LETTER

doi:10.1038/nature12920

A transcriptional switch underlies commitment to sexual development in malaria parasites

Björn F. C. Kafsack[1]†, Núria Rovira-Graells[2,3], Taane G. Clark[4,5], Cristina Bancells[2], Valerie M. Crowley[1,3]†, Susana G. Campino[6], April E. Williams[7], Laura G. Drought[4], Dominic P. Kwiatkowski[6,8], David A. Baker[4], Alfred Cortés[2,3,9] & Manuel Llinás[1,7]†

The life cycles of many parasites involve transitions between disparate host species, requiring these parasites to go through multiple developmental stages adapted to each of these specialized niches. Transmission of malaria parasites (*Plasmodium* spp.) from humans to the mosquito vector requires differentiation from asexual stages replicating within red blood cells into non-dividing male and female gametocytes. Although gametocytes were first described in 1880, our understanding of the molecular mechanisms involved in commitment to gametocyte formation is extremely limited, and disrupting this critical developmental transition remains a long-standing goal[1]. Here we show that expression levels of the DNA-binding protein PfAP2-G correlate strongly with levels of gametocyte formation. Using independent forward and reverse genetics approaches, we demonstrate that PfAP2-G function is essential for parasite sexual differentiation. By combining genome-wide PfAP2-G cognate motif occurrence with global transcriptional changes resulting from PfAP2-G ablation, we identify early gametocyte genes as probable targets of PfAP2-G and show that their regulation by PfAP2-G is critical for their wild-type level expression. In the asexual blood-stage parasites *pfap2-g* appears to be among a set of epigenetically silenced loci[2,3] prone to spontaneous activation[4]. Stochastic activation presents a simple mechanism for a low baseline of gametocyte production. Overall, these findings identify PfAP2-G as a master regulator of sexual-stage development in malaria parasites and mark the first discovery of a transcriptional switch controlling a differentiation decision in protozoan parasites.

From its uptake in a mosquito blood meal to initial infection of red blood cells in the subsequent host, the malaria parasite *Plasmodium falciparum* goes through at least seven key developmental changes (asexual red cell stage → gametocyte → gamete → ookinete → oocyst → sporozoite → liver stage → asexual red cell stage). In all but one case, as the parasite reaches its subsequent niche within the host, differentiation into the appropriate developmental stage is a necessity for continuation of the life cycle. The lone exception occurs once the parasite has started replicating in red blood cells. During the 48-h intraerythrocytic developmental cycle following each new red blood cell invasion, a developmental decision is made that determines whether daughter parasites will continue replicating asexually and maintain the infection of the current host or differentiate into non-dividing male or female gametocytes. Although the latter decision is a dead-end for replication within the current host it is essential for infection of mosquitoes and thus transmission to the next host[5,6].

A recent study on transcriptional variation identified differentially expressed genes linked to early gametocyte development in two stocks (3D7-A and 3D7-B) of the common 3D7 *P. falciparum* parasite line[7]. Within this expression cluster of early gametocyte markers, we noted the presence of a potential transcriptional regulator, PfAP2-G (PFL1085w/PF3D7_1222600; http://www.plasmodb.org), which belongs to the apicomplexan AP2 (ApiAP2) family of DNA-binding proteins (Supplementary Fig. 1) and is conserved among most members of the phylum (Supplementary Fig. 2). ApiAP2 proteins represent the main family of transcriptional regulators in malaria parasites[8] and have thus far been found to regulate several of the parasite's developmental transitions, including ookinete formation[9] and oocyst sporozoite maturation[10] within the mosquito, and development in the mammalian liver[11]. Follow-up quantitative PCR with reverse transcription (qRT–PCR) analysis in blood-stage parasites confirmed higher *pfap2-g* transcript abundance in 3D7-B compared to 3D7-A and also revealed significant variation in expression levels between individual 3D7-B subclones (Fig. 1a). Notably, when gametocyte formation was measured in these lines, *pfap2-g* transcript levels were highly predictive ($R^2 > 0.99$) of relative gametocyte production (Fig. 1b).

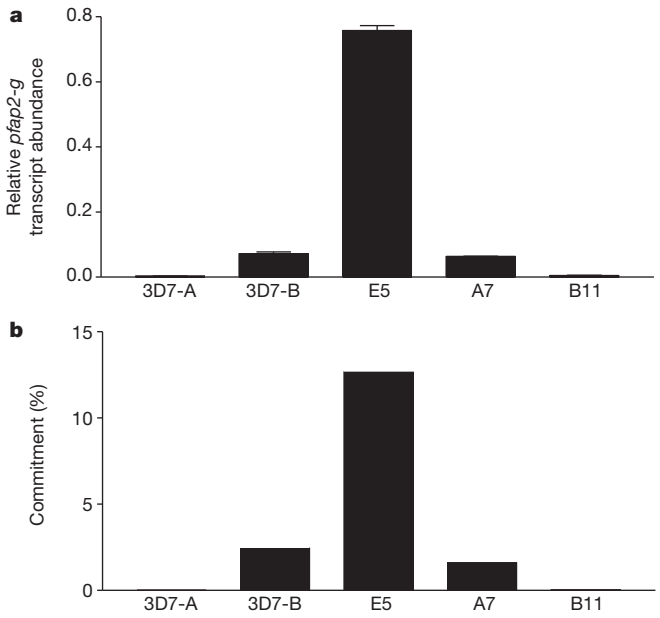


Figure 1 | *pfap2-g* transcript levels mirror gametocyte production. a, *pfap2-g* relative transcript abundance in synchronized (early schizont stage) cultures as measured by qPCR varies significantly between 3D7-A and 3D7-B populations as well as the 3D7-B subclones E5, A7 and B11. Values are normalized against seryl transfer RNA synthetase (PF07_0073) ($n = 3$, standard deviation shown). **b**, Per cent commitment to gametocyte differentiation in these lines mirrors relative *pfap2-g* transcript levels (mean of $n = 2$).

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. [2]Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), Barcelona, 08036 Catalonia, Spain. [3]Institute for Research in Biomedicine (IRB), Barcelona, 08028 Catalonia, Spain. [4]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [5]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [6]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. [7]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [8]Wellcome Trust Sanger Centre for Human Genetics, Oxford OX3 7BN, UK. [9]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, 08010 Catalonia, Spain. †Present addresses: Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA (B.F.C.K.); Department of Molecular Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, State College, Pennsylvania 16802, USA (V.M.C., M.L.).

In a parallel line of inquiry, we screened the well-studied gametocyte non-producer line F12 (refs 5, 6, 12), as well as a second parasite line (GNP-A4) that had also spontaneously lost its ability to produce gametocytes, for mutations in protein-coding regions. Whole genome sequencing of these lines revealed that the only gene containing mutations in both F12 and GNP-A4 was *pfap2-g* (Supplementary Table 1), resulting in the introduction of stop codons upstream of or within the AP2 DNA-binding domain (Fig. 2a and Supplementary Fig. 3). Previous studies identified subtelomeric deletions in the right arm of *P. falciparum* chromosome 9 that are associated with defective gametocyte production[13,14]. The F12 and GNP-A4 clones do not have coding-sequence mutations or deletions within the chromosome 9 region, nor within any of the 16 genes recently implicated in gametocyte development by random transposon mutagenesis[15]. The presence of *pfap2-g* mutations in two independently derived gametocyte non-producer lines provides a second, independent connection between PfAP2-G and gametocyte formation, pointing to this locus as a key determinant of sexual differentiation. Although spontaneous inactivation of *pfap2-g* has occurred repeatedly *in vitro*, no loss-of-function mutations could be found in the genomes of nearly 300 distinct field isolates[16], further underlining its potential importance to transmission.

To directly test the contribution of PfAP2-G function to gametocyte formation, we generated a PfAP2-G null mutant (Δ*pfap2-g*) via double homologous recombination in the high-gametocyte-producing 3D7-B subclone E5 (Fig. 2a, b and Supplementary Fig. 4). As predicted based on our earlier sequencing results, the Δ*pfap2-g* mutant completely lost the ability to produce gametocytes (Fig. 2c). To identify any additional mutations that may have been acquired in the extended process of generating the *pfap2-g* knockout, we sequenced the genomes of both Δ*pfap2-g* and its E5 parent. Apart from the targeted deletion, we found only a limited number of additional mutations within coding regions, none of which are shared with the other non-producer lines that we sequenced or found in genes previously linked to gametocyte development[13–15] (Supplementary Table 1). This combination of forward and reverse genetic evidence strongly implies the essentiality of PfAP2-G for the production of gametocytes in *P. falciparum*. In direct competition cultures Δ*pfap2-g* consistently outgrew its parent E5, consistent with the fact that PfAP2-G action occurs at or before the asexual/sexual decision but not thereafter, as only a failure to initiate gametocytogenesis would provide an *in vitro* growth advantage (Supplementary Fig. 5).

Attempts at generating full-length complementation expression constructs were unsuccessful, probably owing to the considerable length (7.3 kilobases (kb)) of the coding sequence and its very low complexity (21.8% GC and long repeat sequences). As an alternative confirmation for the role of PfAP2-G in gametocyte formation, we made PfAP2-G function ligand-regulatable by appending the FKBP-derived destabilization domain (ddFKBP) to the 3′ end of the endogenous coding sequence (*pfap2-g-ddfkbp*, Supplementary Fig. 6a, b). In the absence of the synthetic ligand Shield-1 (Shld1) the ddFKBP domain is unstable and targets fusion proteins for proteolytic degradation[17,18], thus making PfAP2-G protein levels regulatable by the addition of Shld1 (Supplementary Fig. 6c). Indeed, in the *pfap2-g-ddfkbp* line gametocyte formation was completely dependent on the addition of Shld1, whereas its presence had no effect on gametocyte production by the E5 parent (Fig. 2d, e), demonstrating that PfAP2-G function is essential for gametocyte formation.

On the basis of the localization of haemagglutinin (HA)-tagged PfAP2-G to the parasite nucleus (Fig. 3a and Supplementary Fig. 7) and the fact that several ApiAP2 proteins act as transcriptional regulators, we aimed to identify possible regulatory targets of PfAP2-G. To do this, we compared the global transcriptional pattern over the 48-h intraerythrocytic cycle for the gametocyte-producing parent E5 to those of the mutant non-producers Δ*pfap2-g* and F12. As expected, only a small number of transcripts changed by greater than twofold in both mutants; with four transcripts increasing and 23 transcripts decreasing in abundance (Fig. 3b and Supplementary Table 2). All four upregulated genes are located in subtelomeric regions and have previously been shown to undergo spontaneous transcriptional variation and were therefore not considered further[7]. However, the cluster of downregulated genes is highly enriched for genes expressed during the first stages of gametocyte formation ($P < 0.003$), including some of the earliest known markers of sexual commitment: *pfs16, pfg27/25* and *pfg14.744* (refs 19, 20).

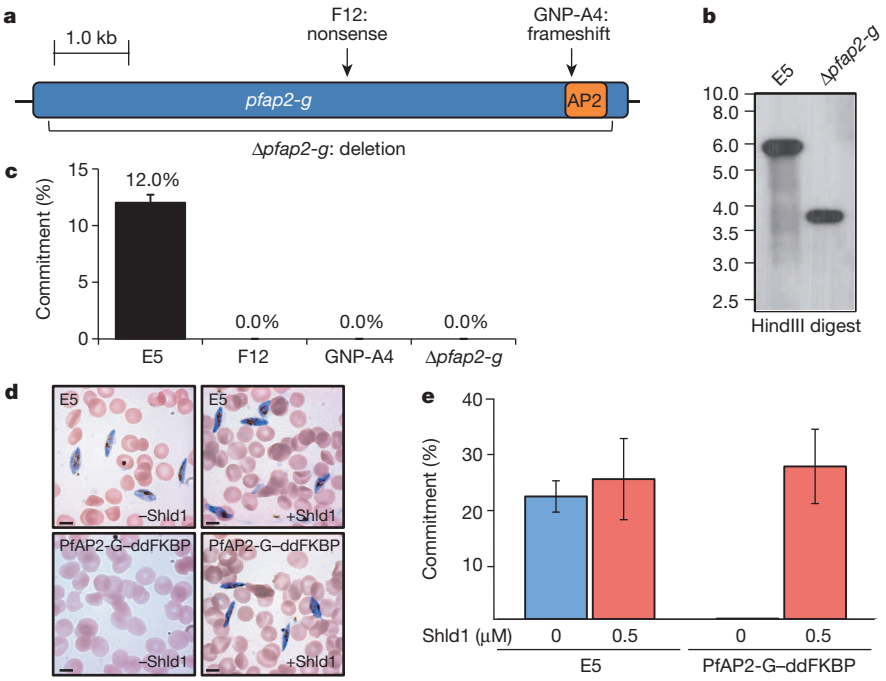


Figure 2 | Disrupting PfAP2-G function results in loss of gametocyte production. a, Positions of *pfap2-g* mutations in the gametocyte non-producer lines F12 and GNP-A4 and the targeted deletion of Δ*pfap2-g*. **b**, Southern blot showing successful disruption of the *pfap2-g* locus by homologous recombination (also see Supplementary Fig. 4). Single replicate. **c**, *pfap2-g* mutants fail to produce gametocytes ($n = 3$, standard error shown). **d**, Ligand-regulatable gametocyte formation in PfAP2-G–ddFKBP (bottom row images) but not in the E5 parent (top row images). Representative of $n = 4$. Scale bars, 5 μm. **e**, Quantification of ligand-regulatable gametocyte formation ($n = 4$, standard error shown).

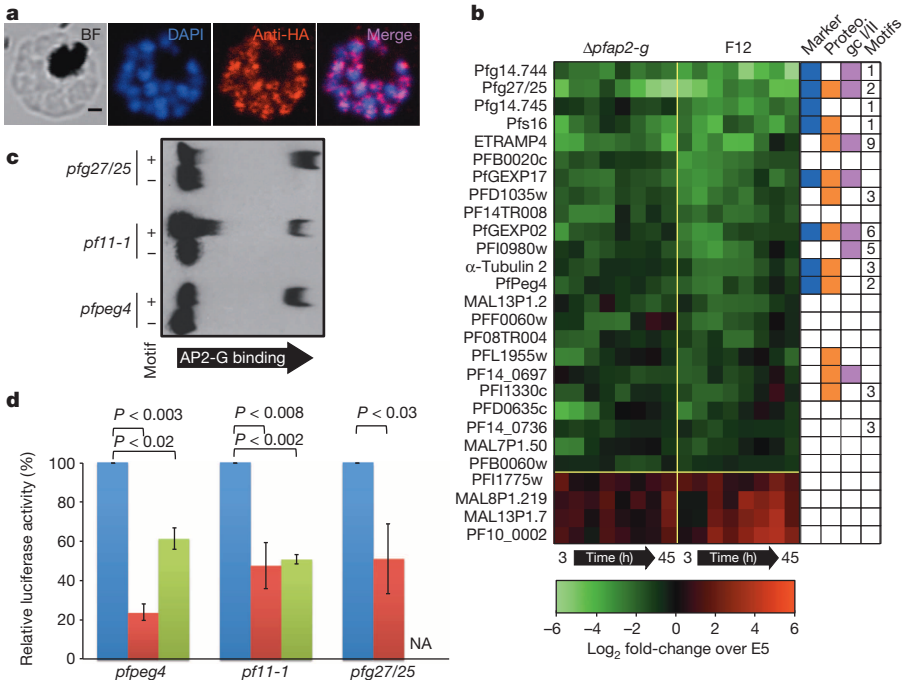


Figure 3 | Identification of PfAP2-G targets. **a**, PfAP2-G–HA×3 localizes to the nuclei of schizonts in asexually growing parasites (see Supplementary Fig. 7 for additional stages). Scale bar, 1 μm. Representative of $n = 8$. BF, bright field; DAPI, 4′,6-diamidino-2-phenylindole. **b**, Relative abundance of transcripts with greater than twofold average difference in both Δ*pfap2-g* and F12 with respect to 3D7-B clone E5 across the intra-erythrocytic developmental cycle at 6-h intervals. Columns on the right indicate whether genes are known gametocyte markers (blue), detected in two or more gametocyte proteomes (orange), enriched in early gametocyte proteome (purple), and the number of PfAP2-G cognate motifs within 2 kb upstream of the start codon. **c**, Binding of the recombinant PfAP2-G AP2 domain to three gametocyte promoters occurs only in the presence of the wild-type cognate motif (+). Representative of $n = 3$. **d**, Relative luciferase activity under the control of wild-type gametocyte promoters in 3D7-B E5 (blue) and Δ*pfap2-g* (red), or in 3D7-B E5 under control of promoters lacking the PfAP2-G motif (green). (18–30 h post-invasion, $n = 3$, standard error is shown, two-sided *t*-test used). NA, not tested.

qRT–PCR measurements of early gametocyte markers confirmed the lower relative abundance levels in Δ*pfap2-g* (Supplementary Fig. 8). Analysis of the upstream regions of most downregulated genes showed that they were also enriched in the DNA motif recognized by PfAP2-G[21] ($P < 0.017$). These results implicate PfAP2-G as a transcriptional switch that controls sexual differentiation by activating the transcription of early gametocyte genes.

Using electrophoretic mobility shift assays we confirmed that the recombinant PfAP2-G DNA-binding domain could interact with three gametocyte promoters in a motif-dependent manner *in vitro* (Fig. 3c). To test whether this interaction occurs within the parasite, we transfected E5 and Δ*pfap2-g* with luciferase reporter constructs under the control of these gametocyte promoters (Fig. 3d). There was a significant reduction in luciferase activity in the Δ*pfap2-g* background compared to its E5 parent for all three constructs. In addition, luciferase levels were also significantly diminished in the parental E5 line when we altered the PfAP2-G recognition sequence in the two promoters tested, indicating that PfAP2-G probably acts as a direct transcriptional activator of the earliest gametocyte genes.

The *pfap2-g* locus shares many features that have been associated with the epigenetic silencing of multigene families in *P. falciparum*[4,22] such as high levels of the H3K9me3 histone modification[3], associated binding of heterochromatin protein 1 (PfHP1)[2,23] and perinuclear localization[3]. On the basis of these data PfAP2-G expression is probably regulated epigenetically by reversible formation of repressive chromatin structures. Interestingly, we find that the pattern of histone modifications at this locus is typical of heterochromatin-silenced genes in both the high gametocyte producer E5 and its low-producing A7 sibling clone (Supplementary Fig. 9a–c). This finding suggests that, in predominantly asexual blood-stage cultures, the *pfap2-g* locus is found in a heterochromatic (silenced) state in the majority of parasites and that the transcriptionally permissive state may only occur in a small number of sexually committed parasites. Indeed, the vast majority of asexually growing PfAP2-G–HA×3 (in which HA×3 denotes a triple HA tag) parasites contained no detectable levels of PfAP2-G by immunofluorescence, whereas a small subpopulation exhibited clear nuclear PfAP2-G staining (Fig. 4a). Every newly formed merozoite within PfAP2-G-expressing schizonts stained positive for PfAP2-G, lending further support to the previous findings that all daughter parasites from a given schizont are commited to the same developmental fate[24]. Furthermore, although the PfAP2-G-positive fraction varied between experiments, it was highly predictive of subsequent gametocyte formation in commitment assays ($R^2 = 0.94$, Fig. 4b).

Stochastic, low-frequency activation would provide a simple mechanism for baseline gametocyte production, which may be modulated in response to environmental stimuli. Furthermore, the presence of insulator-like pairing element sequences—which have been suggested to have an important role in the silencing of *var* genes[25]—flanking the *pfap2-g* locus (Supplementary Fig. 9d) raises the intriguing possibility that the expression of *pfap2-g* may be mutually exclusive with that of the *var* gene family[22]. In addition to chromatin-mediated control, PfAP2-G expression may be autoregulated via binding to the eight instances of the PfAP2-G cognate motifs located 2.1–3.6 kb upstream of the PfAP2-G locus (Supplementary Fig. 10). We have integrated these various regulatory mechanisms into a model of how PfAP2-G expression controls the decision of individual cells to commit to gametocyte formation or to continue along the default pathway of asexual replication (Fig. 4c).

Together with the work of Sinha *et al.*[26] (accompanying manuscript), our results demonstrate that AP2-G is an essential regulator of gametocyte formation in malaria parasites and acts as a developmental switch by activating the transcription of early gametocyte genes. This provides the first insight into the molecular mechanisms controlling the asexual/sexual developmental decision in malaria parasites and unveils new targets in the long-standing aim of interrupting malaria transmission by preventing the formation and/or maturation of the parasite's sexual stages[1]. Last, ligand-regulatable PfAP2-G is not only a powerful new

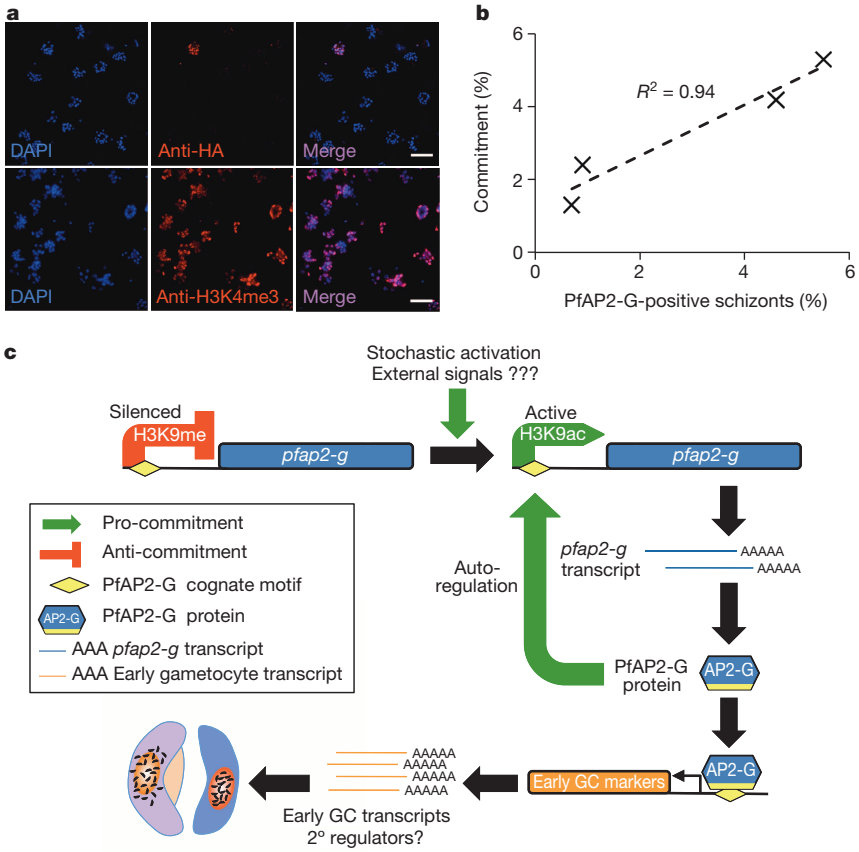


Figure 4 | Activation of PfAP2-G. a, Only a small fraction (1–6%) of asexually growing subclone 9A schizonts (see Supplementary Fig. 7 for details) express detectable levels of PfAP2-G–HA×3 (top row). H3K4me3 staining was performed in parallel to confirm full permeabilization (bottom row). Scale bars, 10 µm. Representative of $n = 4$. **b**, The percentage of PfAP2-G–HA×3-positive cells is highly predictive ($R^2 = 0.94$) of subsequent gametocyte formation levels. **c**, Model of PfAP2-G activation and function.

tool for future studies of sexual-stage development in malaria parasites but also holds great potential for inducible gene expression in general.

METHODS SUMMARY

Parasites and strains. Δ*pfap2-g* knockout parasites were generated by transfection of 3D7-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (*hdhfr*)/negative (*fcu*) selection. *pfap2-g-ddfkbp* parasites were generated by transfection of 3D7-B E5 with pJDD145-pfap2-g (Supplementary Fig. 6). Parasites expressing PfAP2-G-HAx3 were generated by transfection of 3D7-B E5 with pHH1inv-pfap2-g-HAx3 (Supplementary Fig. 7). All parasites were grown in media containing AlbuMAX II and synchronized by standard methods.

Gametocytogenesis. Gametocyte induction was performed according to published methods[27]. For ligand-regulatable gametocytogenesis (Fig. 2d, e), synchronized parasites were set up at 0.5–1.0% late trophozoites in 3% hematocrit on day 0. Cultures were split in two and treated with 0.5 µM Shld1 or solvent control for the remainder of the experiment.

Gel shifts. Electrophoretic mobility shift assays were performed using Light Shift EMSA kits (Thermo Scientific) using 2 µg of protein and 20 fmol of probe.

Microarrays. Starting at 3 h post-invasion, tightly synchronized parasites were collected at eight time points with 6-h intervals. RNA isolation, complementary DNA generation/labelling, array hybridization, and feature extraction was performed as described previously[28]. Cy5-labelled cDNA was hybridized with a common Cy3-labelled reference pool on the *P. falciparum* 8×15K Agilent nuclear expression array (Gene Expression Omnibus (GEO) platform accession GPL17880). Genes were rank ordered by their average relative transcript abundance differences across the eight time points between the wild type (E5) and mutant (F12 or Δ*pfap2-g*).

Luciferase assays. Equal numbers of synchronized, stably transfected parasites were isolated and saponin-lysed (0.05% in PBS) at ~18–30 h post-invasion and assayed using Bright-Glo Luciferase Assay System (Promega).

Next-generation sequencing and analysis. Next-generation sequencing of the 3D7-B subclone E5 and Δ*pfap2-g* was performed using Illumina TruSeq single-end sequencing runs, analysed and visualized as described previously[29]. Genomic DNA for 3D7A, F12 and GNP-A4 was also used for whole genome sequencing at the Sanger Institute using Illumina GA II technology with 76-base paired-end reads. The raw sequence data were processed as described previously[30]. Experimental confirmation of informative genomic variants was performed using capillary sequencing methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 27 November 2013.
Published online 23 February 2014.

1. Wells, T. N. C., Alonso, P. L. & Gutteridge, W. E. New medicines to improve control and contribute to the eradication of malaria. *Nature Rev. Drug Discov.* **8,** 879–891 (2009).
2. 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).
3. 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).
4. Cortés, A., Crowley, V. M., Vaquero, A. & Voss, T. S. A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog.* **8,** e1002943 (2012).
5. Alano, P. *Plasmodium falciparum* gametocytes: still many secrets of a hidden life. *Mol. Microbiol.* **66,** 291–302 (2007).
6. Dixon, M. W. A., Thompson, J., Gardiner, D. L. & Trenholme, K. R. Sex in *Plasmodium*: a sign of commitment. *Trends Parasitol.* **24,** 168–175 (2008).
7. Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.* **22,** 925–938 (2012).
8. Painter, H. J., Campbell, T. L. & Llinás, M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Mol. Biochem. Parasitol.* **176,** 1–7 (2011).
9. Yuda, M. *et al.* Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol. Microbiol.* **71,** 1402–1414 (2009).
10. Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. & Kaneko, I. Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol. Microbiol.* **75,** 854–863 (2010).

11. Iwanaga, S., Kaneko, I., Kato, T. & Yuda, M. Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS ONE* **7,** e47557 (2012).
12. Alano, P. *et al. Plasmodium falciparum*: parasites defective in early stages of gametocytogenesis. *Exp. Parasitol.* **81,** 227–235 (1995).
13. Day, K. P. *et al.* Genes necessary for expression of a virulence determinant and for transmission of *Plasmodium falciparum* are located on a 0.3-megabase region of chromosome 9. *Proc. Natl Acad. Sci. USA* **90,** 8292–8296 (1993).
14. Eksi, S. *et al. Plasmodium falciparum* gametocyte development 1 (*Pfgdv1*) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog.* **8,** e1002964 (2012).
15. Ikadai, H. *et al.* Transposon mutagenesis identifies genes essential for *Plasmodium falciparum* gametocytogenesis. *Proc. Natl Acad. Sci. USA* **110,** E1676–E1684 (2013).
16. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487,** 375–379 (2012).
17. Armstrong, C. M. & Goldberg, D. E. An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*. *Nature Methods* **4,** 1007–1009 (2007).
18. Banaszynski, L. A., Chen, L.-C., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126,** 995–1004 (2006).
19. Pradel, G. Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. *Parasitology* **134,** 1911–1929 (2007).
20. Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Proteomics* **9,** 1437–1448 (2010).
21. 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).
22. Guizetti, J. & Scherf, A. Silence, activate, poise, and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cell. Microbiol.* **15,** 718–726 (2013).
23. Pérez-Toledo, K. *et al. Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of *var* genes. *Nucleic Acids Res.* **37,** 2596–2606 (2009).
24. Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology* **100,** 191–200 (1990).
25. Avraham, I., Schreier, J. & Dzikowski, R. Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **109,** 52 (2012).
26. Sinha, A. *et al.* A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* http://dx.doi.org/10.1038/nature12970 (this issue).
27. Fivelman, Q. L. *et al.* Improved synchronous production of *Plasmodium falciparum* gametocytes *in vitro*. *Mol. Biochem. Parasitol.* **154,** 119–123 (2007).
28. Kafsack, B. F. C., Painter, H. J. & Llinás, M. New Agilent platform DNA microarrays for transcriptome analysis of *Plasmodium falciparum* and *Plasmodium berghei* for the malaria research community. *Malar. J.* **11,** 187 (2012).
29. Straimer, J. *et al.* Site-specific genome editing in *Plasmodium falciparum* using engineered zinc-finger nucleases. *Nature Methods* **9,** 993–998 (2012).
30. Robinson, T. *et al.* Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. *PLoS ONE* **6,** e23204 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank C. Klein, T. Campbell and A. Schieler for technical assistance and are grateful to O. Billker, C. Flueck, J. Kelly, C. Sutherland, A. Vaidya and A. Waters for discussion and reading of the manuscript. We would also like to thank P. Alano for providing *P. falciparum* clone F12, C. Taylor for providing the *P. falciparum* GNP-A4 clone, E. Thompson for isolating *P. falciparum* DNA for whole genome analysis, Z. Gorvett for assistance with confirming single nucleotide polymorphisms in gametocyte non-producing clones, M. Duraisingh for the ddFKBP tagging construct pJDD145, C. Ben Mamoun for the anti-PP2c antibody and D. Goldberg for Shld1. M.L. is funded by National Institutes of Health (NIH) grant R01 AI076276 with support from the Centre for Quantitative Biology (P50GM071508). B.F.C.K. was supported by a Howard Hughes Medical Institute fellowship of the Damon Runyon Cancer Research Foundation. D.A.B. is funded by Wellcome Trust grant ref. 094752 and European Commission FP7 'MALSIG' (ref. 223044). L.G.D. is supported by a Biotechnology and Biological Sciences Research Council CASE PhD studentship with Pfizer as the Industrial partner. A.C. is funded by the Spanish Ministry of Science and Innovation grant SAF2010-20111. V.M.C. was supported by a fellowship from IRB Barcelona. T.G.C. is supported by the Medical Research Council UK (J005398). D.P.K. and S.G.C. are supported through the Wellcome Trust (098051; 090532/Z/09/Z) and the Medical Research Council UK (G0600230). C.B. is supported by the Catalan Government fellowship 2011-BP-B 00060 (AGAUR, Catalonia, Spain).

Author Contributions M.L. managed the overall project with input from B.F.C.K., D.A.B. and A.C. B.F.C.K. generated the Δ*pfap2-g* knockout, PfAP2-G–ddFKBP and luciferase lines and designed, performed and analysed the microarray, gel shift, luciferase and ligand-regulatable gametocytogenesis experiments. V.M.C. performed qRT–PCR validation. A.E.W. prepared Δ*pfap2-g* sequencing libraries and together with B.F.C.K. analysed the sequencing data. D.A.B., T.G.C. and S.G.C. conceived the sequencing of gametocyte non-producer lines F12 and GNP-A4. T.G.C. analysed the gametocyte non-producer sequencing data and L.G.D. confirmed the SNPs by PCR. S.G.C. and D.P.K. carried out and supervised sequencing of gametocyte non-producer lines, respectively. A.C. and N.R.-G. generated E5 and other 3D7-B subclones and respectively supervised and performed the experiments presented in Figs. 1 and 2b, and provided the analysis presented in Supplementary Fig. 1. V.M.C. and N.R.-G. performed and A.C. supervised chromatin immunoprecipitation experiments. C.B. and A.C. generated the PfAP2-G–HA×3 line and carried out immunofluorescence assays and correlations with gametocyte formation. B.F.C.K. wrote the manuscript with major input from M.L., D.A.B. and A.C.

Author Information Microarray data was submitted to the NCBI GEO repository (series accession GSE52030). Next generation sequencing data was submitted to the NCBI Sequence Read Archive (SRA) (study number ERP000190 for samples F12 (ERS011445), 3D7A (ERS011446) and GNP-A4 (ERS011447) and study number SRP035432 for samples E5 (SRS529791) and Δ*pfap2-g* (SRS529811)). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L. (manuel@psu.edu).

METHODS

Parasites and strains. Parasite lines 3D7-A[31], 3D7-B[31] and F12[12] have been described previously. Note that 3D7-A is not the same line as the competent gametocyte producer 3D7A[32], which was used only as a reference genome for next-generation sequence analysis. 3D7-B subclones E5, A7 and B11 were generated by limiting dilution. The gametocyte non-producer line GNP-A4 was generated during an attempt to knockout a phosphodiesterase gene (PfPDEδ, PF14_0672). Integration of the knockout construct by single crossover homologous recombination occurred at the targeted locus but this event was not responsible for the clone's inability to produce gametocytes[33]. A subsequent successful knockout of PfPDEδ produced gametocytes at normal rates and the true phenotype was a significantly lower exflagellation rate than parental parasites owing to a reduced ability of gametes to egress from red blood cells[34]. Δ*pfap2-g* knockout parasites were generated by transfection of 3D7-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (*hdhfr*)/negative (*fcu*) selection using WR99210 and 5-fluoro-cytosine as described previously[35]. Resistant parasites were subcloned and verified by PCR and Southern blot. *pfap2-g-ddfkbp* parasites were generated by transfection of 3D7-B E5 with pJDD145-pfap2-g and selected on WR99210. After subcloning, integration was verified by PCR using a forward primer at position +4,269 and a ddFKBP reverse primer. Displacement of the endogenous downstream sequence was verified using primers at +4,269 and +7,490 with respect to the translation initiation site (Supplementary Fig. 6). Parasites expressing a HA×3-tagged version of PfAP2-G were obtained by transfecting 3D7-B E5 with the plasmid pHH1inv-pfap2-g-HA×3 and cycling twice on/off WR99210 to select for parasites where the plasmid has integrated into the genome. After subcloning by limiting dilution and Southern blot analysis (Supplementary Fig. 7b), a subclone with a single copy of the plasmid integrated at the *pfap2-g* locus (E5-pfap2-g–HA×3 clone 9A) was selected for immunofluorescence assay (IFA) analysis. All parasites were grown in media containing AlbuMAX II and synchronized by standard methods[36].

Knockout and ddFKBP-tagging constructs. Knockout construct: the region from −126 base pairs (bp) to +366 bp and +6,945 to +7,379 bp with respect to the *pfap2-g* initiation codon were cloned into the NcoI/EcoRI and SpeI/SacII sites of pHHT-FCU[35], respectively, to generate pHHT-FCU-pfap2-g. ddFKBP carboxy-terminal tagging construct: *pfap2-g* coding sequence positions +4,740 to +7,296 were cloned into with NotI/XhoI sites of pJDD145[37] (gift from M. Duraisingh).

Luciferase expression constructs. The *hdhfr* selectable marker of pVLbIDh[38] was replaced with blasticidin-S deaminase using the SacI/NotI sites to generate pVL-BSD. The *var7b* promoter was excised with HpaI/KpnI, blunted and re-ligated, destroying these sites. The 1,445 bp, 1,226 bp and 1,159 bp upstream of the *pf11-1*, *pfg27/25* and *pfpeg4* start codons were cloned into the AatII/NcoI sites. The PfAP2-G cognate motifs in the upstream sequences of *pf11-1* (−328 to −323) and *pfpeg4* (−1,138 to 1,131) were converted to adenines using site-directed mutagenesis.

HA×3-tagging construct. The plasmid pHH1inv-pfap2-g-HA×3 was derived from the plasmid E140-0[39]. A triple HA tag (HA×3) was cloned into KpnI-XhoI sites of E140-0, replacing the *eba-140* open reading frame (ORF) and introducing several new restriction sites (plasmid pHH1inv_HA×3). A fragment of the *pfap2-g* ORF from position +6,685 to the stop codon was PCR-amplified and cloned in-frame into KpnI-PstI sites of pHH1inv_HA×3, such that upon integration of the plasmid by single homologous recombination PfAP2-G is expressed as a fusion protein with the HA×3 tag, separated by the sequence YLQ.

Gametocytogenesis. Gametocyte conversion rates for the 3D7-A, 3D7-B and 3D7-B subclones (Fig. 1) were measured by treating synchronized ring-stage parasite cultures at 5% parasitemia with *N*-acetyl-D-glucosamine and counting gametocytaemia 3–4 days later. Gametocyte induction of E5, F12, GNP-A4 and Δ*pfap2-g* (Fig. 2c) was performed according to published methods[27] in media containing 5% AB+ heat-inactivated human serum and 0.25% AlbuMAX II. For ligand-regulatable gametocytogenesis (Fig. 2d, e), synchronized parasites were set up at 0.5–1.0% late trophozoites in 3% hematocrit on day 0. Cultures were split in two and treated with 0.5 μM Shld1 (gift from D. Goldberg) or an equal volume of ethanol solvent control for the remainder of the experiment. Parasitemia was determined on day 3 and 50 mM *N*-acetyl-D-glucosamine was added to all cultures for the remainder of the experiment. Gametocytemia was determined on day 9 and converted into per cent commitment by dividing by day 3 parasitemia. Statistical significance in gametocyte production was determined using unpaired two-sided *t*-tests. Replicates were biological not technical.

Growth competition. E5 and Δ*pfap2-g* parasites were mixed at 1:1 ratio and grown for 7 or more weeks in triplicate. Parasites were diluted 1:20 with uninfected erythrocytes and genomic DNA was isolated whenever parasitemia exceeded 10%. For each time point, gDNA was isolated from the technical replicates, pooled at equal concentration, used for PCR amplification of a 414-nucleotide (nt) (1,718–2,131) region of PFF0275c covering the single nucleotide polymorphism (SNP) (described in Supplementary Table 1) and sequenced using the reverse primer. Difference in the growth rate was determined using the relative sequencing read peak height of A and C across at least 16 replication cycles (32 or more days). Growth rate was fit to the data using:

$$WT_{t=x} = WT_{t=0} / \left(WT_{t=0} + KO_{t=0} \times (1+\Delta g)^{t/2} \right)$$

$$KO_{t=x} = 1 - WT_{t=x}$$

where $WT_{t=x}$ is the relative peak height of cytosine at day x, $KO_{t=x}$ is the relative peak height of adenine at day x, and Δg is the per cent difference in growth rate between Δ*pfap2-g* and E5. 95% confidence intervals were determined using 1.96 × standard error of the mean for the difference in growth rate. Replicates were biological not technical.

Chromatin immunoprecipitation (ChIP). ChIP experiments were performed as described previously[40]. In brief, cultures were synchronized to late trophozoites/schizont stage, saponin-lysed and crosslinked using formaldehyde. Nuclei were released using a Dounce homogenizer (Kimble Chase) and DNA was subsequently fragmented using a Bioruptor (Diagenode). Immunoprecipitations were carried out using commercial antibodies against H3K9ac (Millipore 07-352) and H3K9me3 (Millipore 07-442) and analysed by qPCR using the relative standard curve method. The primers used for ChIP analysis of the *pfap2-g* locus amplify positions (relative to the start codon) −4,954 to −4,875 (5′-1), −1,412 to −1,302 (5′-2), −449 to −351 (5′-3), +3,874 to +3,979 (ORF-1), +5,318 to +5,433 (ORF-2) and +8,492 to +8,632 (3′-1). Primers for the control genes *clag3.1* (primer pair 5, beginning of the ORF), *clag3.2* (primer pair 5, beginning of the ORF), *ama-1* (primer pair 2, beginning of the ORF) and the *var* gene PFL1950w (upstream region, presumably 5′ untranslated region) have been described before[40,41]. Replicates were biological not technical.

Western blots. At the late trophozoite stage synchronized parasites were treated with 0.5 μM Shld1 or ethanol solvent control for 36 h. Proteins were sequentially extracted as described previously[2], separated on 4–12% polyacrylamide gels (Life Technologies) and assayed using anti-HA tag (Roche Diagnostics 11 867 423 001), anti-histone 3 (Abcam ab1791) and anti-PfPP2C (gift from C. Ben Mamoun). Replicates were biological not technical.

qRT–PCR. *pfap2-g* transcript abundance measurements were carried out as previously described[7] using a primer set designed to amplify positions +3,874 to +3,979 and normalized to seryl-tRNA synthetase abundance. Statistical significance was determined using a two-sided *t*-test. Replicates presented are technical and representative of several biological replicates.

IFAs. IFAs were performed on smears of E5-pfap2-g–HA×3-9A cultures synchronized to different stages. Air-dried smears were fixed for 10 min with 1% formaldehyde and permeabilized for 10 min in 0.1% Triton X-100 in PBS. Experiments performed on smears fixed with 90% acetone/10% methanol yielded identical results (not shown). Smears were incubated with rabbit anti-HA (1:100; Life technologies 71-5500) or rabbit anti-H3K4me3 (1:10,000; Millipore 05-745) antibodies. Secondary anti-rabbit antibodies were conjugated with Alexa Fluor 488 (Life technologies A-11034). Nuclei were stained with DAPI. Importantly, no wild-type E5 parasites were positive for staining with anti-HA antibody, and secondary antibody controls also yielded no signal. Preparations were observed under a confocal Leica TCS-SP5 microscope with LAS-AF image acquisition software and were processed using ImageJ software. The proportion of HA-positive schizonts was determined by counting >3,000 schizonts (identified by DAPI staining) for each experiment. The gametocyte conversion rate was measured for each of the same parasite cultures used to quantify the proportion of schizonts positive for anti-HA by IFA. Replicates were biological not technical.

Gel shifts. Electrophoretic mobility shift assays were performed using Light Shift EMSA kits (Thermo Scientific) as previously described using 2 μg of protein and 20 fmol of probe[21]. Biotinylated double-stranded probes were designed using the 24-nt flanking the PfAP2-G motif of the indicated upstream sequence. Probe sequences were as follows with capital letters indicating the AP2-G motifs and lowercase letters indicating the flanking sequences: *pfg27/25*: 5′-ttattagtatctGTACAC attggtatttgt-3′, *pf11-1*: 5′-tatatatatattGTACACatacatgtagtt-3′, *pfpeg4*: 5′-gacaataa agaaGTGTACACatatatcaataa-3′. The motifs were replaced by an equal number of adenines for 'no motif' probes. Replicates were conducted using the same materials on separate days.

Transcription profiling and associated analysis. Starting at 3 h post-invasion, tightly synchronized parasites were collected at eight time points with 6-h intervals. RNA isolation, cDNA generation/labelling, array hybridization and feature extraction was performed as described previously[28]. Cy5-labelled cDNA was hybridized

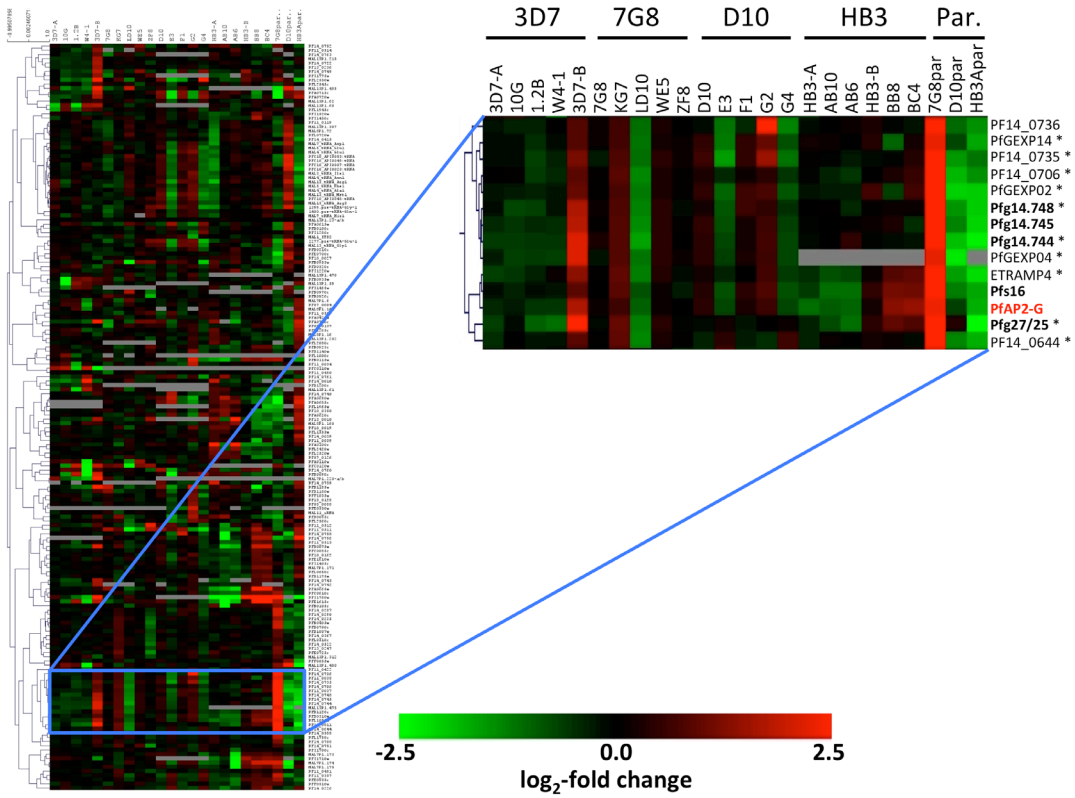
with a common Cy3-labelled reference pool on the *P. falciparum* 8×15K Agilent nuclear expression array (Gene Expression Omnibus (GEO) platform ID GPL17880). Relative transcript abundance was determined using a shared Cy3-labelled reference pool. All microarray data was submitted to the NCBI GEO repository (series accession number GSE52030). Genes were ordered by their average relative transcript abundance differences across the eight time points between the wild type (E5) and mutant (F12/Δ*pfap2-g*). Occurrences of the trimmed (6-nt) PfAP2-G motif were mapped using ScanACE to intergenic regions up to 2,000 bp upstream of the start codon as previously described[21] (see Supplementary Fig. 10 for motif). Significant enrichments of proteomic evidence and PfAP2-G motif occurrence were calculated using an unpaired two-sided *t*-test comparing the occurrences within the cluster of downregulated genes and their frequency genome-wide. Results were validated by qRT–PCR for a subset of downregulated genes using the primers in Supplementary Table 3 and methods described above. Statistically significant differences in relative expression levels were determined by two-sided *t*-test.

Luciferase assays. Equal numbers of synchronized, stably transfected parasites were isolated and saponin-lysed (0.05% in PBS) at ~18–30 h post-invasion and assayed using Bright-Glo Luciferase Assay System (Promega) as per the manufacturer's protocol on a Synergy H1 (Bio-Tek) plate reader. Statistical significance was determined using unpaired two-sided *t*-tests. Replicates were biological not technical.

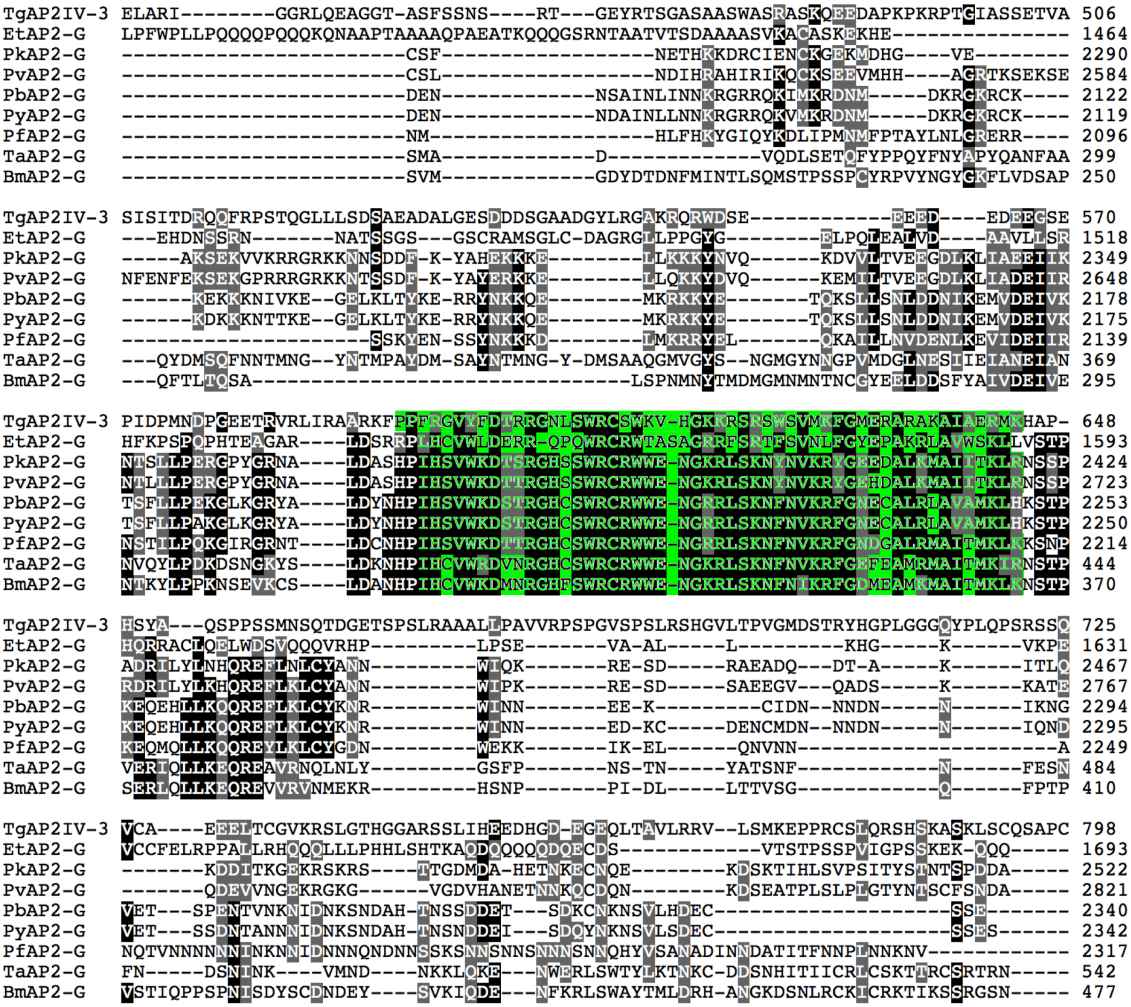
Next-generation sequencing and analysis. Genomic DNA was extracted (10 μg each) from E5, *Δpfap2-g*, GNP-A4 and F12 parasite lines. This genomic DNA was used to generate barcoded sequencing libraries for an Illumina TruSeq single-end sequencing run, analysed and visualized as described previously[29]. Genomic DNA for 3D7A[32], F12[12] and GNP-A4[33] was also used for whole genome sequencing at the Sanger Institute using Illumina GA II technology with 76-base paired-end reads. The raw sequence data were processed as described previously[30]. In brief, the raw data for each isolate was mapped onto the 3D7 reference genome (version 3) using the SMALT short read alignment algorithm[42]. High-quality SNPs and insertions and deletions (supported by bidirectional reads, and error rates less than one per 1,000 bp) in unique genomic regions were called using SAMtools (http://samtools.sourceforge.net). Regions of interest were inspected using the Artemis alignment viewer (http://www.sanger.ac.uk/resources/software/artemis/), and polymorphisms compared to publically available sequence data[16,30,43] processed as described above. Experimental confirmation of informative genomic variants was performed using capillary sequencing methods.

31. Cortés, A., Benet, A., Cooke, B. M., Barnwell, J. W. & Reeder, J. C. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. *Blood* **104,** 2961–2966 (2004).
32. Walliker, D. *et al.* Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236,** 1661–1666 (1987).
33. Taylor, C. J. *The role of two cyclic nucleotide phosphodiesterases in the sexual development of* Plasmodium falciparum *and* Plasmodium berghei. PhD thesis, Univ. London (2007).
34. Taylor, C. J., McRobert, L. & Baker, D. A. Disruption of a *Plasmodium falciparum* cyclic nucleotide phosphodiesterase gene causes aberrant gametogenesis. *Mol. Microbiol.* **69,** 110–118 (2008).
35. Maier, A. G., Braks, J. A. M., Waters, A. P. & Cowman, A. F. Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination. *Mol. Biochem. Parasitol.* **150,** 118–121 (2006).
36. Ménard, R. *Malaria: Methods and Protocols. Methods in Molecular Biology* Vol. 923, 2nd edn, 3–15 (Humana Press, Springer, 2013).
37. Farrell, A. *et al.* A DOC2 protein identified by mutational profiling is essential for apicomplexan parasite exocytosis. *Science.* **335,** 218–221 (2012).
38. Calderwood, M. S., Gannoun-Zaki, L., Wellems, T. E. & Deitsch, K. W. *Plasmodium falciparum var* genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J. Biol. Chem.* **278,** 34125–34132 (2003).
39. Cortés, A. *et al.* Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog.* **3,** e107 (2007).
40. Crowley, V. M., Rovira-Graells, N., Ribas de Pouplana, L. & Cortés, A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol. Microbiol.* **80,** 391–406 (2011).
41. Jiang, L. *et al.* Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. *Proc. Natl Acad. Sci. USA* **107,** 2224–2229 (2010).
42. SMALT - Wellcome Trust Sanger Institute. (http://www.sanger.ac.uk/resources/software/smalt/).
43. The Wellcome Trust Sanger Institute SRA Study ERP000190 (http://www.ebi.ac.uk/ena/data/view/ERP000190).

doi:10.1038/nature12920



Supplementary Figure 1: PfAP2-G clusters with early markers of gametocyte development. Clustering analysis of the expression patterns of 176 clonally variant genes (rows) in 21 parasite lines (columns) (see[1] for details) revealed co-clustering of PfAP2-G (in red) with known markers of early gametocyte development (Pfs16, Pfg27/25, Pfg14.744, Pfg14.745, Pfg14.748, in bold)[2,3] as well as numerous genes enriched in the early gametocyte proteome (marked with *)[4]. Values are the $\log_2$ of the expression fold-change relative to the average within a given comparison. Published gene names are shown when available.

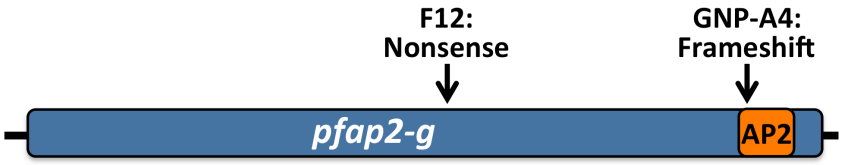


Supplementary Figure 2: Sequence conservation among PfAP2-G apicomplexan orthologs centers on the ApiAP2 DNA-binding domain. Apicomplexan orthologs (based on reciprocal BLAST analysis) were aligned using Clustal omega[5]. The AP2 domain (Pfam PF00847) is indicated in green and greater (50+%) conservation and identity is indicated with grey and black background respectively. No ortholog could be identified in *Cryptosporidium spp*.

(Gene IDs: *T. gondii* TgAP2IV-3: TGME49_318610 (ToxoDB), *E. tenella* EtAP2-G: ETH_00001325 (GeneDB), *P. knowlesi* PkAP2-G: PKH_143910 (PlasmoDB), *P. vivax* PvAP2-G: Translation of PviS_CM000455 1749161-1757917*, *P. berghei* PbAP2-G: PBANKA_143750 (PlasmoDB), *P. yoelii* PyAP2-G: PYYM_1441600 (PlasmoDB), *P. falciparum* PfAP2-G: PF3D7_1222600 (PlasmoDB), *T. annulata* TaAP2-G: XP_952218.1 (NCBI nr), *B. microti* BmAP2-G: CCF73073.1 (NCBI nr))

*The *P. vivax* ortholog (PVX_123760 at PlasmoDB) is misannotated, ORF and homology extends 7884nt (2628aa) upstream of the annotated start codon.

A



B **Clone GNP-A4: Insertion of an 'a' nucleotide results in a frameshift and a STOP codon at position 2190**

```
gga aga cgt tta agt aaa aaa ttt taa
 G   R   R   L   S   K   K   F   stop2190
```

Parental 3D7 with an asparagine residue at position 2190

```
gga aga cgt tta agt aaa aat ttt aat gtt aaa aga ...
 G   R   R   L   S   K   N   F   N2190V   K   R
```

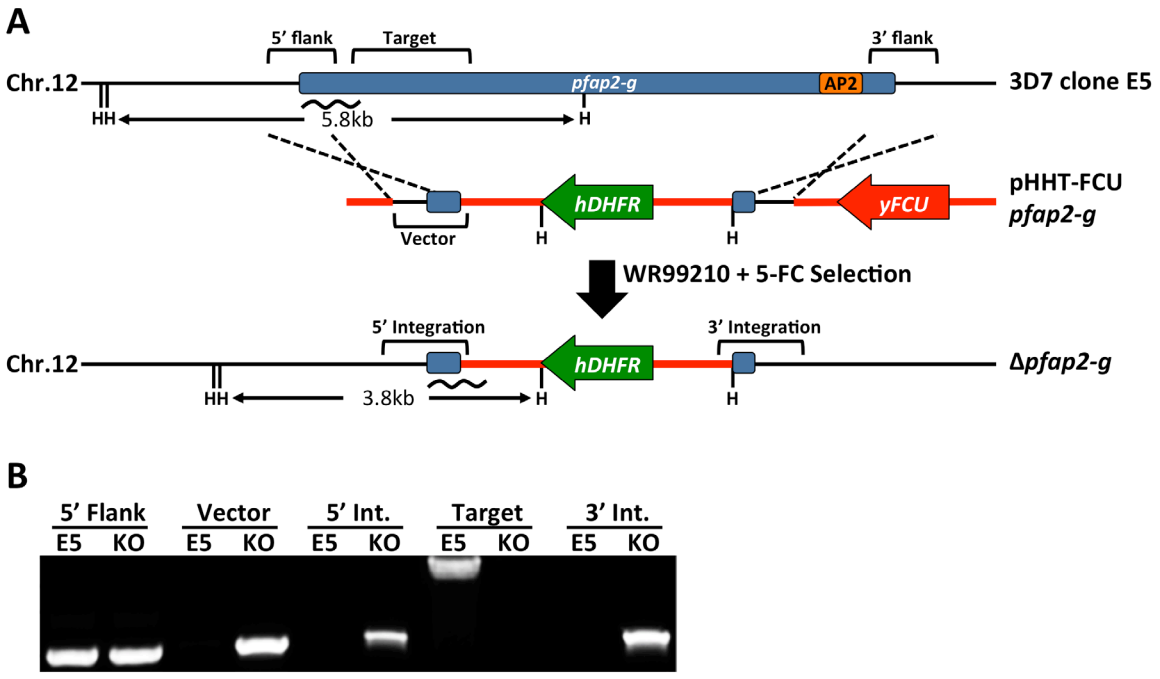
C **Clone F12: a-to-c SNP converts serine 1308 to a STOP codon**

```
gaa tta aat aat tat cca ata aat taa
 E   L   N   N   Y   P   I   N   stop1308
```

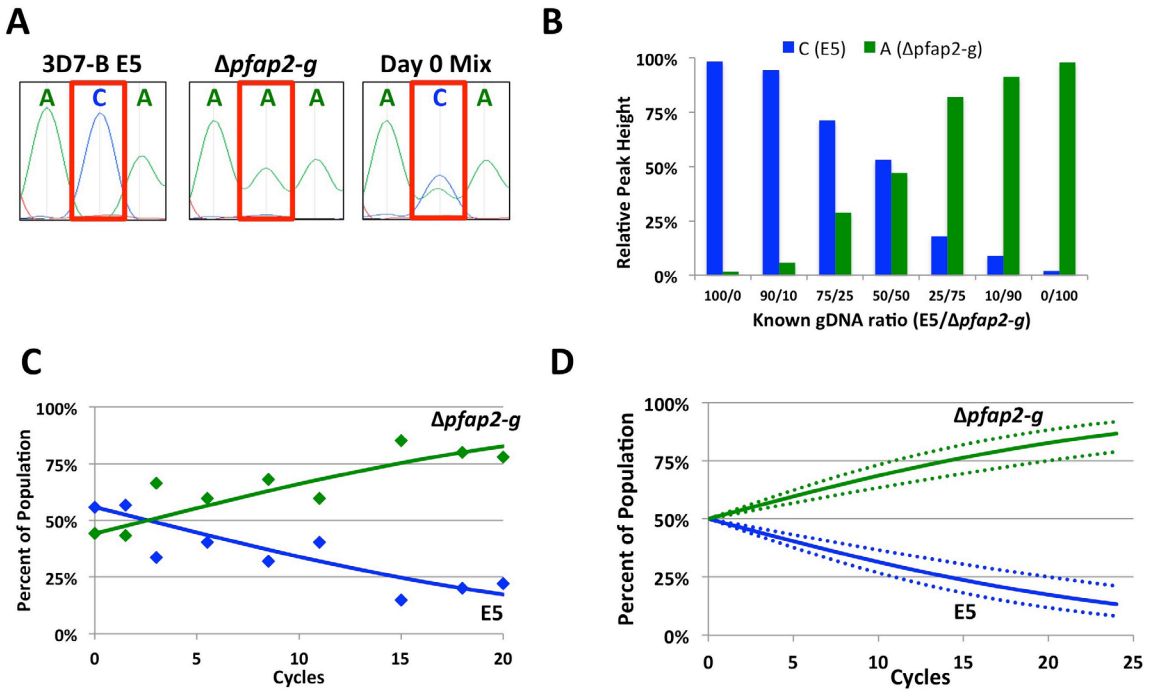
Parental 3D7 with a serine residue at position 1308

```
gaa tta aat aat tat cca ata aat tca aca caa gat tat ...
 E   L   N   N   Y   P   I   N   S1308T   Q   D   Y
```

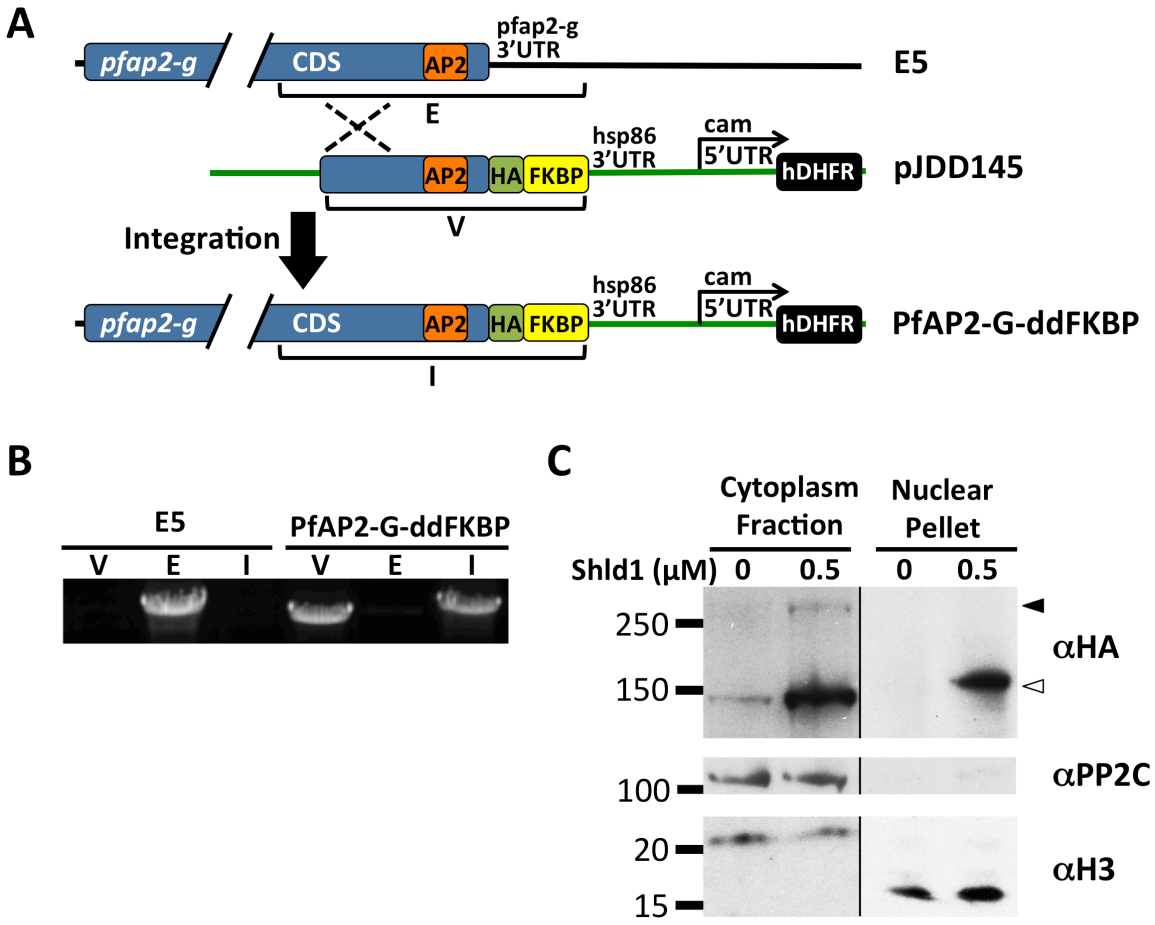
Supplementary Figure 3: *pfap2-g* mutations in gametocyte non-producer lines.
A) Positions of *pfap2-g* mutations in gametocyte non-producer lines F12 and GNP-A4. **B)** A single nucleotide insertion disrupts the coding sequence of *pfap2-g* in the gametocyte non-producer GNP-A4. **C)** A non-sense mutation disrupts the coding sequence of *pfap2-g* in the gametocyte non-producer F12.



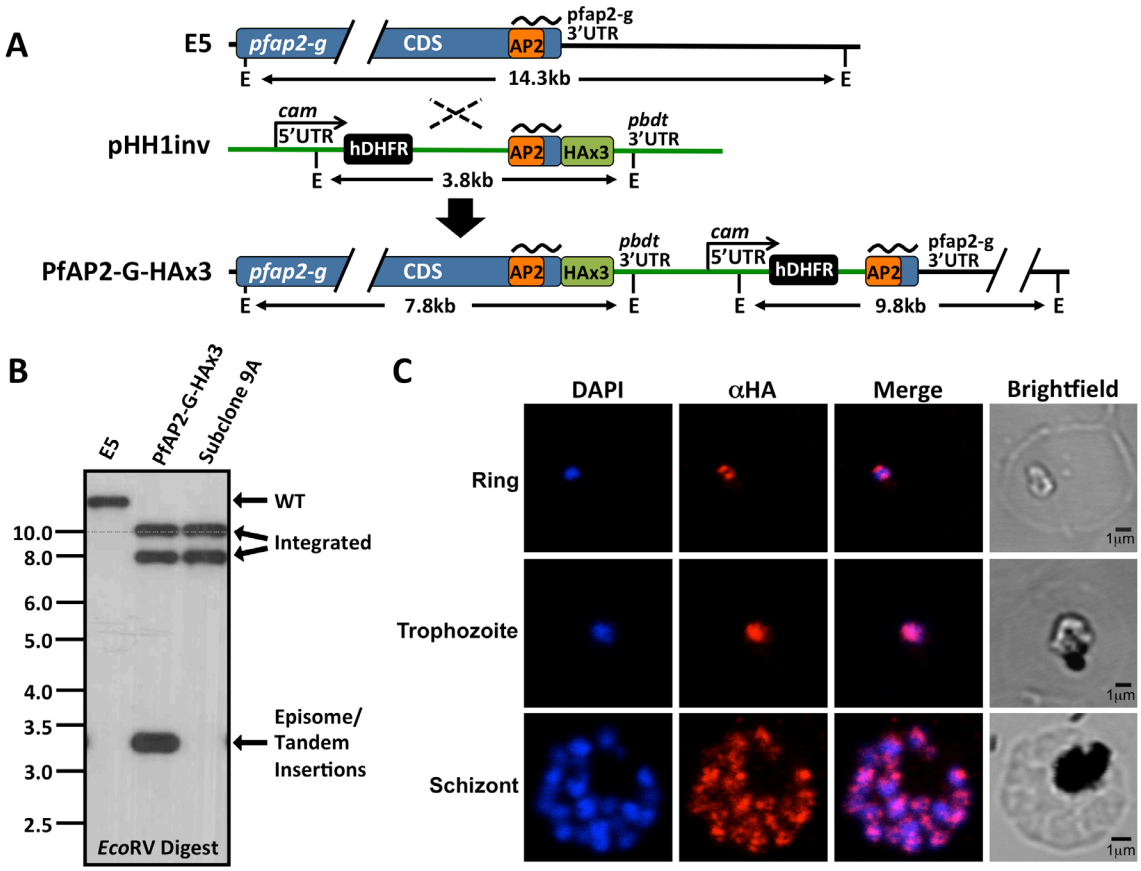
Supplementary Figure 4: PfAP2-G knockout strategy and validation. **A)** Overview of PfAP2-G knockout strategy using positive/negative selection for replacement by double homologous integration (dashed lines). Positions of HindIII restriction sites (H) and probe (undulating line) used for Southern blot analysis (See Figure 2B) are also shown. **B)** PCR validation of PfAP2-G knockout by double homologous recombination with amplified regions indicated by the brackets in the overview. The 5' targeting flank is present in both the E5 parent and Δ*pfap2-g* while the knockout construct along with its integration at the 5' and 3' targeting regions can only be detected in Δ*pfap2-g.* Conversely, the region targeted for deletion is only detectable in E5 but not Δ*pfap2-g.* Representative of n=3.



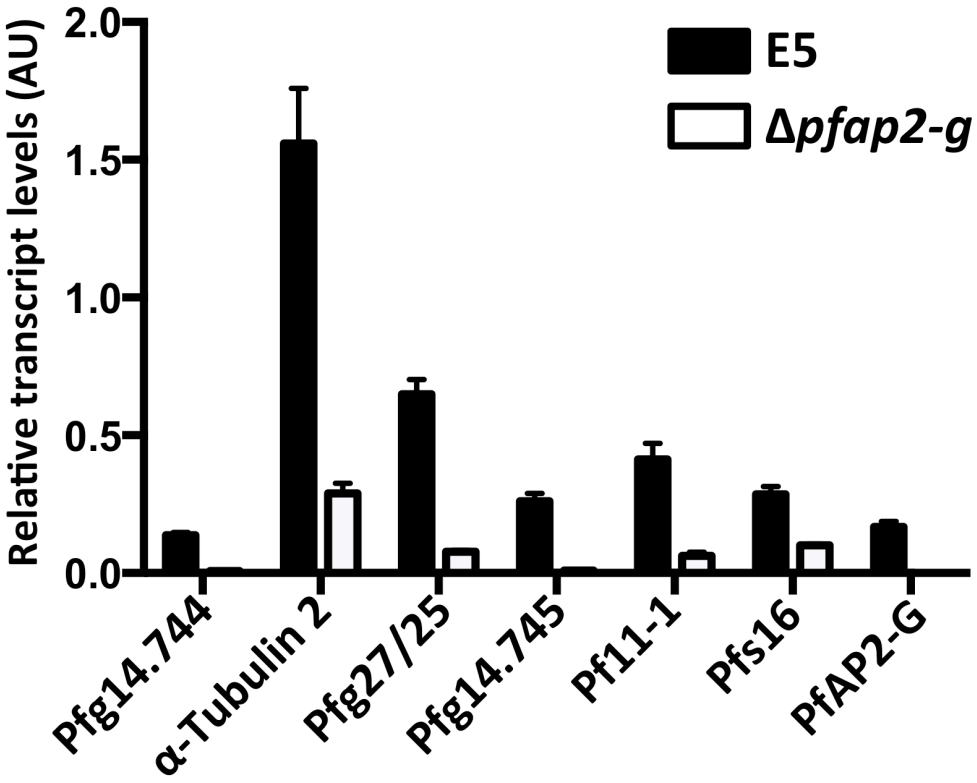
Supplementary Figure 5: Growth competition between 3D7-B E5 and Δ*pfap2-g*. **A)** Example sequencing trace peaks for the PFF0275c SNP. **B)** Validation of peak heights as an accurate measure of E5 (blue) to Δ*pfap2-g* (green) ratios (mean of n=2). **C)** Representative growth competition between 3D7-B E5 and Δ*pfap2-g*. **D)** Projected growth of equal E5 (blue) to Δ*pfap2-g* (green) mixture based on the measured 8.1% difference in growth rates (n=3). Dotted lines indicate the 95% confidence interval based on the observed 1.3% standard error in the growth rate measurements.



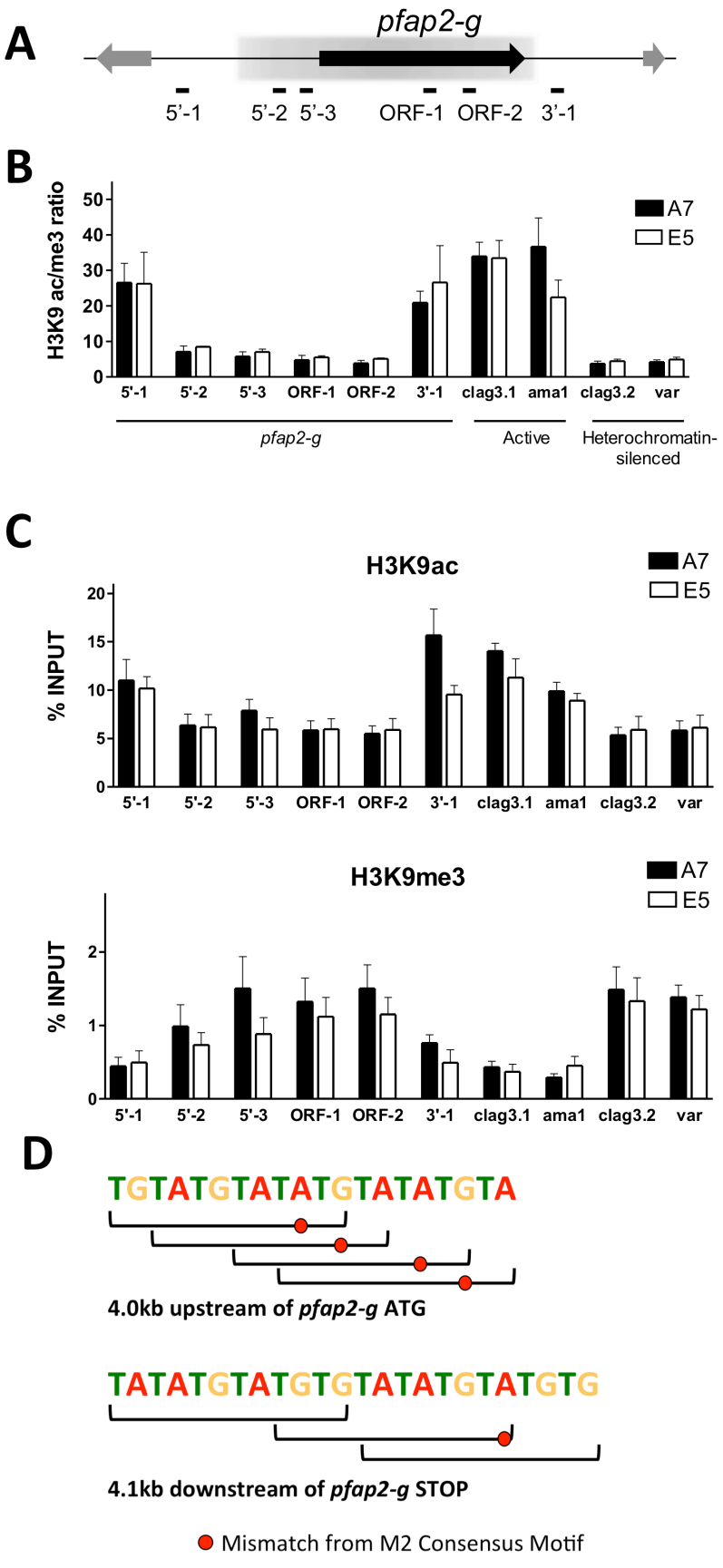
Supplementary Figure 6: Ligand-regulatable PfAP2-G-ddFKBP. **A)** Overview of the strategy for C-terminal tagging of PfAP2-G with HAx3-ddFKBP. PCR products detecting the pJDD145-pfap2-g vector (V), unmodified endogenous locus (E), and homologous integration (I) of the tagging construct are indicated with brackets. **B)** PCR validation of C-terminal tagging of the *pfap2-g* coding sequence with ddFKBP using the regions indicated in the overview above. Representative of n=2. **C)** PfAP2-g-ddFKBP was stabilized by 0.5 µM Shld1 in both the cytoplasmic fraction (cytoplasmic anti-PP2C loading control) and the nuclear pellet (nuclear anti-histone3 loading control) but not detectable in high-salt nuclear extracts. Full length PfAP2-G-ddFKBP (299.5 kDa) (black arrow) is N-terminally processed to a major band of ~150 kDa (white arrow). Representative of n=3.



Supplementary Figure 7: Nuclear localization of epitope-tagged PfAP2-G. A) Overview of PfAP2-G HAx3-tagging strategy. Distances between EcoRV restriction sites (E) and the probe used for Southern blot analysis (undulating line) are shown. Nucleotide distances not drawn to scale. **B)** Southern blot analysis demonstrating single copy integration of pHH1inv-pfap2-g-HAx3 into the E5 genome via single homologous recombination in the 9A subclone used for IFA. The position of DNA size markers (in kb) is shown. Single replicate. **C)** Immunolocalization of PfAP2-G-HAx3 subclone 9A parasites synchronized to ring, pigmented trophozoite and schizont stages, scale bar = 1μm. Representative of n=8 biological replicates at various time points.

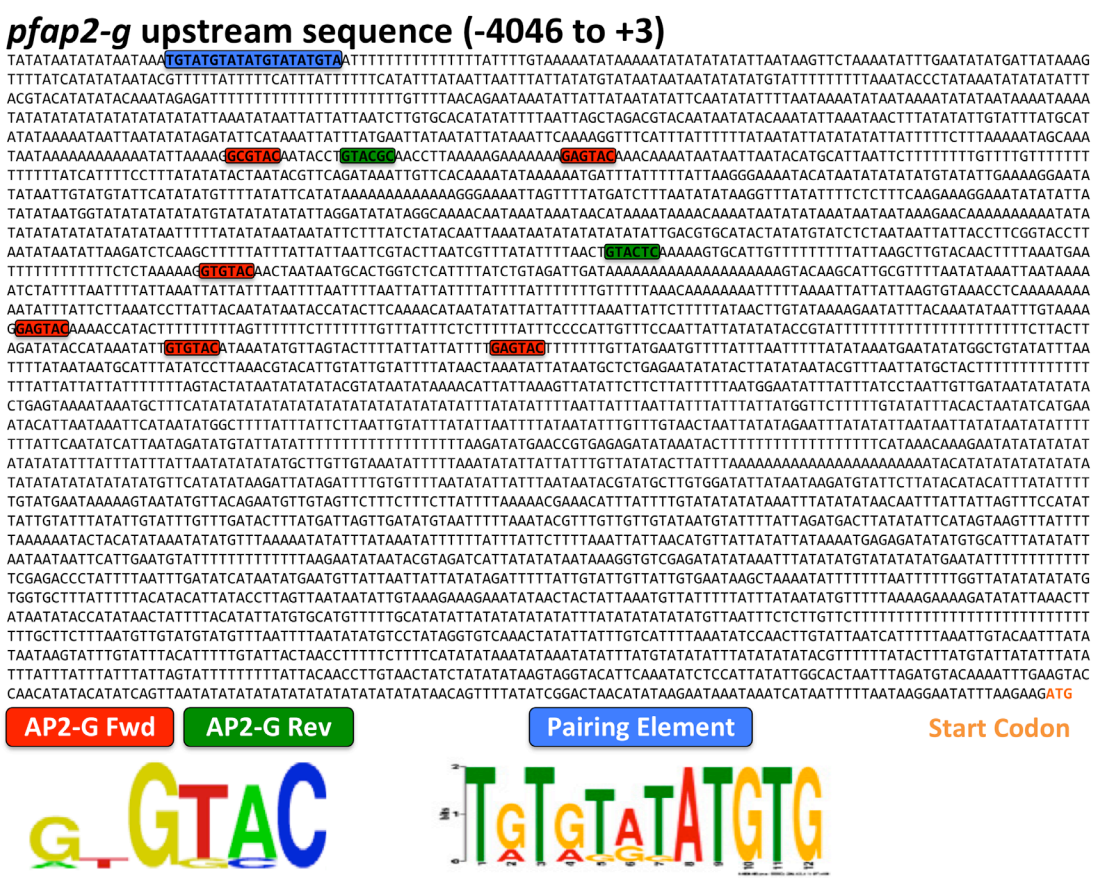


Supplementary Figure 8: Reduced early gametocyte transcript levels in late rings/early trophozoites in *Δpfap2-g* compared to the E5 parental line measured by quantitative RT-PCR. All six early gametocyte genes assayed show significant reductions in relative steady-state RNA levels (n=3, displayed in arbitrary units (AU) after normalization to seryl-tRNA synthetase transcript abundance, standard error shown).



Supplementary Figure 9: *pfap2-g* locus shows a heterochromatic conformation in bulk cultures of both A7 and E5 subclones.

A) Schematic showing the position of the primers used for ChIP analysis of the *pfap2-g* locus. Block arrows indicate the position of genes. The shaded region indicates the position of the *pfap2-g* heterochromatin domain as defined by the distribution of H3K9me3 and HP1[6,7] (taken from www.plasmodb.org.) **B)** ChIP results expressed as the ratio of % input for H3K9ac divided by % input for H3K9me3 (n=3, standard error shown). Parasite lines A7 and E5 are subclones of 3D7-B, which stably expresses *ama1* and *clag3.1* (active controls) but keeps *clag3.2* and the *var* PFL1950w/ PF3D7_1240300 silenced[8-9]. **C)** Same results expressed as % input for H3K9ac and H3K9me3. **D)** The *pfap2-g* locus is flanked by arrays of insulator-like pairing elements based on close matches to the M2 motif as described in Avraham *et al*.[10].



Supplementary Figure 10: Potential *pfap2-g* upstream regulatory elements. Forward (red) and reverse (green) PfAP2-G cognate motifs and the upstream pairing element are shown with respect to the *pfap2-g* start codon (orange). Truncated six-base PfAP2-G cognate motif[11] and M2 Pairing Element motif[10] are shown for reference.

Supplementary Figure Legend References:

1. Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res* **22,** 925–938 (2012).
2. Pradel, G. Proteins of the malaria parasite sexual stages: expression, function and potential for transmission blocking strategies. *Parasitology* **134,** 1911–1929 (2007).
3. Eksi, S. *et al.* Identification of a subtelomeric gene family expressed during the asexual–sexual stage transition in *Plasmodium falciparum*. *Mol Biochem Parasitol* **143,** 90–99 (2005).
4. Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Molecular & Cellular Proteomics* **9,** 1437–1448 (2010).
5. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539 (2011).
6. 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).
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. Crowley, V. M., Rovira-Graells, N., Ribas de Pouplana, L. & Cortés, A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol Microbiol* **80,** 391–406 (2011).
9. Cortés, A. *et al.* Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog* **3,** e107 (2007).
10. Avraham, I., Schreier, J. & Dzikowski, R. Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite *Plasmodium falciparum*. *Proc Natl Acad Sci USA* **109,** E3678-E3686 (2012).
11. 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).

Supplementary Table 3: Primers used for qRT-PCR validation of microarray results

Name	Sequence
PF14_0744_F2	AATCCTTGTATCCAAAGAAATGT
PF14_0744_R2	TGAAGAAGAATAATTTGCAGAAG
PFD1050w_F2	GAAATTGTTGATGTATGTTTGGA
PFD1050w_R2	TCTAATAATAAACAACCAAGACC
PF11_0037_F2	TTAGAAATTTGTATGAAGATGAGT
PF11_0037_R2	TAACATTGCTTCTACTCATACCA
PF14_0745_F2	GATGCTCGATATTTTTATACTGA
PF14_0745_R2	GTATTTCACTTAAGGAACATAGAA
PF10_0374_F2	CAAGAGTACATTGTATATTGTGG
PF10_0374_R2	CATTATCATCAATATTTACCATC
PFD1035w_F2	GAAAAGTTATGGAGATACAATCA
PFD1035w_R2	TCATATTCTTCAATTCTGATTCG
PFL1085w_F2	GCTACCTTACTAATGTATTAAGTC
PFL1085w_R2	CATATTATTAAAGAATTGGTTCACG
PF07_0073_F	AAGTAGCAGGTCATCGTGGTT
PF07_0073_R	TTCGGCACATTCTTCCATAA