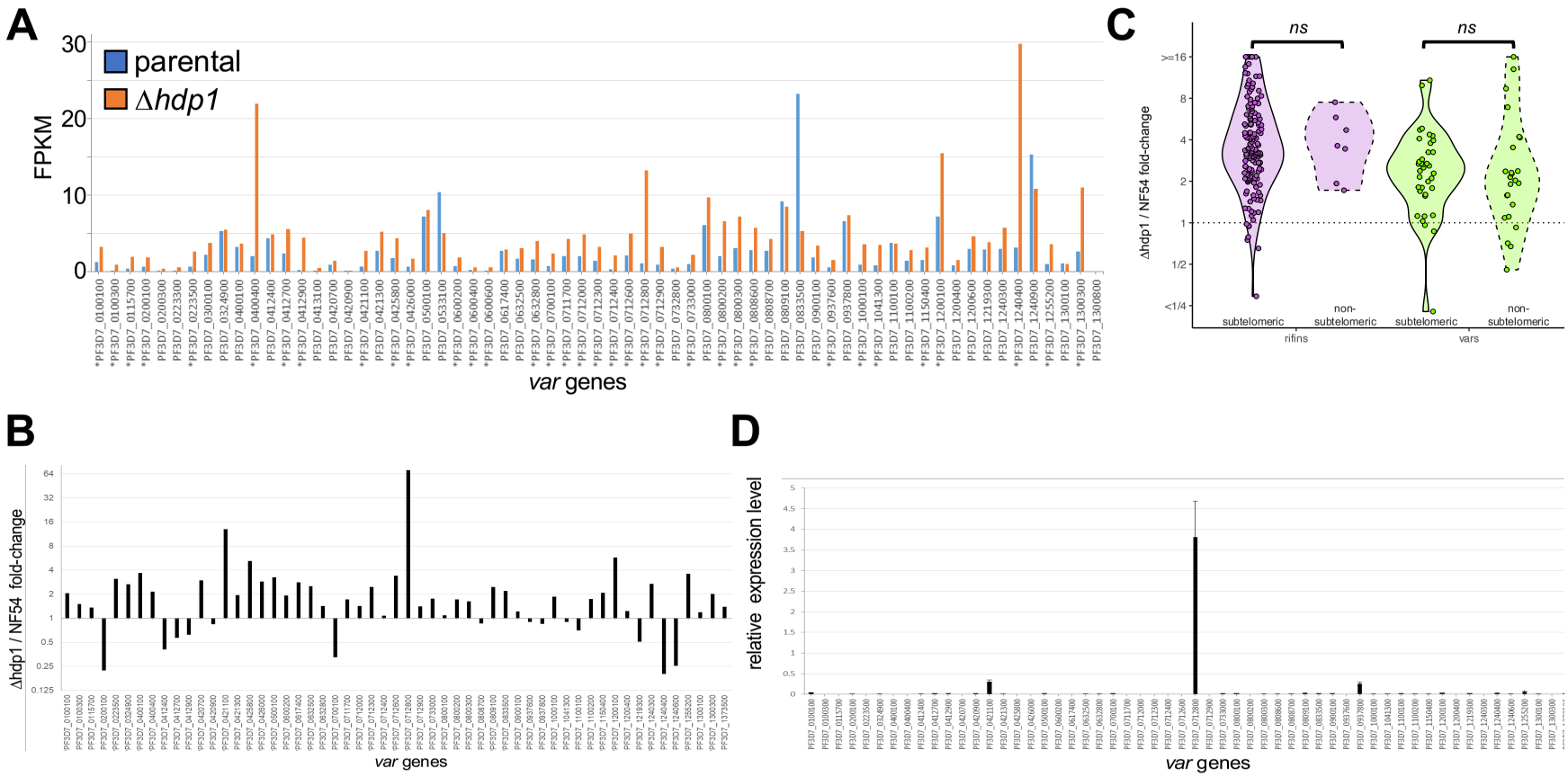

fig. S8: *var* gene expression is altered in early Δ*hdp1* gametocytes but not asexual blood stages. (A) Normalized abundance of reads uniquely mapping to *var* genes in day 2 gametocytes from Δ*hdp1* (orange) or parental NF54 parasites (blue). Significantly upregulated genes are marked with asterisks. (n=2) **(B)** qRT-PCR confirmation of *var* gene up-regulation in Δ*hdp1* vs NF54 day 2 gametocytes. **(C)** Upregulation of *rifin* (purple) and *var* genes (green) in Δ*hdp1* day 2 gametocytes is independent of chromosomal position in subtelomeric or non- subtelomeric heterochromatin clusters. **(D)** qRT-PCR analysis of *var* transcript abundance (normalized seryl tRNA synthetase expression) in Δ*hdp1* asexual ring stages found expression a single dominant var as expected indicating that mutually exclusive expression remains unchanged in asexual stages. (n=2)

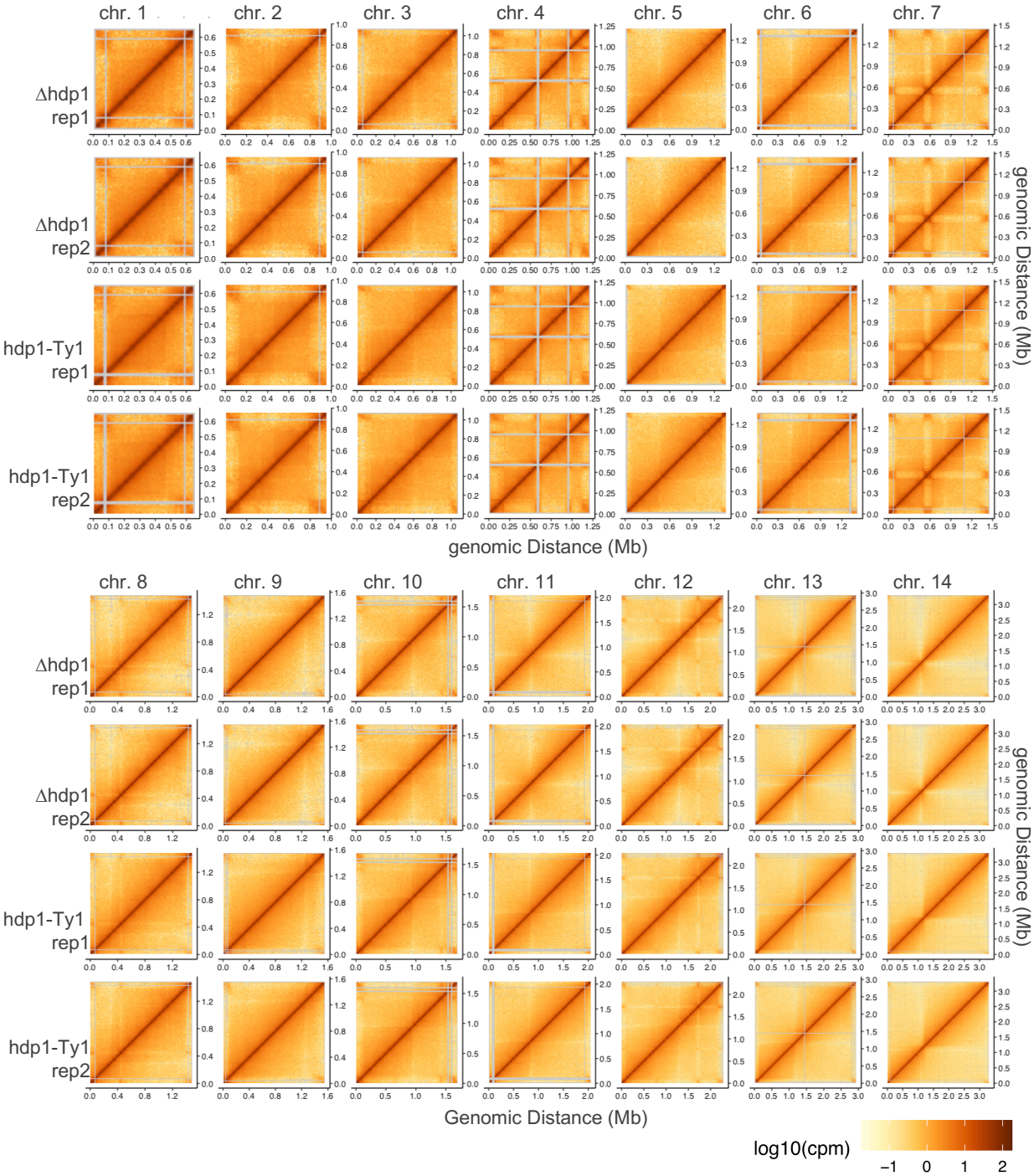

fig. S9: Intra-chromosomal HiC maps from Δ*hdp1* and *hdp1-Ty1* of day 2 gametocytes. Color scale indicates the normalized log10 interaction frequency (counts per million). Gray bars indicate bins of high-homology that could not be uniquely mapped with high confidence.

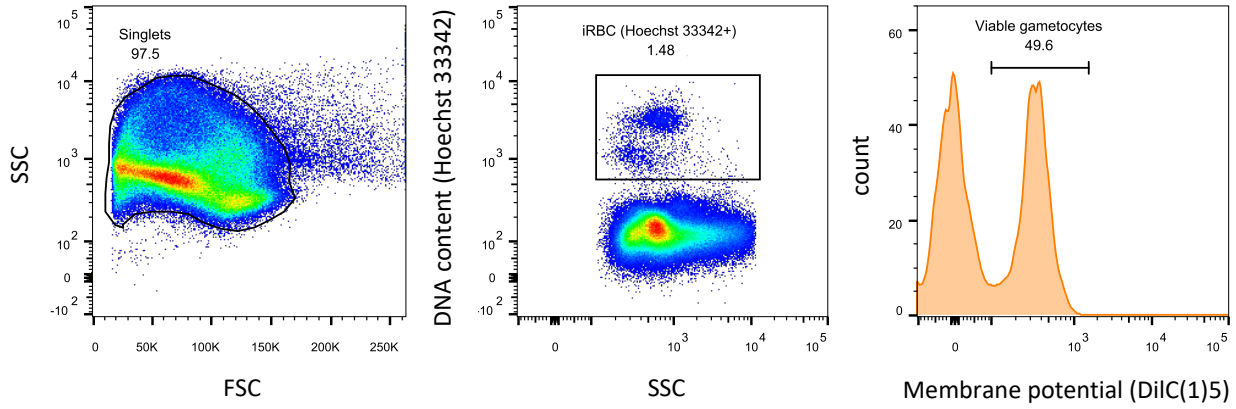

fig. S10: Gating schema for viable gametocytes. Populations were gated for single cells based on forward (FSC) and side scatter (SSC). Viable gametocytes were identified based on DNA content and mitochondrial membrane potential based on Hoechst33342 and DilC(1)5 staining.