nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-021-01045-0

Check for updates

The transcriptional regulator HDP1 controls expansion of the inner membrane complex during early sexual differentiation of malaria parasites

Riward A. Campelo Morillo[1], Xinran Tong[1], Wei Xie[2], Steven Abel[3], Lindsey M. Orchard[4], Wassim Daher[5], Dinshaw J. Patel[2], Manuel Llinás[4,6], Karine G. Le Roch[3] and Björn F. C. Kafsack[1]✉

Transmission of *Plasmodium falciparum* and other malaria parasites requires their differentiation from asexual blood stages into gametocytes, the non-replicative sexual stage necessary to infect the mosquito vector. This transition involves changes in gene expression and chromatin reorganization that result in the activation and silencing of stage-specific genes. However, the genomes of malaria parasites have been noted for their limited number of transcriptional and chromatin regulators, and the molecular mediators of these changes remain largely unknown. We recently identified homeodomain protein 1 (HDP1) as a DNA-binding protein, first expressed in gametocytes, that enhances the expression of key genes critical for early sexual differentiation. The discovery of HDP1 marks a new class of transcriptional regulator in malaria parasites outside of the better-characterized ApiAP2 family. Here, using molecular biology, biochemistry and microscopy techniques, we show that HDP1 is essential for gametocyte maturation, facilitating the necessary upregulation of inner membrane complex components during early gametocytogenesis that gives *P. falciparum* gametocytes their characteristic shape.

To complete its life cycle, *Plasmodium falciparum*, the most widespread and virulent of the human malaria parasites, 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. Infection of the vector requires asexual blood stages to differentiate into non-replicating, male and female gametocytes that can infect the mosquito once taken up during a blood meal.

All differentiation requires the repression and activation of genes that underlie the specific phenotypes of origin and destination cell types, respectively. To ensure complete commitment to one cell type or another, these transitions often involve a bistable switch that controls the activity of a single master regulator at the top of the transcriptional cascade that underlies the differentiation programme[1–4]. On activation, the master regulator initiates the broader downstream changes in gene expression by altering the expression of additional transcriptional regulators and changing their access to cell-type-specific promoters via chromatin reorganization.

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 trigger a transcriptional feedback that drives its expression to high levels, thereby locking cells into the sexual differentiation gene expression programme[5,9,10]. Under conditions that impair heterochromatin maintenance, this feedback loop is activated in a larger fraction of cells, thus increasing the frequency of sexual differentiation[8,11,12].

Commitment to this sexual differentiation programme involves substantial changes in gene expression and redistribution of heterochromatin during the early stages of gametocyte development[13,14]. While AP2-G is critical for the initiation of sexual differentiation, it is expressed only during a small window that begins with sexually committed schizonts and ends after the first 48 h of gametocyte development[10,15]. This means that many of the expression changes during the prolonged process of gametocyte development depend on additional transcriptional regulators, and is consistent with our previous observations that AP2-G upregulates a number of putative transcription factors and chromatin remodelling enzymes[9,10].

While most species of malaria parasites form spherical gametocytes that mature in 2–6 days, sexual differentiation in *P. falciparum* takes 12–14 days and produces gametocytes with the characteristic falciform morphology that gives the parasite its name. Unsurprisingly, this prolonged maturation is accompanied by a wide array of gene expression changes[16–19]. However, relatively little is known about the transcriptional regulators downstream of AP2-G that mediate these changes. Compared to other single-celled eukaryotes, DNA-binding proteins are notably under-represented in the genomes of malaria parasites[20]. Most belong to the ApiAP2 family[20], but only a small number have been shown to function

[1]Department of Microbiology and Immunology, Weill Cornell Medicine, New York, NY, USA. [2]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [3]Department of Molecular, Cell and Systems Biology, University of California Riverside, Riverside, CA, USA. [4]Department of Biochemistry and Molecular Biology, and Huck Center for Malaria Research, Pennsylvania State University, University Park, PA, USA. [5]Dynamique des Interactions Membranaires Normales et Pathologiques, UMR5235 CNRS, INSERM, Université de Montpellier, Montpellier, France. [6]Department of Chemistry, Pennsylvania State University, University Park, PA, USA. ✉e-mail: bjk2007@med.cornell.edu

specifically during gametocyte development of *P. falciparum*[21]. In the rodent malaria parasite *Plasmodium berghei*, PbAP2-G2 functions as a transcriptional repressor of asexual-specific gene expression[6,22,23] while PbAP2-FG (PyAP2-G3 in *Plasmodium yoelii*) was shown to mediate upregulation of female-specific transcripts[24,25].

In this study, we identify HDP1, a previously uncharacterized DNA-binding protein that is absent from asexual blood stages and is first expressed during early sexual differentiation. The development of HDP1-deficient gametocytes arrests at the transition from stage I to stage II, and ends in loss of viability. Analysis of gene expression and HDP1 binding shows that this protein functions as a positive transcriptional regulator of genes essential for gametocyte development, including those critical for the expansion of the inner membrane complex (IMC) that gives *P. falciparum* gametocytes their characteristic shape.

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 3,078-amino acid protein with a C-terminal homeodomain-like domain, containing a helix-turn-helix structural motif commonly involved in DNA binding (Extended Data Fig. 1a)[26]. Syntenic orthologs of *hdp1* could be readily identified in other malaria parasites with homology to the homeodomain-like domain also found among the coccidia but apparently absent from other apicomplexan clades (Extended Data Fig. 1b). Analysis of *hdp1* transcript levels in *P. falciparum* blood stages by quantitative PCR with reverse transcription (qRT–PCR) detected only minimal expression in cultures of asexual blood stages, which always contain small numbers of gametocytes, but showed substantial upregulation during the early stages of gametocytogenesis (Extended Data Fig. 1c). AP2-G, the transcriptional master switch that controls the initiation of the sexual differentiation gene expression programme[5,9,10,27], binds at two sites located upstream of *hdp1* in early gametocytes[10], consistent with our hypothesis that AP2-G activates additional regulators of gene expression during early gametocytogenesis.

To determine the subcellular localization of HDP1, we inserted an N-terminal Halo tag at the endogenous *hdp1* locus (Extended Data Fig. 2a) to avoid possible interference with the predicted DNA-binding domain (DBD) located at the very C terminus. As expected, based on transcript abundance data, 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 at the HDP1 C terminus with either green fluorescent protein (GFP) or a triple Ty1 epitope tag (*hdp1-gfp* and *hdp1-Ty1*; Extended Data Fig. 2b,c) yielded parasite lines that produce gametocytes indistinguishable in numbers and morphology from the wild-type parent (Fig. 1a,b), despite the proximity to the putative DBD. To test whether N-terminal tagging would result in loss of HDP1 function, we generated a Δ*hdp1* knockout line for comparison by replacing 1.4 kb at the 5' end of the *hdp1* locus with a selectable marker cassette (Extended Data Fig. 2d). 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 (Extended Data Fig. 3a) but are unable to complete gametocyte development (Fig. 1a–c).

Because the length of the *hdp1* coding sequence rendered genetic complementation infeasible, we generated inducible HDP1 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 (*hdp1-glmS*; Extended Data Fig. 2e)[28]. In the absence of glucosamine, the resulting *hdp1-glmS* gametocytes expressed HDP1 protein at levels comparable to *hdp1-Ty1* parasites lacking the ribozyme, and produced gametocytes indistinguishable from wild type in both number and morphology (Extended Data Fig. 4). 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 the number of falciform gametocytes by 80% in *hdp1-glmS* parasites, recapitulating the phenotype of lines *halo-hdp1* and Δ*hdp1* (Fig. 1g,h). Since *hdp1* transcript levels remain relatively constant during gametocyte development (Extended Data Fig. 1c), we wanted to test whether *hdp1* transcripts were required throughout gametocyte maturation. Knockdown of *hdp1* levels with glucosamine reduced gametocyte maturation before day 5 but not thereafter, indicating that sufficient HDP1 protein had been produced by that time to support gametocyte maturation (Extended Data Fig. 4).

HDP1 is a chromatin-associated nuclear protein expressed in gametocytes. An earlier study reported a variety of subcellular localizations based on antibodies raised against a low-complexity region of HDP1 (Extended Data Fig. 1a), 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 (Extended Data Fig. 1a) and localization of the *Toxoplasma gondii* ortholog to the nucleus of tachyzoites (Extended Data Fig. 5). To resolve this apparent disagreement, we carried out live-cell and immunofluorescence microscopy of *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. According to its corresponding author, the antisera used in that study are unfortunately no longer available, precluding a direct comparison. HDP1 protein was undetectable in asexual blood stages by both microscopy and immunoblotting (Extended Data Fig. 6 and Fig. 2c), but showed increasing expression from day 2 of gametocytogenesis (stages I–II) onward, reaching maximal levels by day 5 (stage III) that remained steady until day 8 (stage IV) (Fig. 2c). Analysis of subcellular compartments from day 5 *hdp1-Ty1* gametocytes found HDP1 almost exclusively in the nuclear fraction, with about 70% resistant to solubilization of up to 600 mM NaCl, indicating a tight association with chromatin (Fig. 2d) and validating HDP1 as a nuclear protein.

Loss of HDP1 leads to dysregulation of gene expression in early gametocytes. Based on the nuclear localization and chromatin binding capacity of HDP1, we wanted to test whether it plays a role in the regulation of gene expression during early gametocytogenesis. To identify changes in expression that may be responsible for the aberrant development of Δ*hdp1* gametocytes, rather than looking at the consequences of the subsequent loss of viability, we decided to analyze the transcriptome of Δ*hdp1* and parental NF54 on day 2 of gametocytogenesis, when HDP1 is first detectable in wild-type gametocytes but before any change in viability or morphology occurs in Δ*hdp1* gametocytes. (Fig. 1d, Extended Data Fig. 3b and Datasets 1 and 2). Global comparison of transcript abundances found that most genes were expressed at similar levels, including canonical markers of early gametocytes such as *pfs16* and *gexp5* (Fig. 3a). This confirms our earlier observation that HDP1-deficient parasites initiate gametocyte development at wild-type rates and that changes in gene expression are not due to change in viability at this point. However, when compared to the parent line, Δ*hdp1* gametocytes had significant reductions in the transcript levels of 156 genes and increased levels of 103. Reassuringly, *hdp1* showed the great-

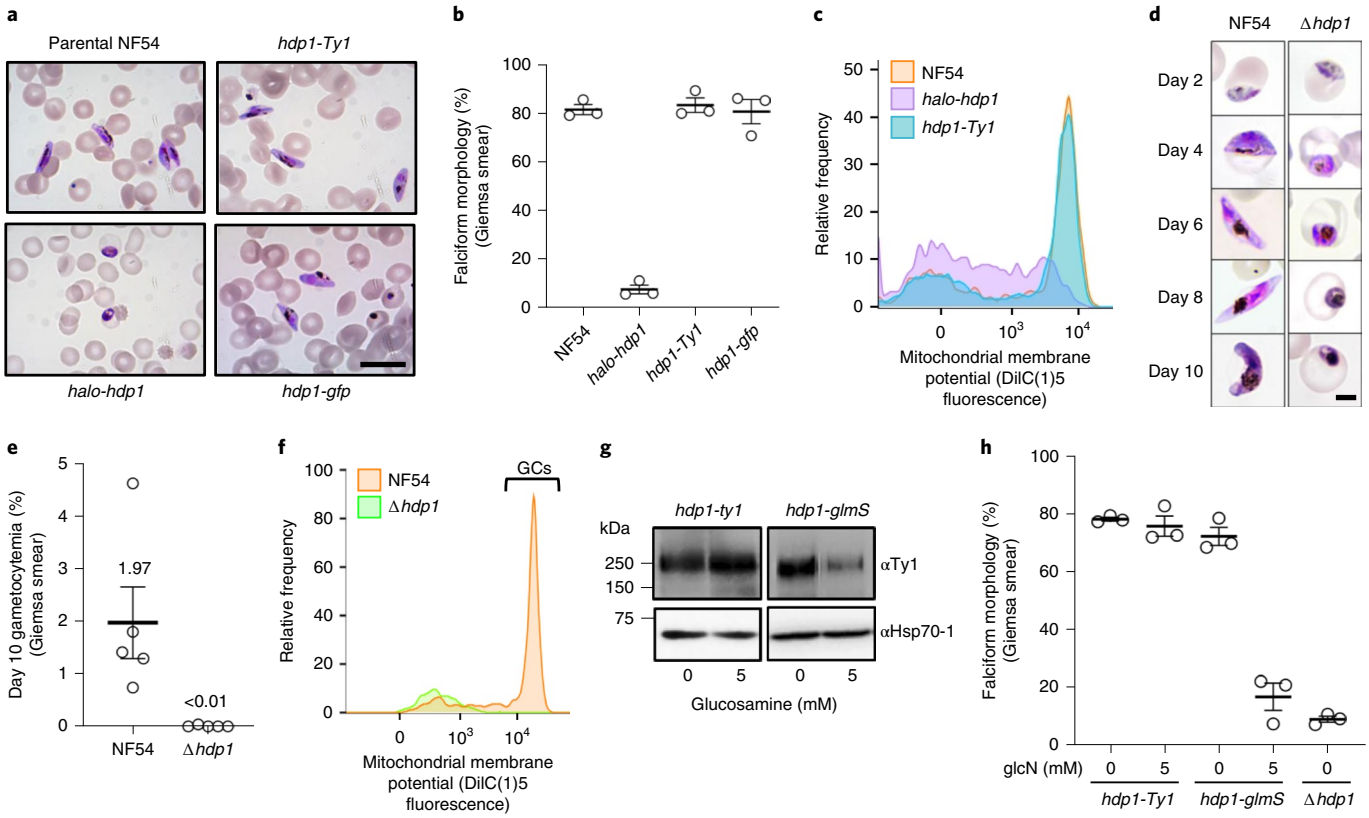


Fig. 1 | Loss of HDP1 function disrupts gametocyte maturation. a–c, N-terminal tagging of endogenously encoded HDP1 (*halo-hdp1*) blocked maturation of gametocytes (GCs), while C-terminal tagging of the endogenous locus with either GFP (*hdp1-gfp*) or a triple Ty1 (*hdp1-Ty1*) epitope had no effect on gametocyte morphology or viability. All images and plots are representative of $n = 3$ biologically independent samples. **a**, Light microscopy images of gametocytes on day 5 of development. Scale bar, 20 µm. **b**, Dot plot showing percentage of falciform day 5 gametocytes. Data are presented as mean values ± s.e.m. **c**, Flow cytometry plots showing gametocyte 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. **d**, Light microscopy images of development progression of gametocytes (NF54 versus Δ*hdp1*) from day 2 until day 10 of maturation. Scale bar, 3 µm. **e**, Dot plot showing gametocytemia of NF54 and Δ*hdp1* lines on day 10 of gametocyte maturation. Data are presented as mean values ± s.e.m. of $n = 5$ biologically independent samples. **f**, Flow cytometry plot showing viability on day 10 of gametocyte maturation. Data are representative of $n = 3$ biologically independent samples. **g**, Glucosamine-inducible knockdown of HDP1 in *hdp1-Ty1* and *hdp1-glmS* day 5 gametocytes. Representative of $n = 3$ biologically independent samples. **h**, Percentage of falciform day 5 gametocytes in response to 5 mM glucosamine (glcN). Data are presented as mean values ± s.e.m. of $n = 3$ biologically independent samples.

est decrease in expression. Gene set enrichment analysis (GSEA) found that transcripts encoding components of the IMC were significantly over-represented among downregulated genes (Fig. 3b,c and Dataset 3), while transcripts from heterochromatin-associated, multicopy gene families were significantly over-represented among upregulated genes (Fig. 3d), including members of gene families *var*, *rifin*, *stevor* and *PHISTa/b/c* (Fig. 3e). Several of these heterochromatin-silenced families exhibit transcriptional variation due to expression switching between members, but this does not lead to the broad upregulation across all family members we observed here. Due to their general lack of expression in wild-type cells, the observed fold changes for these genes were often substantial and the absolute increase was generally small but, nevertheless, significantly above levels in wild-type cells.

HDP1 recognizes a GC-rich DNA motif in vitro. Since HDP1 is an integral component of chromatin and has homology to homeo-like domains that typically mediate protein–DNA interactions, we evaluated whether the HDP1 DBD recognizes DNA in a sequence-specific manner using a protein-binding microarray (PBM). Recombinant HDP1 DBD was highly enriched on probes containing the palindromic hexamer GTGCAC (Fig. 4a and Extended Data Fig. 7). Since homeodomains often bind DNA as dimers[26,30], we carried out isothermal titration calorimetry to measure the interaction of HDP1 with double-stranded DNA containing a tandem motif with a 5-base-pair (bp) spacer that places the motifs one helical turn apart. The results indicated that binding was saturated at a 2:1 protein/DNA molar ratio with a dissociation constant of 2.8 µM, indicating that in vitro DNA recognition by the HDP1 DBD occurred primarily as a dimer (Fig. 4b). We subsequently confirmed dimeric binding using DNA gel-shift assays with double-stranded DNA probes containing either no match, a single binding motif or a tandem motif. When compared to the tandem-motif probe, the gel-shift of the single-motif probe was substantially weaker but identical in size (Fig. 4c), again consistent with DNA recognition as a dimer even when only a single motif is present.

HDP1 binds GC-rich motifs upstream of a subset of gametocyte-expressed genes in vivo. Since our experiments clearly show that HDP1 can bind DNA and is tightly associated with nuclear chromatin (Fig. 2d), we set out to determine HDP1 genome-wide distribution. To do this we performed chromatin immunoprecipitation sequencing (ChIP–seq) with anti-GFP antibodies on nuclei from *hdp1-gfp* and *hdp1-Ty1* gametocytes on day 5 of development, with the latter serving as a negative control. We identified 1,003 regions significantly enriched for HDP1-GFP binding

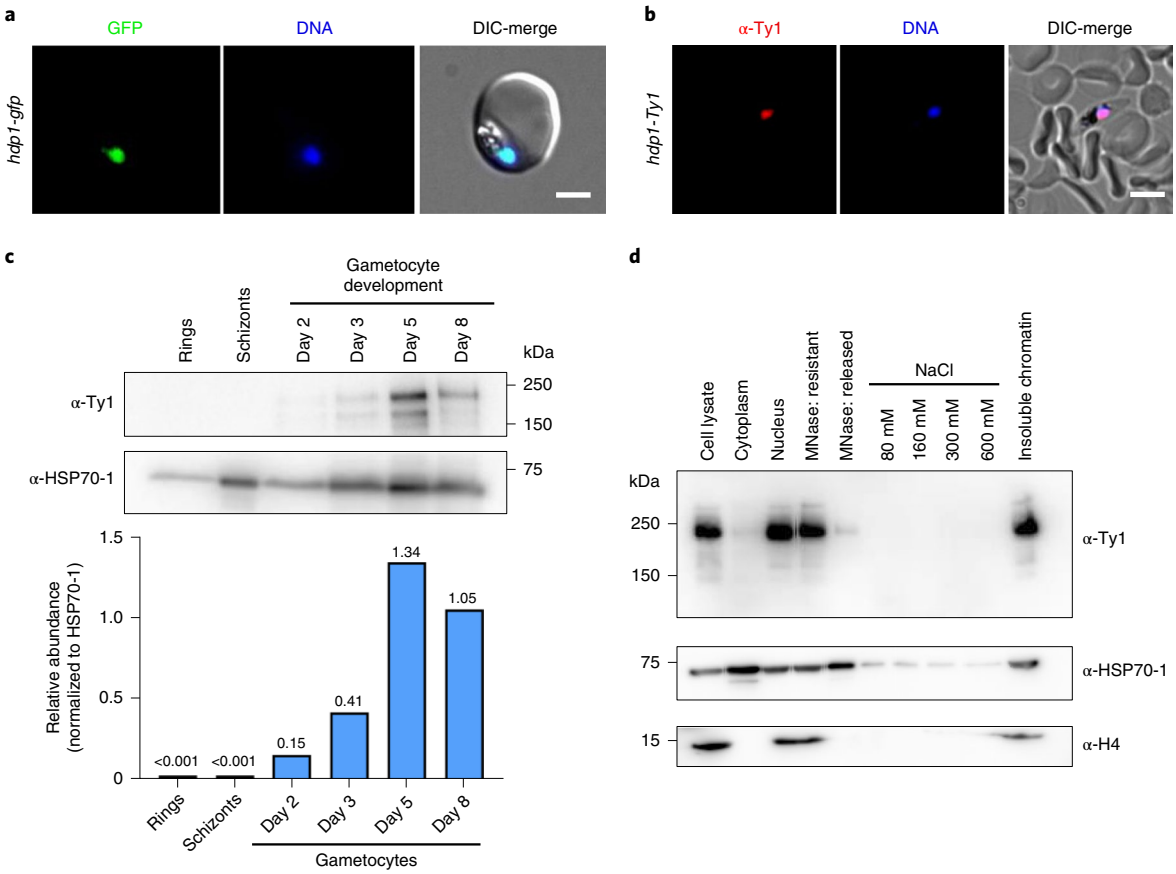


Fig. 2 | HDP1 is a chromatin-associated protein expressed in gametocytes. a, Live-cell fluorescence microscopy of *hdp1-gfp* gametocytes on day 5 of maturation stained with the DNA dye Hoechst 33342 (blue). DIC-merge, differential interference contrast merged with Hoechst and GFP fluorescence. Scale bar, 3 µm. Representative of $n=2$ biologically independent samples. **b**, Immunofluorescence microscopy of *hdp1-Ty1* gametocytes on day 5 of maturation costained with anti-Ty1 antibodies (red) and Hoechst 33342 (blue). Scale bar, 5 µm. Representative of $n=2$ biologically independent samples**. c**, Immunoblotting of parasite lysates in asexual stages and during gametocyte maturation shows expression of HDP1 during the stages of gametocytogenesis. Numbers above bars in the bottom panel represent relative abundance of HDP1-Ty1 normalized to HSP70-1. Representative of $n=3$ biologically independent samples. **d**, Immunoblot of cytoplasmic and nuclear extracts of *hdp1-Ty1* gametocytes on day 5 of maturation, stained with antibodies against the Ty1 epitope tag, histone H4 and HSP70-1. Representative of $n=2$ biologically independent samples.

containing 1,188 binding summits (Dataset 4). Most of these summits (85%) occur within the upstream regions of genes, with the greatest enrichment occurring just upstream of the annotated transcription start site (TSS) (Fig. 4d); 59% of significantly downregulated genes were found to have upstream HDP1 binding regions, compared with only 18% of those upregulated in Δ*hdp1* gametocytes (Fig. 4g). The prevalence of HDP1 binding near TSSs of genes with reduced expression in gametocytes lacking HDP1 strongly suggests that it functions as a transcriptional activator.

To determine whether HDP1 binding in these regions involved recognition of a specific DNA sequence, we carried out motif enrichment analysis of the 100 bp flanking the HDP1 binding summits. This identified two highly enriched sequence motifs, referred to hereafter as Motifs A and B (Fig. 4e). Motif A (GTGCACAC, enrichment $P=1\times10^{-138}$) is a GC-rich 8-mer that closely matched the motif obtained by PBM (Fig. 4a). Motif B ([GTA]TGTA[CT][GA]TAC, enrichment $P=1\times10^{-122}$) is a 11-mer with greater sequence flexibility that differs substantially from the PBM motif. However, the sequence space covered by PBM allows identification of only 8-bp motifs or shorter. Moreover, ChIP–seq of DNA-binding protein frequently identifies new binding motifs, possibly as the result of interaction with other proteins or domains. Instances of Motifs A and B could be identified within 61.8% and 47.8% of HDP1-bound regions, respectively, with 78.1% of HDP1 peaks having at least one motif.

The instances within HDP1-bound regions of both Motif A and B occurring were centred on the ChIP–seq summits (Fig. 4f), confirming their recognition by HDP1 in vivo. The genome-wide number of these motifs exceeds those found to be occupied by HDP1, indicating that HDP1 is probably recruited to the occupied subset based on interaction with other proteins or excluded from others.

A comparison of expression levels of genes with upstream HDP1-bound regions to all genes found that HDP1-bound genes are significantly more highly expressed (two-sided Wilcoxon rank test $P=3.9\times10^{-13}$). HDP1 binding occurred across a wide range of expression levels, but was less frequent upstream of silent or lowly expressed genes (Fig. 4h). Interestingly, while the HDP1 DBD preferentially recognized a tandem motif in vitro, virtually all genome-wide binding sites contain only a single motif. Furthermore, the only perfected instances of the tandem motif, which are found at the centromeric end of the subtelomeric repeats, showed no significant enrichment, possibly because these regions are not accessible for binding due to heterochromatin formation.

HDP1 enhances the expression of genes encoding IMC components and required for IMC expansion in early gametocytes. The failure of Δ*hdp1* gametocytes to elongate from spherical stage I gametocytes into the oblong stage II (Fig. 1d) was strikingly similar to the phenotype described for knockdown of PhIL1, an IMC

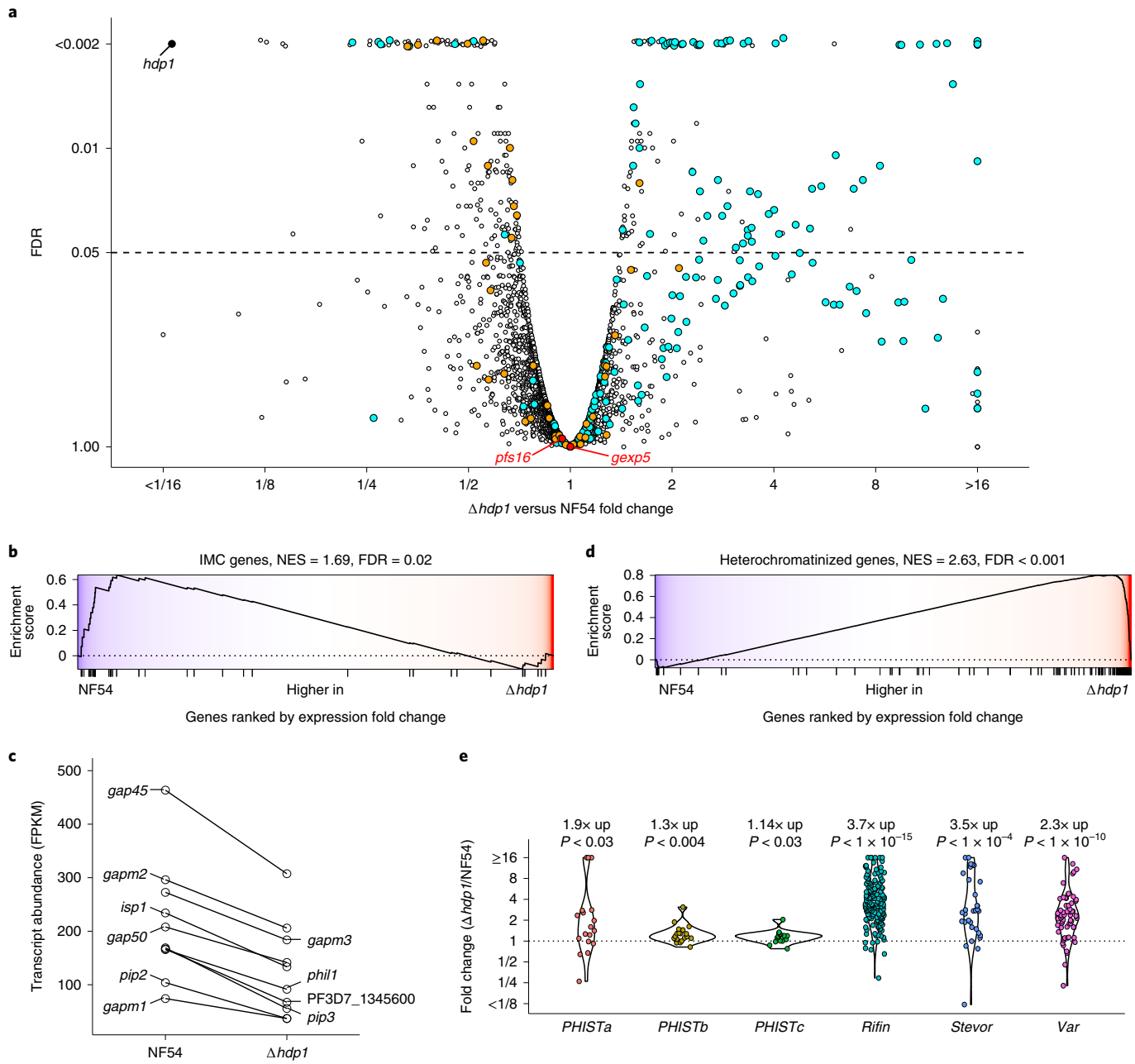


Fig. 3 | Disruption of HDP1 results in leaky expression of heterochromatin-associated genes and reduced expression of IMC genes in early gametocytes. a, Genome-wide comparison of differential gene expression in Δ*hdp1* and parental NF54 gametocytes on day 2 of gametocytogenesis (stage I, $n = 2$ biologically independent samples). *hdp1* (black), heterochromatin-associated genes (cyan), IMC genes (orange) and the two canonical early gametocyte markers, *pfs16* and *gexp5* (red), are highlighted. The horizontal line indicates the FDR = 0.05 significance cut-off. **b**,**c**, GSEA indicated significant downregulation of IMC genes, as determined by enrichment score (**b**) and transcript abundance (**c**). **d**,**e**. GSEA indicated global upregulation of heterochromatin-associated genes, as determined by enrichment score (**d**) and fold change for Δ*hdp1* over NF54 (**e**). **b**,**d**, Genes with detectable expression are ranked along the *x* axis by expression fold change, from highest in NF54 to highest in Δ*hdp1*, with tick marks indicating the ranking of IMC or heterochromatinized genes, respectively. The *y* position of the line indicates the enrichment score as a sliding window for genes highest in NF54 to those highest in Δ*hdp1*. The overall normalized enrichment score (NES) and FDR for each gene set are shown above each plot. **e**, Geometric mean fold changes and *P* values (two-sided, one-sample *t*-test) are indicated.

protein required for the expansion of the IMC in early *P. falciparum* gametocytes[31]. Given that genes encoding PhIL1 and other IMC components were highly enriched among genes with reduced expression in HDP1-deficient gametocytes, we examined HDP1 binding at these loci in greater detail. We detected HDP1 binding upstream of 11 out of the 13 IMC genes with reduced expression in Δ*hdp1* (Fig. 5a and Extended Data Fig. 8), and confirmed significant downregulation for all 11 genes tested by qRT–PCR (Fig. 5b). Using antibodies against PhIL1 (generous gift of P. Malhotra), we found that PhIL1 expression was greatly reduced in Δ*hdp1* gametocytes (Fig. 5c). PhIL1 expression and extension of the IMC during early gametocytogenesis were also clearly impaired on knockdown of HDP1 (Fig. 5d,e), confirming that both PhIL1 expression in early gametocytes and IMC extension require HDP1.

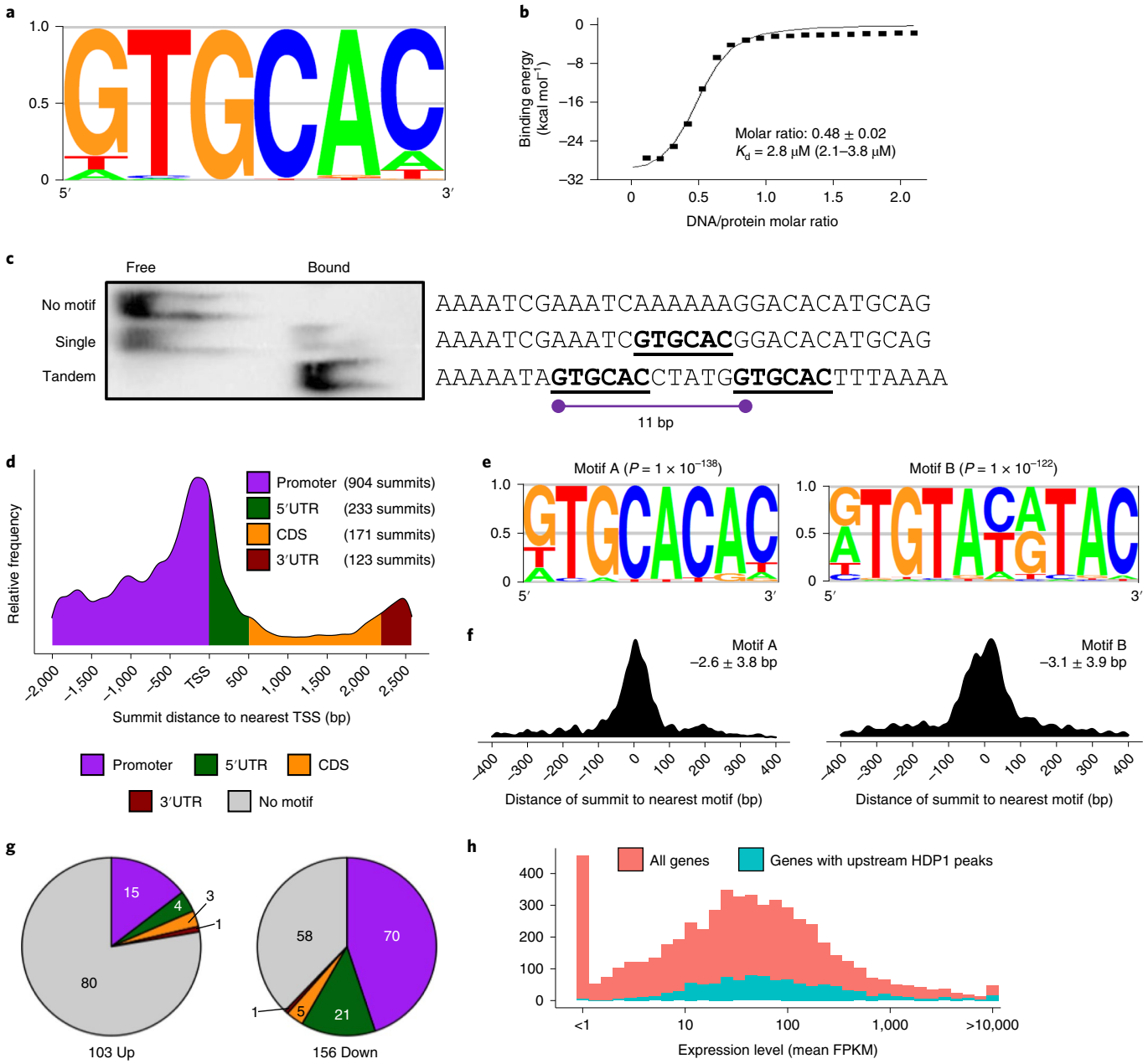


Fig. 4 | HDP1 binds near the TSS of genes expressed in early gametocytes. a, Maximum enrichment DNA motif for the GST-HDP1 DBD as determined by PBM. The *y* axis shows the fraction of each base at that position. **b**, Isothermal calorimetry indicates that the HDP1 DBD domain recognizes DNA as a dimer. Data are representative of $n = 2$ independent experiments. **c**, Optimal gel-shift was observed for probes containing a tandem motif with a 5-bp spacer compared with those with either a single or no motif. Data are representative of $n = 3$ independent experiments. **d**, Distribution of gene-associated HDP1 ChIP-seq binding sites ($n = 2$ biologically independent samples). Relative distance to TSS is shown on the *x* axis. Length of gene features shown is the median length of those with HDP1 binding summits. Summits found within regions mapping to two adjacent genes (242/1,188 summits) were counted for both. CDS, nucleotide coding sequence. **e**, Highly enriched sequence motifs within 100 bp centred on HDP1-GFP ChIP-seq summits; *y* axis shows the fraction of each base at that position. **f**, Density graph of the distance between instances of Motif A (left) or Motif B (right) within HDP1-bound regions to the nearest ChIP-seq summit. Mean distance ± s.e.m. is also shown. **g**, Number of differentially regulated genes based on location of HDP1-bound regions. **h**, Histogram of mean expression levels in early gametocytes of NF54 for all genes (red) and for those with upstream (promoter or 5′ UTR) HDP1 motifs (teal). Note the lack of genes with HDP1 upstream peaks that have very low expression. Hoechst, Hoechst 33342 staining.

HDP1 binds upstream of its own locus and two ApiAp2 genes. Our data indicate that upstream binding of HDP1 promotes the expression of IMC genes necessary for gametocyte maturation. Intriguingly, we also found that the *hdp1* locus itself has multiple upstream HDP1 binding sites (Fig. 5f), pointing to the possibility of a transcriptional feedback loop that may sustain HDP1 expression in gametocytes. This could explain the progressive increase in HDP1 during gametocytogenesis and why the consequences of losing HDP1 expression become apparent when expression is still quite low (Fig. 2c). Additionally, two loci encoding gametocyte-expressed ApiAP2 proteins were found to have HDP1 binding peaks upstream (Fig. 5g), and their expression was significantly reduced in HDP1-deficient gametocytes. These include the gene encoding AP2-O2 which, in the later stages of gametocyto-

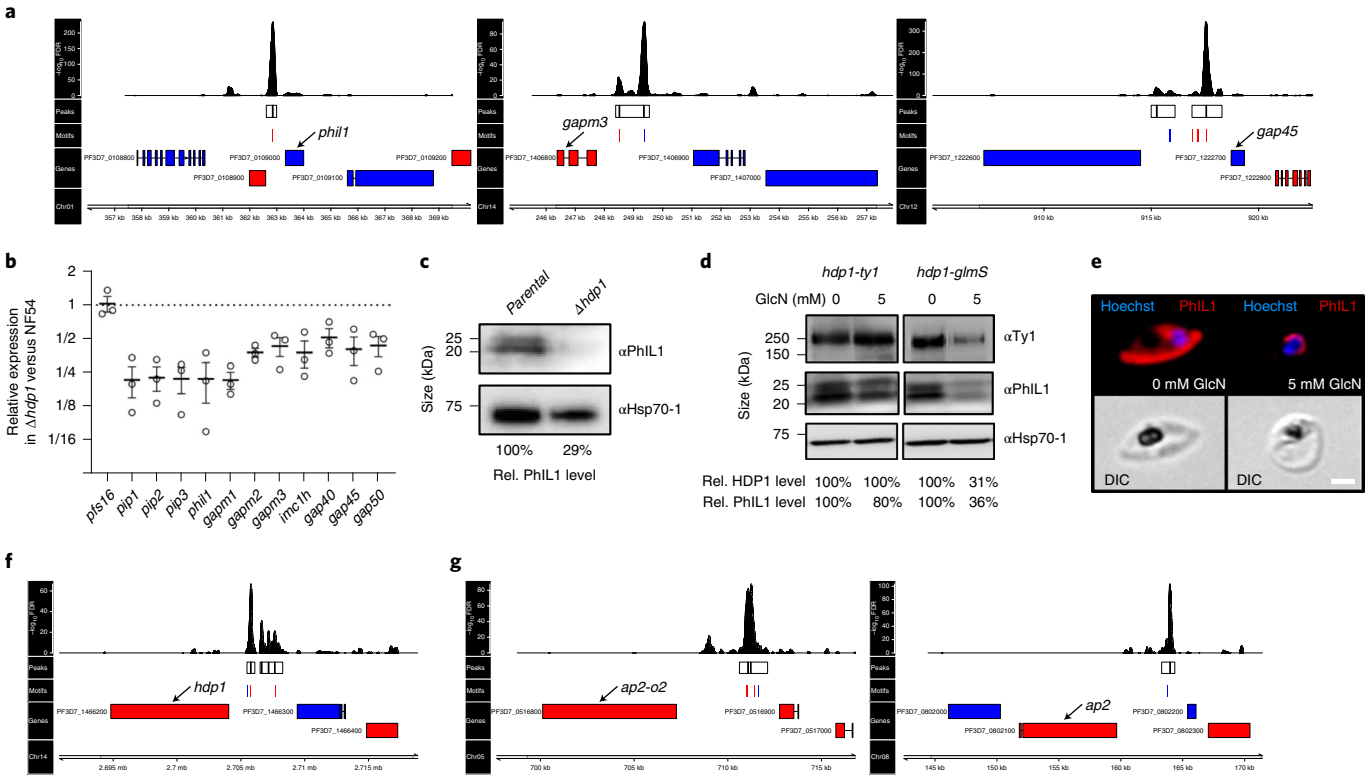


Fig. 5 | HDP1 is essential for expansion of the IMC in early gametocytes. a, Example of HDP1 binding sites upstream of genes encoding IMC proteins. Histogram track shows the significance of enrichment by position. Regions of significant enrichment are shown as boxes, with black vertical lines indicating peak summits within each peak. Instances of Motif A, Motif B and overlapping motifs within peaks are shown in red, blue and purple, respectively. Genes encoded in forward or reverse orientation are shown in blue or red, respectively. Combined estimate of $n=2$ biologically independent samples. **b**, Validation of downregulation of genes encoding IMC genes in HDP1 knockout parasites by qRT-PCR. Data are presented as mean values ± s.e.m of $n=3$ biologically independent samples**. c**, PhIL1 expression in parental and Δ*hdp1* gametocytes. Hsp70-1 abundance shown as the loading control. Representative result of $n=2$ biologically independent samples. **d**, HDP1 and PhIL1 protein levels in *hdp1-ty1* and *hdp1-glmS* day 5 gametocytes under 0 and 5 mM glucosamine. Hsp70-1 abundance shown as the loading control. Representative result of $n=2$ biologically independent samples. **e**, Immunofluorescence microscopy of PhIL1 distribution in day 5 gametocytes of *hdp1-glmS* under 0 and 5 mM glucosamine. Data are representative of $n=2$ biologically independent samples. Scale bar, 3 µm. **f**,**g**, HDP1 binding sites upstream of genes encoding HDP1 itself (**f**) and two ApiAP2 proteins (**g**). Tracks are the same as in **a**.

genesis, is required for the upregulation of transcripts essential for ookinete development.

Discussion

Our understanding of the regulatory machinery that underlies 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 switch of this developmental decision. However, it is becoming clear that the role of AP2-G is largely constrained to the initiation of the transcriptional programme that drives the nearly-2-week-long process of gametocyte development[15]. 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 showed that HDP1 is a nuclear DNA-binding protein that functions as a regulator of gene expression in early gametocytes and is essential for their development. Our experiments show that HDP1 is an integral component of chromatin in gametocytes and preferentially binds to a GC-rich motif near the transcription start site of target genes. Loss of HDP1 expression results in failure to upregulate a limited set of genes during the early stages of gametocytogenesis, most of which have upstream HDP1 binding sites, supporting the role of HDP1 as a positive transcriptional regulator. However, the expression of effected genes was substantially reduced but not completely lost in Δ*hdp1* gametocytes, indicating that upstream binding of HDP1 functions in concert with additional transcriptional regulators to achieve the necessary level of expression.

HDP1 bound upstream of most of the genes negatively impacted by *hdp1* disruption. Interestingly Motif A is similar to those recognized by ApiAP2 proteins SIP2 and API2-I in vitro[32–34], suggesting that DNA-binding specificity in vivo relies on regions beyond the DBD or is mediated through interaction with other proteins[35]. Motif instances were also found upstream of the *hdp1* locus, indicating the possibility of a positive transcriptional feedback loop that allows enhancement by HDP1 of its own expression and progressive increase in its protein levels. The role of HDP1 in regulating itself and other transcriptional regulators to advance or sustain the transcriptional programme underlying gametocytogenesis will certainly warrant additional investigation.

The development of gametocytes lacking HDP1 aborts just before the transition from stage I to II, which is marked by the onset of IMC elongation that gives the *P. falciparum* gametocyte its eponymous sickle shape. Instead, HDP1-deficient gametocytes remain spherical and lose viability over the next few days. This phenotype closely resembles one described following knockdown of the IMC

protein PhIL1 in early gametocytes. Indeed, we found that HDP1 binds upstream of *phil1* and other genes involved in expansion of the IMC in early gametocytes, and that HDP1 is required for full expression of these genes. HDP1 also enhances the expression of *mdv1*, another gene known to be essential during the early stages of gametocyte maturation[36]. However, the observed 70% reduction in *mdv1* expression cannot explain the complete arrest of gametocyte maturation following loss of HDP1, because low-level expression of *mdv1* is sufficient for the normal development of female gametocytes[36] while all Δ*hdp1* gametocytes are arrested and die during early gametocytogenesis.

While HDP1 is not expressed in *P. falciparum* asexual blood stages and is essential during early gametocyte development, whether it is required in other parasite stages remains to be determined.

Many genes with upstream HDP1 binding have maximal expression in stage V gameotcytes and even in ookinetes (Dataset 4). Since the knockdown system used in this study regulates expression at the transcript level, we can infer that *hdp1* messenger RNA is not required during the later stages of gametocyte maturation, but HDP1 protein expressed during the earlier stages of gametocytogenesis may still be required later. Interestingly, we found that HDP1 enhances the expression of at least two other DNA-binding proteins, indicating that it may also be involved in a cascade of transcriptional regulation that underlies gene expression changes during late gametocytogenesis and beyond. Homeodomain-like proteins have been implicated in mating processes of other haploid protozoa, such as *Dictyostelium*[37]. Because previous transcriptomic studies indicate that HDP1 is also expressed in ookinetes[38], HDP1 may well play a role in subsequent stages of the parasite life cycle. Moreover, while expression of HDP1 is gametocyte specific in *P. falciparum* blood stages, disruption of its *P. berghei* ortholog (PBANKA_1329600) significantly impaired asexual blood stage replication[39]. This suggests that the gametocyte-specific function of HDP1 may have evolved in the Laveranian clade of malaria parasites, which uniquely produce falciform gametocytes.

Additional studies will be required to elucidate the structure–function relationship of HDP1 and identify key interaction partners. Only 2% of its protein sequence comprises the homeodomain-like DNA-binding domain, while the remainder contains no other identifiable domains and conservation is generally weak. 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 functions remains unclear. Similarly, identification of interactions with other nuclear proteins and genomic locations will offer important insights into its function.

Methods

Parasite culture. Unless otherwise noted, *P. falciparum* parasites were grown in 0.5% AlbuMAX II supplemented malaria complete medium using established cell culture techniques[40], at 3% haematocrit and <3% parasitemia. Strains expressing selectable markers were maintained under constant drug selection. *Toxoplasma* tachyzoites were cultured as described in ref. [41].

Gametocyte induction and isolation. Gametocytes were induced synchronously as previously described in ref. [9]. Gametocyte maturation was monitored by Giemsa-stained thin blood smears and gametocytemia was counted on day 5 of development. Sexual commitment rate was calculated by dividing day 5 gametocytemia by day 1 parasitemia, counted before the addition of N-acetyl-D-glucosamine. For knockdown experiments in the *hpd1-glmS* line, at the gametoring stage either 5 mM glucosamine or solvent control was added. Gametocytes were purified from culture at the required development stage using magnetic columns (LS columns, Miltenyi Biotec).

Generation of transgenic strains. Transfection of ring-stage parasites was performed as previously described[42]. Genome editing was performed by CRISR–Cas9 technology using the system described in ref. [43]. Flanking homology regions were PCR amplified using Advantage Genomic LA polymerase (Takara) from NF54 genomic DNA (Supplementary Table 1 gives a list of primers used) and cloned into the *AflII* and *SpeI* sites of pL6 (carrying selectable marker hDHFR) by Gibson assembly. *Plasmodium* codon-optimized sequences for Halo and triple Ty1 epitope tags were synthetized as gene-Blocks (Genewiz). Double-stranded single-guide RNA blocks were synthetized as single-stranded DNA and then annealed and cloned into the *XhoI* site of pL6. The absence of undesired mutations in homology regions and sgRNA was confirmed by Sanger sequencing. Genomic DNA from transfectant parasites was isolated using the QIAamp DNA blood Kit (Qiagen, catalogue no. 51106), and diagnostic PCR performed with Taq Phusion DNA polymerase (Invitrogen). The parasite line TGME49_233160-HA was generated as part of an earlier study by tagging of the endogenous locus in *T. gondii* strain RH-ku80ko as described in ref. [41].

Flow cytometric analysis of gametocyte viability. Gametocytes were stained with 16 µM Hoechst 33342 and 50 nM DilC(1)5 for 30 min 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 1,000 gametocytes[44]. Analysis was carried out using FlowJo 10. The gating strategy is shown in Extended Data Fig. 9.

Nuclear extract preparation and chromatin high-salt fractionation. Nuclear isolation and extraction was carried out based on ref. [45], with minor modifications. Briefly, parasites released from red blood cells (RBCs) by saponin treatment (0.01%) were lysed with ice-chilled CLB (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM EDTA pH 8.0, 1 mM EGTA pH 8.0, 0.65% NP-40, 1 mM DTT, 1× Roche Complete protease inhibitors cocktail). Nuclei were pelleted at 3,000*g* for 20 min at 4 °C and the 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, 1× Roche Complete protease inhibitors cocktail EDTA-free) and treated with 5 U of micrococcal nuclease (Thermo Fisher, catalogue no. 88216) for 30 min in a water bath at 37 °C. Soluble and insoluble nuclear fractions were recovered by centrifugation at 3,000*g* for 10 min at 4 °C. The insoluble nuclear fraction was 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 phenylmethanesulfonyl fluoride (PMSF), 1× Roche Complete protease inhibitors cocktail) supplemented with increasing NaCl concentrations (80–600 mM) while rotating at 4 °C for 30 min. All supernatants were recovered by centrifugation at 700*g* for 4 min at 4 °C, and the last remaining pellet was resuspended in 1× PBS supplemented with protease inhibitors cocktail; 5% of each fraction was prepared for immunoblotting to check the quality of fractionation.

Immunoblotting. For SDS–polyacrylamide gel electrophoresis (SDS–PAGE), total protein lysates were prepared using saponin-lysed parasites resuspended with 1× Laemmli loading buffer diluted in 1× PBS supplemented with 1× Roche Complete protease inhibitors cocktail. Protein samples were separated in 4–15% polyacrylamide gels using Precision Plus Protein Standards as an MW ladder (Bio-Rad) and transferred to 0.2 µm Immobilion-P transfer membrane (Millipore, catalogue no. ISEQ00010) using a Bio-Rad transfer system. Membranes were blocked in 5% skim milk/1× Tris-buffered saline (TBS)-Tween20 for 1 h at room temperature (RT). Primary and secondary antibodies were prepared in 3% skim milk/1× TBS-Tween20 and incubated for 1 h at RT. Membranes were washed four times with 1× TBS-Tween20 for 10 min, after incubation of primary and secondary antibodies. The following primary antibodies were used in this study: anti-Ty1 BB2 mouse (1:2,500; Invitrogen, catalogue no. MA5-23513), anti-PhIL1 rabbit (1:5,000 (ref. [46])), anti-PfHsp70 rabbit (1:5,000; StreesMarq Biosciences, catalogue no. SPC-186D) and anti-Histone 4 rabbit (1:2,000; Diagenode, catalogue no. C15410156-50). Horseradish peroxidase-conjugated anti-mouse and anti-rabbit antibodies were used (1:5,000, Millipore). Immunoblots were incubated with the chemiluminescent substrate SuperSignal West Pico PLUS (Thermo Fisher, catalogue no. 34578) following the manufacturer's directions. Chemiluminescent images were obtained using an Azure c300 digital imaging system (Azure Biosystems).

Live-cell and immunofluorescence microscopy. For live-cell microscopy of *hdp1-gfp* and NF54 blood stages, infected RBCs were stained with 16 µM Hoechst 33342 in incomplete medium for 15 min at 37 °C, and imaged with identical exposure settings at ×1,000 magnification using a Leica DMI6000 microscope with differential interference contrast bright-field optics, DAPI and GFP filter cubes. For immunofluorescence microscopy of *hdp1-Ty1* and *hdp1-glmS* gametocytes, cells were immobilized on glass slides with Concanavalin A (5 mg ml^{-1}; Sigma) as described in ref. [47], then fixed with a solution of 4% paraformaldehyde/0.0075% glutaraldehyde for 20 min at 37 °C. Parasites were permeabilized with 0.1% Triton X-100 for 15 min at RT followed by blocking with 3% bovine serum albumin (BSA). Primary antibodies (anti-Ty1 BB2 mouse 1:1,000, anti-PhiL1 rabbit 1:400) were allowed to bind for 1 h in 3% BSA/PBS followed by three washes with blocking buffer for 5 min. Secondary antibodies were diluted at 1:500 (anti-mouse-Alexa 546 and anti-rabbit-Alexa 488, Invitrogen) in fresh blocking buffer containing 16 µM Hoechst 33342 and incubated for 1 h. Z-stacks of stained specimens were collected at ×1,000 magnification using a Leica DMI6000 microscope with differential interference contrast bright-field optics, DAPI and red fluorescent protein 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 carried out as previously described[41].

Protein expression and purification. Expression of the recombinant HDP1 DBD motif was done using the glutathione S-transferase (GST) gene fusion system (GE Healthcare). The plasmid pGEX-4T-1 was used as backbone for cloning of the codon-optimized sequence comprising the final 87 amino acids of HDP1. Plasmid pGEX-GST-HDP1 DBD was transformed into competent *E. coli* strain BL21 (DE3) (NEB), and protein expression performed following the manufacturer's directions with some modifications. Briefly, an overnight culture was inoculated with one bacterial colony in 2× YT medium supplemented with the corresponding antibiotic. The following day, cultures were diluted 1:100 with fresh medium and kept at 30 °C with vigorous agitation. Bacterial growth was monitored until cultures reached the exponential phase. At this point, isopropyl-ß-d-thiogalactopyranoside (IPTG, 1 mM final concentration) was added and the culture was kept for a further 2 h at 30 °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, 1× protease inhibitors cocktail) supplemented with Lysozyme (1 mg ml^{-1}, Sigma). To remove bacterial DNA from our putative DNA-binding protein, lysate was treated with polyethyleneimine (0.1% v/v)[48]. Lysate was sonicated and cleared by centrifugation at 14,000*g* for 30 min at 4 °C. Protein extract was recovered, and GST-HDP1 DBD purified using the Pierce GST Spin purification kit (catalogue no. 16106) following the manufacturer's directions. Proteins of interest were dialysed using Slide-A-Lyzer Dialysis Cassette 10,000 MWCO (Thermo Scientific, catalogue no. 66810) and concentrated using Amicon Ultra Centrifugal filters 10,000 MWCO (Millipore, catalogue no. UFC901024). Purity was verified by Coomassie staining after SDS–PAGE, and concentration was measured by Bradford assay.

PBM. GST-HDP1–DBD binding was analyzed twice on PBMs as previously described[49,50]. In this study two different universal PBM arrays (Agilent AMADIDs 016060 v.9 and AMADID 015681 v.11) were used, covering all contiguous 8-mers, as well as gapped 8-mers spanning up to ten positions. Binding of purified GST-HDP1–DBD fusion proteins was visualized on PBM 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. An enrichment cutoff score 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 <0.45 were obtained.

Isothermal titration calorimetry. Sequences encoding HDP1 aa2991–3078 were cloned into the MCS1 of the pRSFDuet-1 vector (Novagen) engineered with an N-terminal His-SUMO tag. Proteins were expressed in *E. coli* strain BL21 CodonPlus (DE3)-RIL (Stratagene). Bacteria were grown in Luria-Bertani medium at 37 °C to optical density (OD_{600}) = 0.8 and induced with 0.4 mM IPTG at 18 °C overnight. Cells were collected by centrifugation at 5,000*g* and lysed via sonication in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 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 a 5-ml HisTrap FF column (GE Healthcare) and eluted using the lysis buffer supplemented with 500 mM imidazole. The elution was dialysed at 4 °C overnight against the buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol) with ULP1 protease added (laboratory stock). The sample was then reloaded on the HisTrap FF column to remove the His-SUMO tag. The flowthrough was loaded on a 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 $MgCl_2$, 1 mM DTT). High-purity eluting fractions were detected by SDS–PAGE and concentrated to around 20 mg ml^{-1}. Protein was flash-frozen in liquid nitrogen and stored at −80 °C.

All binding experiments were performed on a Microcal ITC 200 calorimeter. Purified HDP1 DBD proteins were dialysed 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 and dissolved in the same buffer. Assays were performed with 1 mM DNA duplexes containing the tandem motif (TAGTGCACCTATGGTGCACTT) with 0.1 mM HDP1 DBD protein. The exothermic heat of each reaction was measured by sequential injection of 2-µl DNA duplexes into protein solution, spaced at intervals of 180 s. Titration was performed according to standard protocol at 20 °C, and data were fitted using the program Origin 7.0.

Gel-shift assay. 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[33]. Biotinylated double-stranded DNA probes were synthesized (Thermo Fisher) with the indicated sequence.

RNA extraction, complementary DNA synthesis and qRT–PCR. Total RNA from saponin-lysed parasites was extracted using Trizol (Invitrogen) and the Direct-Zol RNA MiniPrep Plus kit (Zymo Research). cDNA was prepared from 100–500 ng of total RNA (pretreated with 2 U of DNase I, amplification grade) using the 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 (Supplementary Table 1). Quantities were normalized to seryl-tRNA synthetase (PF3D7_0717700).

RNA sequencing. Following gametocyte induction, highly synchronous cultures containing committed schizonts were added to fresh RBCs and allowed to reinvade for 12 h, before 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 from NF54 and two independent Δ*hdp*1 clones 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). RNA concentrations were measured using the NanoDrop system (Thermo Fisher Scientific). Preparation of the RNA sample library and RNA sequencing (RNA-seq) were performed by the Genomics Core Laboratory at Weill Cornell Medicine. Ribosomal RNA was removed from total RNA using the Ribo Zero Gold for human/mouse/rat kit (Illumina). Using the TruSeq RNA Sample Library Preparation v.2 kit (Illumina), RNA was fragmented into small pieces using divalent cations under elevated temperature. 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 ligation of adaptors. The products were then purified and enriched with PCR to create the final cDNA library. Libraries were pooled and sequenced on an Illumina HiSeq4000 sequencer with 50 single-end cycles. Read files were checked for quality using FASTQC v.0.11.5 (https://github.com/s-andrews/FastQC). Reads were trimmed to remove low-quality positions and adaptor sequences using cutadapt (v.1.16)[51]. Reads were mapped against the *P. falciparum* 3D7 reference genome v.40 (ref. [52]) using STAR aligner (v.2.61)[53], and nuclear-encoded genes were analyzed for differential gene expression with cufflinks (v.2.2.1)[54]. Genes with false discovery rate (FDR) ≤ 0.05 and mean fragments per kilobase exon per million mapped reads (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 the purposes of fold-change calculation. GSEA was carried out with the FGSEA v.1.16.0 Bioconductor v.3.12 package[55] with FDR cutoff ≤0.05.

ChIP–seq. *hdp1-gfp* and *hdp1-Ty1* gametocytes were harvested on day 5 of development, isolated from the middle interface of a 30/35/52.5% percoll gradient, washed in 1× PBS and then spun down for 5 min at 2,500*g*. Pelleted parasites were resuspended in 500 µl of lysis buffer (25 mM Tris-HCl pH 8.0, 10 mM NaCl, 2 mM AESBF, 1% NP-40, 1× protease inhibitors cocktail) and incubated for 10 min at RT. Parasite lysates were homogenized by passage through a 26G needle 15 times. Samples were crosslinked by the addition of formaldehyde to a final percentage of 1.25, followed by further homogenization by passage through a 26G needle ten times and incubation for 25 min at RT with continuous mixing. Crosslinking was quenched by the addition of glycine to a final concentration of 150 mM, followed by incubation for 15 min at RT and then 15 min at 4 °C with continuous mixing. Samples were spun for 5 min at 2,500*g* at 4 °C, and crosslinked parasite pellets were washed once with 500 µl of ice-cold wash buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 2 mM AESBF, 1× protease inhibitors cocktail). Samples were stored at −80 °C until further use. Crosslinked parasite pellets were thawed on ice and resuspended in 1 ml of nuclear extraction buffer (10 mM HEPES, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM AEBSF, 1× protease inhibitor cocktail). After 30 min of incubation on ice, Igepal-CA-630 was added to a final concentration of 0.25% and homogenized by passage through a 26G needle seven times. The nuclear pellet was extracted by centrifugation at 2,500*g*, then resuspended in 130 µl of shearing buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl pH 7.5, 1× protease inhibitor cocktail) and transferred to a 130-µl Covaris sonication microtube. The sample was then sonicated using a Covaris S220 Ultrasonicator for 8 min (duty factor, 5%; intensity peak power, 140; cycles per burst, 200; bath temperature, 6 °C). Samples were transferred to ChIP dilution buffer (30 mM Tris-HCl pH 8.0, 3 mM EDTA, 0.1% SDS, 300 mM NaCl, 1.8% Triton X-100, 1× protease inhibitor cocktail, 1× phosphatase inhibitor tablet) and centrifuged for 10 min at 16,000*g* at 4 °C, retaining the supernatant. For each sample, 13 µl of protein A agarose/salmon sperm DNA beads was washed three times with 500 µl of ChIP dilution buffer (without inhibitors) by centrifuging for 1 min at 800*g* at room temperature, then the buffer was removed. For preclearing, diluted chromatin samples were added to the beads and incubated for 1 h at 4 °C with rotation, then pelleted by centrifugation for 1 min at 800*g*. Supernatant was extracted carefully to avoid removal of any beads; 10% of the sample was removed

for use as input, and 2 µg of anti-GFP antibody (Abcam ab290, anti-rabbit) was added to the remaining sample with incubation overnight at 4 °C with rotation. For each sample, 25 µl of protein A agarose/salmon sperm DNA beads was washed with ChIP dilution buffer (no inhibitors), blocked with 1 mg ml^{-1} BSA for 1 h at 4 °C then washed three more times with buffer. Next, 25 µl of washed and blocked beads was added to the sample with incubation for 1 h at 4 °C with continuous mixing, to collect the antibody/protein complex. Beads were pelleted by centrifugation for 1 min at 800*g* at 4 °C. The bead/antibody/protein complex was then washed for 15 min with rotation using 1 ml of each of the following buffers twice: low-salt immune complex wash buffer (1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), high-salt immune complex wash buffer (1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), LiCl immune complex wash buffer (0.25 M LiCl, 1% Igepal, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) and TE wash buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The complex was then eluted from the beads by the addition of 250 µl of freshly prepared elution buffer (1% SDS, 0.1 M sodium bicarbonate), twice, with rotation for 15 min. We added 5 M NaCl to the elution, and crosslinking was reversed by heating at 45 °C overnight followed by the addition of 15 µl of 20 mg ml^{-1} RNase A with incubation for 30 min at 37 °C. Next, 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris-HCl pH 7.5 and 2 µl of 20 mg ml^{-1} proteinase K were added to the elution and incubated for 2 h at 45 °C. DNA was recovered by phenol/chloroform extraction and ethanol precipitation, using one volume of phenol/chloroform/isoamyl alcohol (25:24:1) mixture twice and chloroform once, followed by the addition of 1/10 volume of 3 M sodium acetate pH 5.2, two volumes of 100% ethanol and 1/1,000 volume of 20 mg ml^{-1} glycogen. Precipitation was allowed to occur overnight at −20 °C. Samples were centrifuged at 16,000*g* for 30 min at 4 °C, washed with fresh 80% ethanol then centrifuged again for 15 min with the same settings. The pellet was air-dried and resuspended in 50 µl of nuclease-free water. DNA was purified using Agencourt AMPure XP beads. Libraries were then prepared from this DNA using a KAPA library preparation kit (KK8230) and sequenced on a NovaSeq 6000 machine.

Analysis of ChIP–seq data. Reads in sequenced libraries were checked for quality with FASTQC v.0.11.9 (https://github.com/s-andrews/FastQC) and adaptor and quality trimmed with trimmomatic v.0.39 (ref. [51]). Properly paired reads were then aligned against the *P. falciparum* 3D7 genome v.51 with bwa v.0.7.17 (ref. [56]) and sorted by name using samtools v.1.13 (ref. [57]). Differential genome-wide enrichment between HDP1-GFP and HDP1-Ty1 samples was then calculated using macs2 v.2.2.7.1 (ref. [58]) to generate Poisson-based FDR and fold-enrichment tracks and significant peaks. Peak annotation and visualization were carried out in R with the packages GenomicRanges v.1.44.0 (ref. [59]), Gviz v.1.36.2 (ref. [60]) and ChIPpeakAnno v.3.26.2 (ref. [61]) from Bioconductor[62] using *P. falciparum* 3D7 genome annotation v.51 from PlasmoDB. Gene promoters were defined as 2 kb upstream of the annotated TSS, and 'upstream region' is defined as the promoter + 5′ UTR. Motif enrichment for 100 bp centred on HDP1-GFP ChIP–seq summits was carried out using the function HOMER2 v.4.11 findMotifsGenome[63].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Processed and raw high-throughput sequencing data have been deposited in the NCBI Gene Expression Omnibus under accession nos. GSE189197 (RNA-seq) and GSE189151 (ChIP–seq). The pipelines for RNA-seq and ChIP–seq data analysis are 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 Information. Additional data related to this paper may be requested from the authors. *P. falciparum* lines generated in this study are available at cost on request. Source data are provided with this paper.

Received: 17 November 2020; Accepted: 8 December 2021;
Published online: 27 January 2022

References

1. Park, B. O., Ahrends, R. & Teruel, M. N. Consecutive positive feedback loops create a bistable switch that controls preadipocyte-to-adipocyte conversion. *Cell Rep.* **2**, 976–990 (2012).
2. Bhattacharya, S. 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. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Stochastic switching of cell fate in microbes. *Annu. Rev. Microbiol.* **69**, 381–403 (2015).
4. Satory, D., Gordon, A. J., Halliday, J. A. & Herman, C. Epigenetic switches: can infidelity govern fate in microbes? *Curr. Opin. Microbiol.* **14**, 212–217 (2011).
5. Kafsack, B. F. C. et al. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 248–252 (2014).
6. Sinha, A. et al. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* **507**, 253–257 (2014).
7. Lopez-Rubio, J. J., 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* **5**, 179–190 (2009).
8. Brancucci, N. M. B. et al. Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe* **16**, 165–176 (2014).
9. Poran, A. et al. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95–99 (2017).
10. Josling, G. A. et al. Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat. Commun.* **11**, 1503 (2020).
11. Coleman, B. I. et al. A *Plasmodium falciparum* histone deacetylase regulates antigenic variation and gametocyte conversion. *Cell Host Microbe* **16**, 177–186 (2014).
12. Filarsky, M. et al. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* **359**, 1259–1263 (2018).
13. Bunnik, E. M. et al. Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages. *Nat. Commun.* **9**, 1910 (2018).
14. Fraschka, S. A. 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 (2018).
15. Bancells, C. et al. Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium falciparum*. *Nat. Microbiol.* **4**, 144–154 (2019).
16. Young, J. A. et al. The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem Parasitol.* **143**, 67–79 (2005).
17. Painter, H. J., Carrasquilla, M. & Llinás, M. Capturing in vivo RNA transcriptional dynamics from the malaria parasite *Plasmodium falciparum*. *Genome Res.* **27**, 1074–1086 (2017).
18. van Biljon, R. et al. Hierarchical transcriptional control regulates *Plasmodium falciparum* sexual differentiation. *BMC Genomics* **20**, 920 (2019).
19. Kent, R. S. et al. Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite *Plasmodium berghei*. *Nat. Microbiol.* **3**, 1206–1213 (2018).
20. Balaji, S., Babu, M. M., Iyer, L. M. & Aravind, L. 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).
21. Modrzynska, K. 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).
22. Yuda, M., Iwanaga, S., Kaneko, I. & Kato, T. Global transcriptional repression: an initial and essential step for *Plasmodium* sexual development. *Proc. Natl Acad. Sci. USA* **112**, 12824–12829 (2015).
23. Singh, S. et al. The PfAP2-G2 transcription factor is a critical regulator of gametocyte maturation. *Mol. Microbiol.* **115**, 1005–1024 (2021).
24. Zhang, C. et al. Systematic CRISPR-Cas9-mediated modifications of *Plasmodium yoelii* ApiAP2 genes reveal functional insights into parasite development. *mBio* **8**, e01986-17 (2017).
25. Yuda, M., Kaneko, I., Iwanaga, S., Mura, Y. & Kato, T. Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. *Mol. Microbiol.* **113**, 40–51 (2019).
26. Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma* **125**, 497–521 (2016).
27. Llorà-Batlle, O. et al. Conditional expression of PfAP2-G for controlled massive sexual conversion in *Plasmodium falciparum*. *Sci. Adv.* **6**, eaaz5057 (2020).
28. Prommana, P. et al. Inducible knockdown of *Plasmodium* gene expression using the glmS ribozyme. *PLoS ONE* **8**, e73783 (2013).
29. Nixon, C. P. 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. Pradhan, L. et al. Crystal structure of the human NKX2.5 homeodomain in complex with DNA target. *Biochemistry* **51**, 6312–6319 (2012).
31. Parkyn Schneider, M. et al. Disrupting assembly of the inner membrane complex blocks *Plasmodium falciparum* sexual stage development. *PLoS Pathog.* **13**, e1006659 (2017).
32. Flueck, C. et al. A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathog.* **6**, e1000784 (2010).
33. Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O. & Llinás, M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6**, e1001165 (2010).
34. Santos, J. M. et al. Red blood cell invasion by the malaria parasite is coordinated by the PfAP2-I transcription factor. *Cell Host Microbe* **21**, 731–741 (2017).
35. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471 (2020).
36. Furuya, T. et al. Disruption of a *Plasmodium falciparum* gene linked to male sexual development causes early arrest in gametocytogenesis. *Proc. Natl Acad. Sci. USA* **102**, 16813–16818 (2005).

37. Hedgethorne, K. et al. Homeodomain-like DNA binding proteins control the haploid-to-diploid transition in *Dictyostelium*. *Sci. Adv.* **3**, e1602937 (2017).
38. López-Barragán, M. J. et al. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* **12**, 587 (2011).
39. Bushell, E. et al. Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell* **170**, 260–272 (2017).
40. Moll, K., Ljungström, I., Perlmann, H. & Scherf, A. *Methods in Malaria Research* (MR4/ATCC, 2008).
41. Morlon-Guyot, J. et al. A proteomic analysis unravels novel CORVET and HOPS proteins involved in *Toxoplasma gondii* secretory organelles biogenesis. *Cell Microbiol.* **20**, e12870 (2018).
42. Rug, M. & Maier, A. G. in *Malaria* (ed. Menard R.) 75–98 (Humana Press, 2012).
43. Ghorbal, M. et al. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819–821 (2014).
44. Tanaka, T. Q. & Williamson, K. C. A malaria gametocytocidal assay using oxidoreduction indicator, alamarBlue. *Mol. Biochem. Parasitol.* **177**, 160–163 (2011).
45. Flueck, C. et al. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* **5**, e1000569 (2009).
46. Saini, E. et al. Photosensitized INA-labelled protein 1 (PhIL1) is novel component of the inner membrane complex and is required for *Plasmodium* parasite development. *Sci. Rep.* **7**, 15577 (2017).
47. Mehnert, A.-K., Simon, C. S. & Guizetti, J. Immunofluorescence staining protocol for STED nanoscopy of *Plasmodium*-infected red blood cells. *Mol. Biochem. Parasitol.* **229**, 47–52 (2019).
48. RR, B. Use of polyethyleneimine in purification of DNA-binding proteins. *Meth. Enzymol.* **208**, 3–10 (1991).
49. Berger, M. F. et al. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
50. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* **4**, 393–411 (2009).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
52. Warrenfeltz, S. et al. EuPathDB: the Eukaryotic Pathogen Genomics Database Resource. *Methods Mol. Biol.* **1757**, 69–113 (2018).
53. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
54. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
55. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/060012v3 (2021).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
57. Li, H. et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
59. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
60. Hahne, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. *Methods Mol. Biol.* **1418**, 335–351 (2016).
61. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
62. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
63. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

Acknowledgements

We thank P. Malhotra for the generous gift of anti-PhIL1 antibodies, V. Carruthers for the generous gift of anti-Ty1 antibodies and the Weill Cornell Medicine genomics core for technical support; and K. Deitsch and J. King for valuable feedback on the manuscript. This work was supported by startup funds from Weill Cornell Medicine: to B.F.C.K., nos. 1R01 AI141965 and 1R01 AI138499; to M.L., no. 1R01 AI125565; and to K.G.L.R., nos. 1R01 AI136511 and R21 AI142506-01; by the University of California, Riverside (no. NIFA-Hatch-225935, to K.G.L.R.); and by support from the Mathers Foundation (to D.J.P.).

Author contributions

Conceptualization was the responsibility of B.F.C.K. Methodology was performed by B.F.C.K., R.A.C.M., W.X. and K.G.L.R. R.A.C.M., X.T., W.X., L.M.O. and W.D. carried out investigations. Software, formal analysis and data curation were provided by B.F.C.K. R.A.C.M. wrote the original draft. B.F.C.K. and R.A.C.M. wrote, reviewed and edited the article. Visualization was conducted by R.A.C.M. and B.F.C.K. B.F.C.K., K.G.L.R., M.L. and D.J.P. supervised the project. Project administration and funding acquisition were carried out by B.F.C.K.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-021-01045-0.

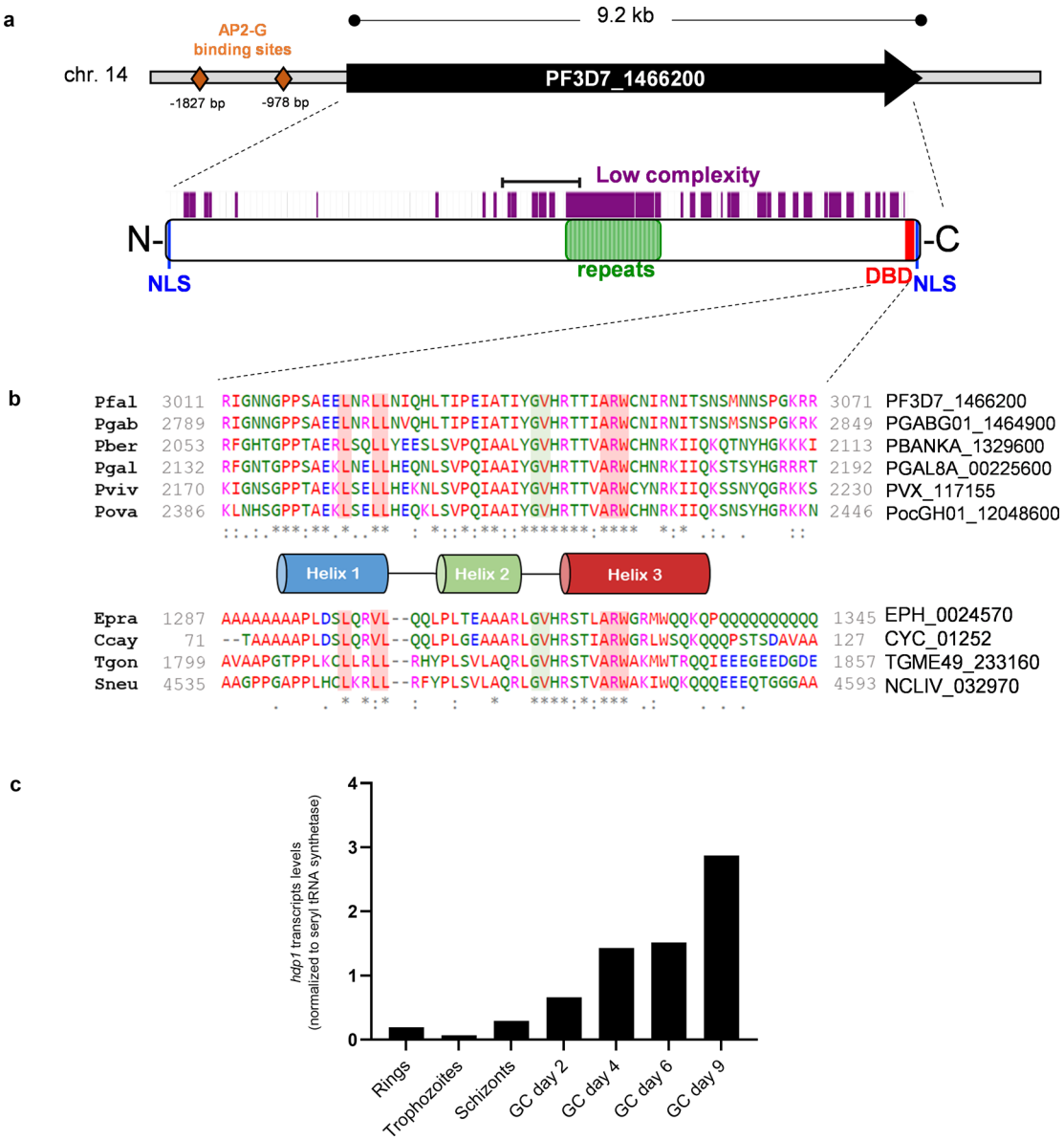
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-01045-0.

Correspondence and requests for materials should be addressed to Björn F. C. Kafsack.

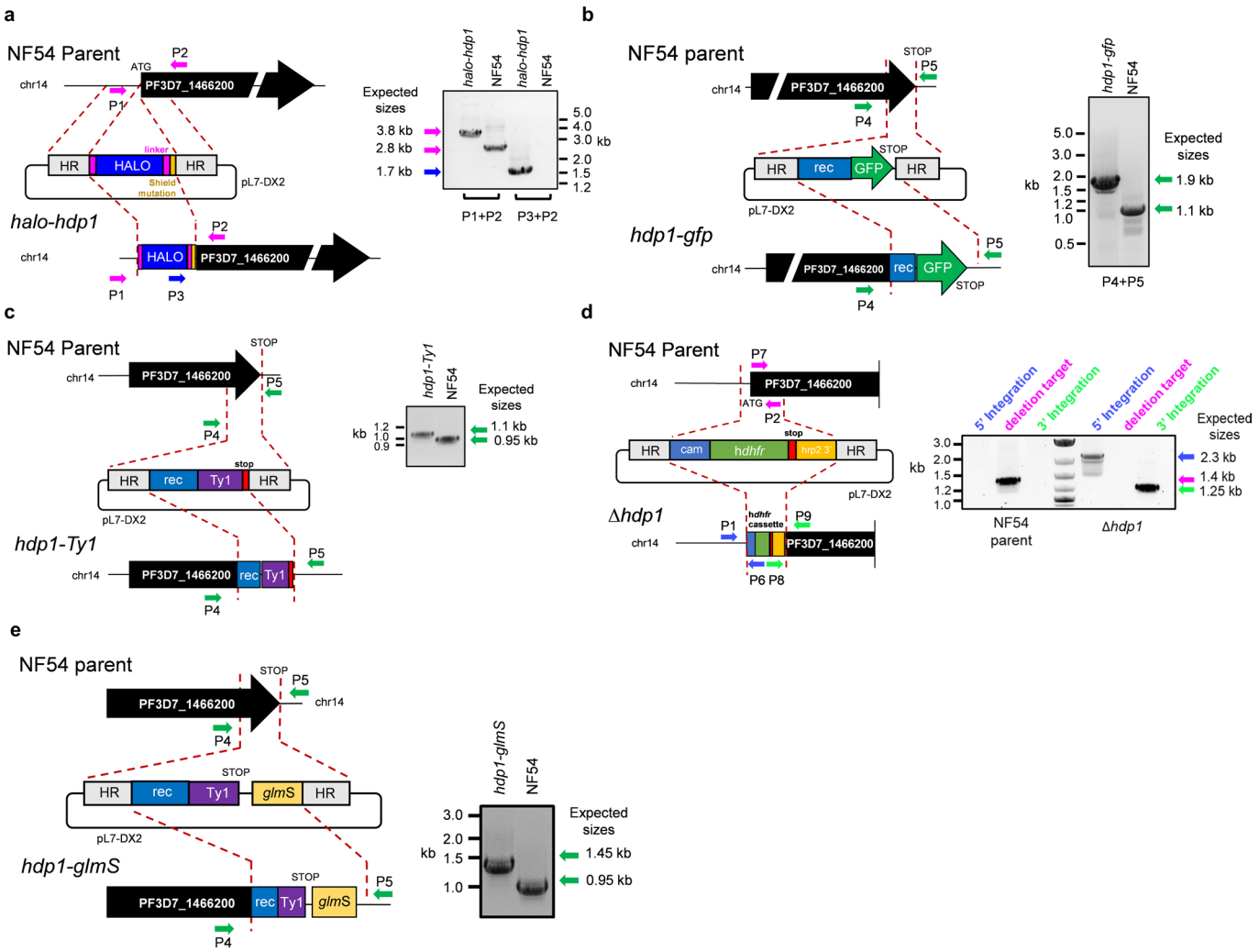
Nature Microbiology thanks Leann Tilley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

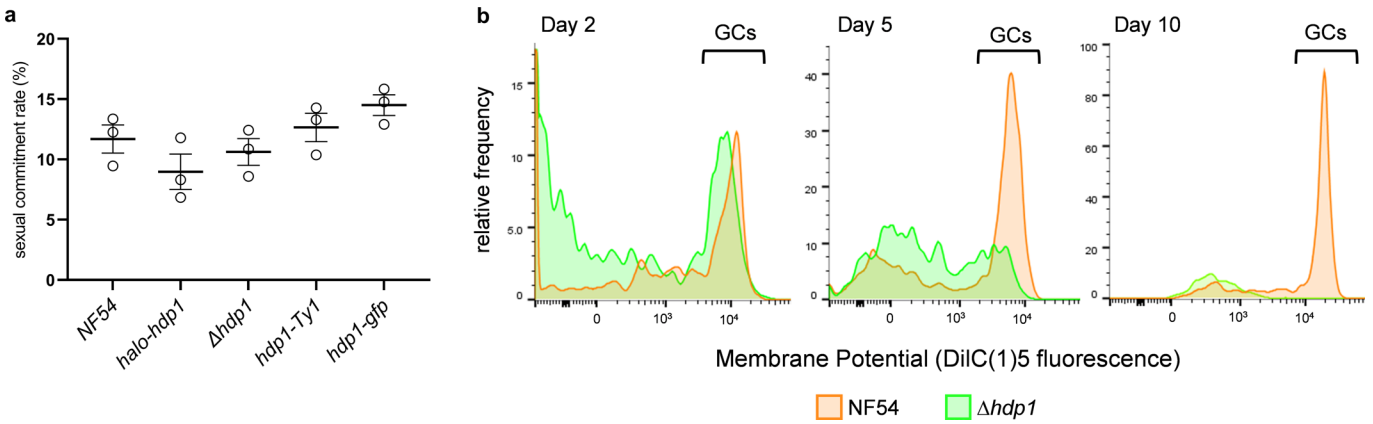
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



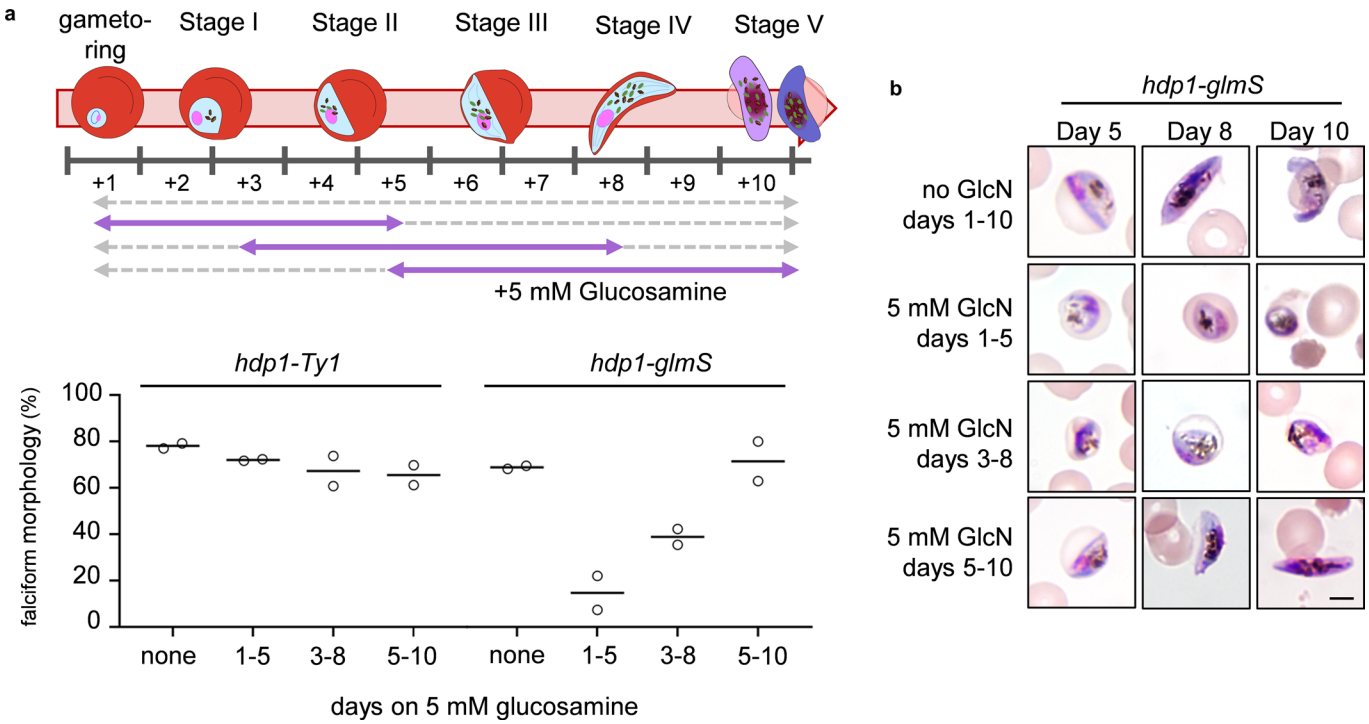
Extended Data Fig. 1 | The predicted DNA-binding protein HDP1 is expressed in gametocytes. (a) The single exon locus Pf3D7_1466200 encodes a large 3078aa protein with a predicted C-terminal Helix-Turn-Helix DNA-binding domain (DBD) and two nuclear localization signals (NLS). Multiple AP2-G binding sites are found in its 2 kb 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 (data are presented as mean values of n = 2 biologically independent samples). Transcript levels were normalized to Seryl-tRNA synthetase.



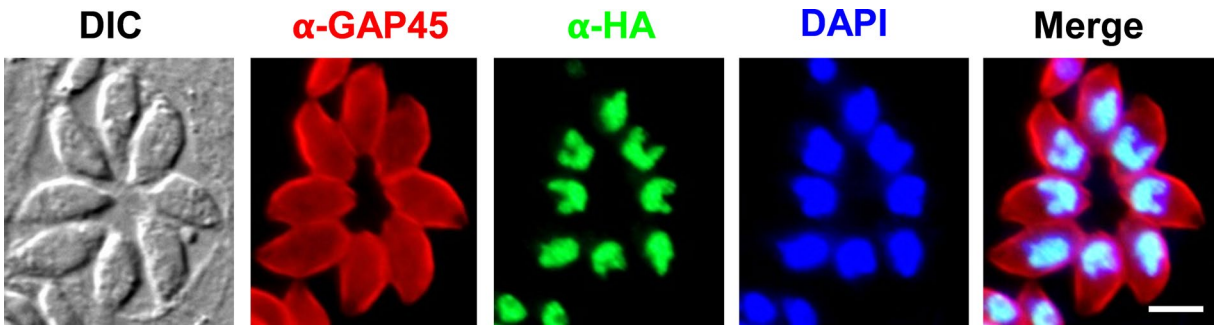
Extended Data Fig. 2 | Validation of engineered parasite lines. (a) Generation of *halo-hdp1* parasites by CAS9 genome editing. Insertion of the N-terminal HALO tag at the 5′ end of *hdp1* coding sequence was confirmed by PCR and checked for mutations by Sanger sequencing of the 3.8 kb PCR product (not 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 1.9 kb 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 1.1 kb 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).



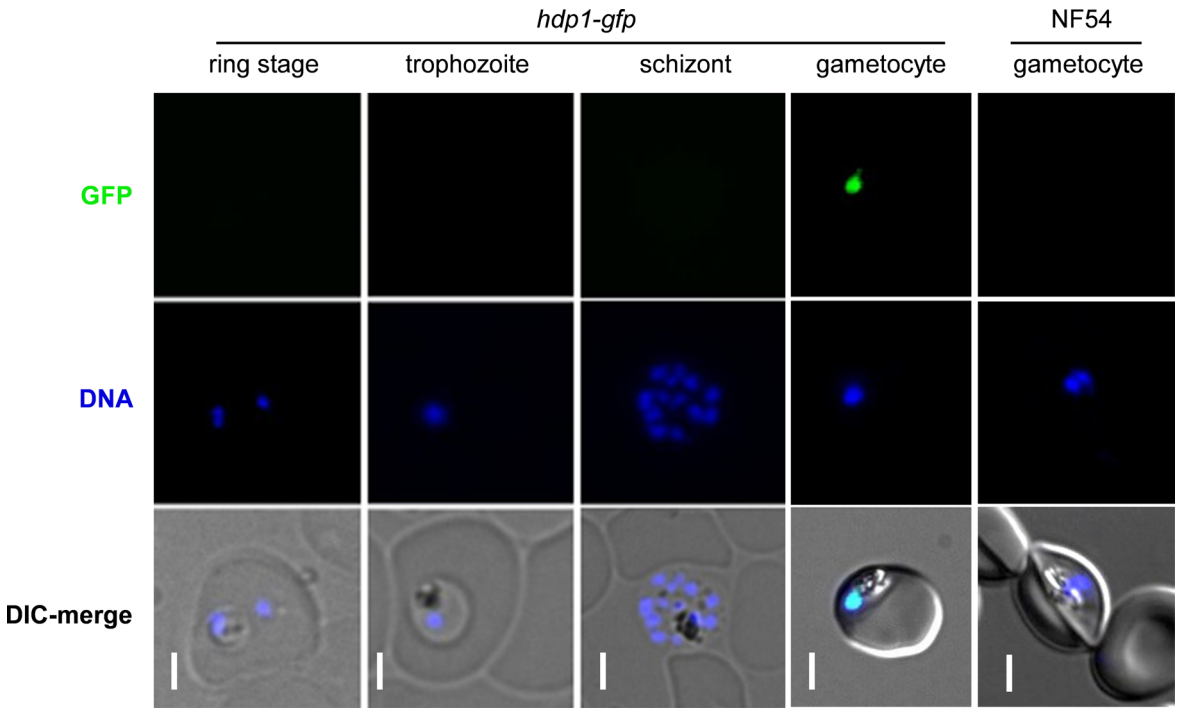
Extended Data Fig. 3 | Loss of HDP1 does not alter the sexual commitment frequency or Stage I gametocyte viability. (a) The sexual commitment frequency (day 5 gametocytes per day 1 ring stages) is not significantly affected in *halo-hdp1* and Δ*hdp1* parasites. Data are presented as mean values ± s.e.m. of n = 3 biologically independent samples. **(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. Flow cytometry plots are representative of n = 2 biologically independet samples.



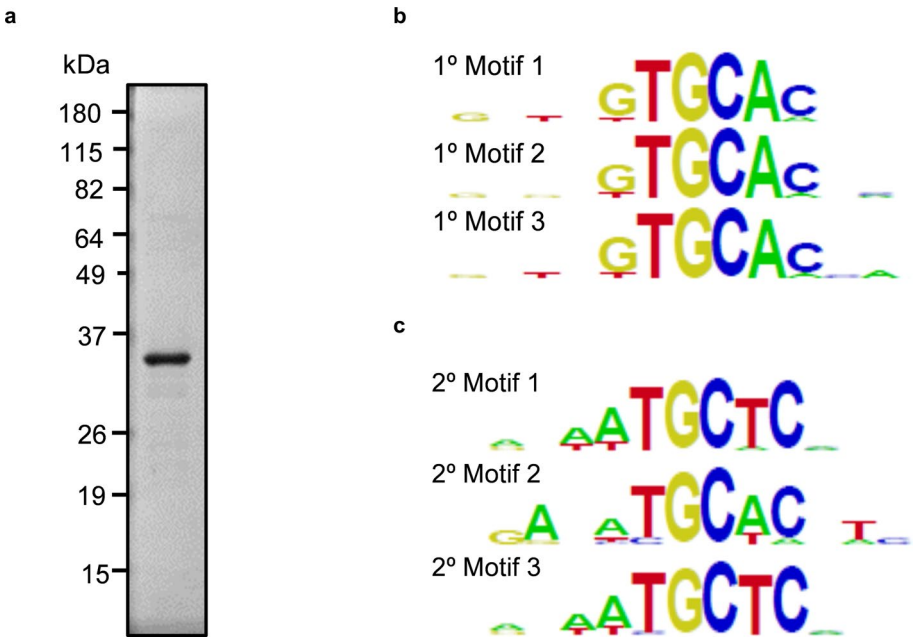
Extended Data Fig. 4 | Inducible knockdown of HDP1 reduces gametocyte maturation in early but not late gametocytes. (a) Percentage of falciform gametocytes on Day 10 in response to 5 mM glucosamine on days 1-5, 3-8, 5-10, or in the absence of glucosamine for *hdp1-Ty1* or *hdp1-glmS* parasites. Data are presented as mean values ± s.e.m of n = 2 biologically independent samples. **(b)** Representative morphology of *hdp1-glmS* gametocytes in response to 5 mM glucosamine on days 1-5, 3-8, 5-10, or in the absence of glucosamine, among two independent experiments. Scale bar, 3 µm.



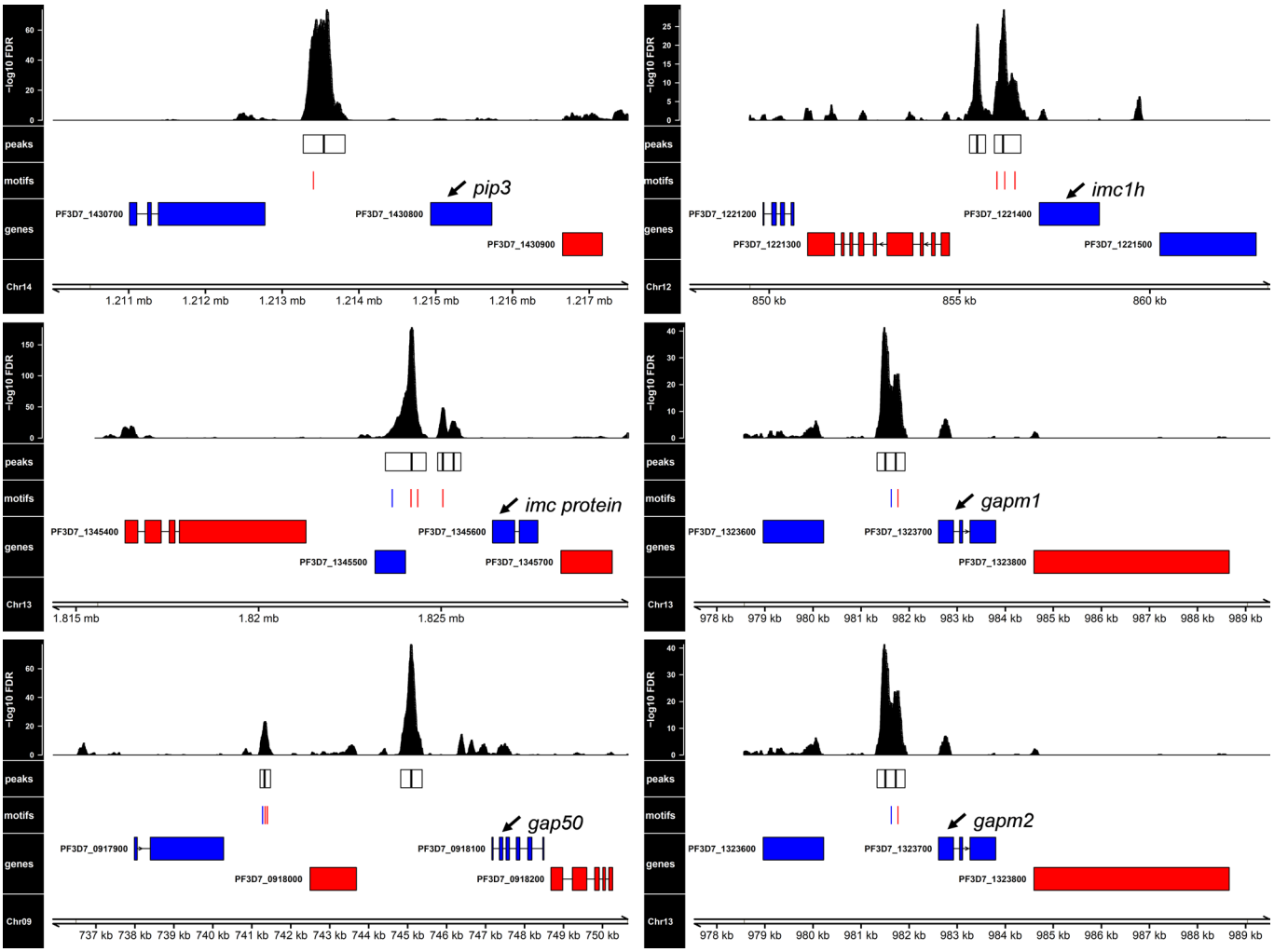
Extended Data Fig. 5 | Immunofluorescence microscopy localizes the HA-tagged ortholog TGME49_233160 to the nucleus of *Toxoplasma gondii* tachyzoites. Staining of *TgME49_233160-HA* intracellular tachyzoites with anti-HA (TgME49_23316, green), anti-GAP45 (inner membrane complex, red), and DAPI (DNA, blue). Scale bar, 5 μm. Image is representative of n = 2 independent experiments.



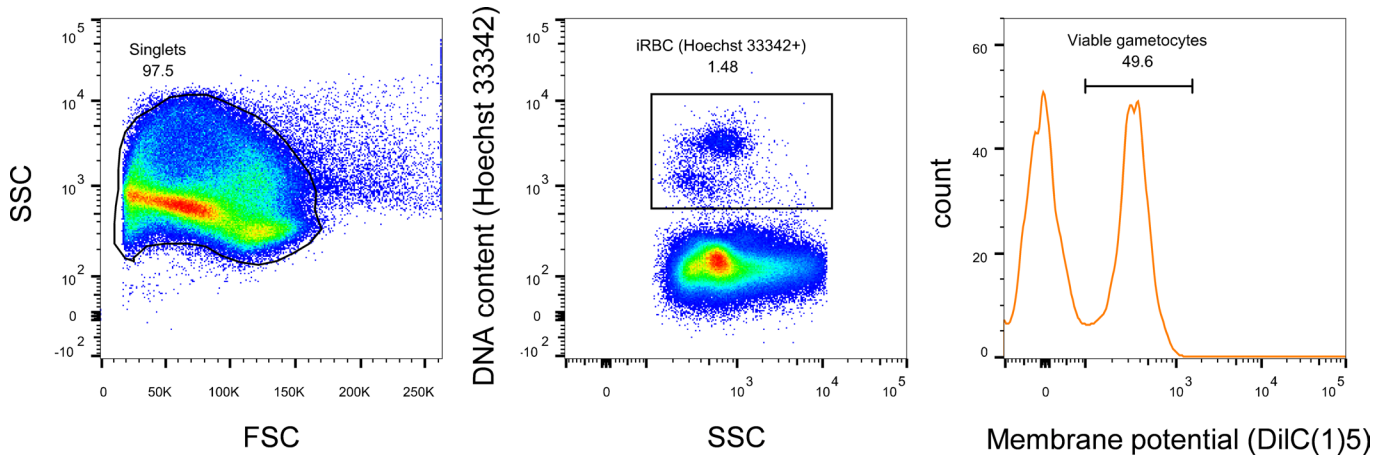
Extended Data Fig. 6 | HDP1-GFP localized to the nucleus of *hdp1-gfp* gametocytes. No signal was observed in *hdp1-gfp* asexual blood stages or gametocytes of the untagged NF54 parent line. Exposure and brightness/contrast settings are uniform across the images shown. Images are representative of n = 2 biologically independent experiments. Scale bar, 3 µm.



Extended Data Fig. 7 | The HDP1 DNA-binding domain tandem GC-rich motif. (a) Coomassie stain of the recombinant 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.



Extended Data Fig. 8 | Genes encoding inner membrane complex genes with significantly reduced expression in HDP1 knockout parasites have upstream HDP1 binding sites. Remaining HDP1 binding sites upstream of genes encoding inner membrane complex proteins. Histogram track shows the significance of enrichment by position. Regions of significant enrichment are shown as boxes with black vertical lines indicating peak summits within each peak. Instances of Motif A, Motif B, or overlapping motifs within peaks are shown in red, blue and purple, respectively. Genes encoded in forward or reverse orientation are shown in blue or red, respectively. Combined estimate of two biologically independent samples.



Extended Data Fig. 9 | 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.

nature portfolio

Corresponding author(s): Björn Kafsack

Last updated by author(s): Nov 9, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Leica LAS v2.7.0.9329, Azure digital imaging1.0,FlowJo CE v7.5.11
Data analysis	ImageJ v1.52.n, DeconvolutionLab2 v08.05.2017, PBMAnalysisSuite v1.0, Origin v7.0, FASTQC v0.11.5, cutadapt v1.16, STAR aligner v2.61 , cufflinks v2.2.1, HOMER2 v4.11, R v4.0.0, Bioconductor v3.12, R packages: fgsea v1.16.0, scales 1.1.1, rtracklayer 1.14.2, plyr 1.14.2, data.table 1.14.2, dplyr 1.0.7, reshape2 1.4.4, ggplot 3.3.5, ggrepel 0.9.1, FloJo v10, trimmomatic v0.39, bwa v0.7.17, samtools v1.13, macs2 v2.2.7.1, GenomicRanges v1.44.0, plyranges 1.14.2, Gviz v1.36.2, ChIPpeakAnno v3.26.2. see https://github.com/KafsackLab/HDP1 for analysis pipelines.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Processed and Raw high throughput sequencing data have been deposited in the NCBI Sequence Read Archive under accession number GSE189197 (RNA-seq) and GSE189151 (ChIP-seq). The pipelines for RNA-seq and ChIP-seq data analysis are 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. P. falciparum lines generated in this study are available at cost upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were set at 3 according to accepted standards in the field, except in cases where the cost of high throughput sequencing limited analysis to 2 independent samples.
Data exclusions	No data was excluded
Replication	All assays were performed with at least 2 independent samples with the most performed in triplicate.
Randomization	Does not apply since since experiments are based on genetic lines
Blinding	samples were not blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Ty1 Tag Monoclonal Antibody (clone BB2) Mouse IgG1, Invitrogen Catalog # MA5-23513 Anti-GFP antibody ChIP Grade (polyclonal) Rabbit IgG, abcam Catalog # ab920 Anti-PhIL1 antibody (polyclonal) Rabbit, non-commercially available (gift from Dr. Pawan Malhotra, ICGEB New Delhi) Anti-HSP70 (P.falciparum) antibody (polyclonal) Rabbit, StreesMarq Biosciences Catalog # SPC186 H4pan (Human) Antibody ChIP Grade (polyclonal) Rabbit, Diagenode Catalog # C15410156-50
Validation	Specificity of antibodies to the Ty1 epitope tag was confirmed by parallel IFA & WB analysis of untagged lines. Specificity of antibodies to GFP was confirmed by WB analysis of untagged lines.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NF54 (MRA-1000) was obtained from BEI resources
Authentication	Newly generated transgenic lines were authenticated based on line specific PCRs and drug sensitivity profiles.
Mycoplasma contamination	Lines were tested for mycoplasma when generated and found negative. Lines have not been tested thereafter.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Processed and Raw high throughput sequencing data have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE189151.
Files in database submission	fastq read file and bedgraph tracks
Genome browser session (e.g. UCSC)	no longer applicable

Methodology

Replicates	n=2 biologically independent samples from hdp1-gfp and hdp1-Ty1 parasite lines were used in this assay, with the latter serving as negative control.
Sequencing depth	Libraries were sequenced in 101bp Paired-End mode with a mean coverage between 46 reads/base and 127 reads/base.
Antibodies	2 µg of anti-GFP antibody (Abcam ab290, anti-rabbit) were used per sample
Peak calling parameters	Properly paired reads were aligned against the P. falciparum 3D7 genome v51 with bwa v0.7.17 and sorted by name using samtools v1.13. Differential genome-wide enrichment between the HDP1-GFP and HDP1-Ty1 samples was then calculated using macs2 v2.2.7.1 to generate poisson-based FDR and fold-enrichment tracks and significant peaks.
Data quality	Reads in sequenced libraries were checked for quality with FASTQC v0.11.9 (https://github.com/s-andrews/FastQC), adapter- and quality-trimmed with trimmomatic v0.39
Software	FASTQC v0.11.9; trimmomatic v0.39 ; bwa v0.7.17; samtools v1.13; macs2 v2.2.7.1; Peak annotation and visualization were carried out in R using the GenomicRanges v1.44.0, Gviz v1.36.2, and ChIPpeakAnno v3.26.2 packages from Bioconductor using P. falciparum 3D7 genome annotation v51 from PlasmoDB; HOMER2 v4.11 findMotifsGenome

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Gametocytes were stained with 16 µM Hoechst33342 and 50 nM DilC(1)5 for 30 minutes at 37℃.
Instrument	Cytek DxP12
Software	Collection FlowJo CE v7.5.11, Analysis: FlowJo v10,
Cell population abundance	Gametocytes occupied between 0.5-2% of erythrocytes analyzed for wild-type levels. Less for some of the mutant lines with gametocytogenesis phenotypes. No cell sorting was performed.
Gating strategy	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.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.