A transcriptional switch underlies commitment to sexual development in malaria parasites

Björn F. C. Kafsack, Núria Rovira-Graells, Taane G. Clark, Cristina Bancelles, Valerie M. Crowley, Susana G. Campino, April E. Williams, Laura G. Drought, Dominic P. Kwiatkowski, David A. Baker, Alfred Cortés & Manuel Llinás

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. 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 prone to spontaneous activation. 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.

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. 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 api-complex apoAP2 (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 and have thus far been found to regulate several of the parasite’s developmental transitions, including ookinete formation and oocyst sporozoite maturation within the mosquito, and development in the mammalian liver. 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 $n = 2$).

---

1Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. 2Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), Barcelona, 08036 Catalonia, Spain. 3Institute for Research in Biomedicine (IRB), Barcelona, 08028 Catalonia, Spain. 4Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. 5Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. 6Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. 7Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, 08010 Catalonia, Spain. 8Wellcome Trust Sanger Centre for Human Genetics, Oxford OX3 7BN, UK. 9Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. 10Welcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. 11Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. 12Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. 13Division 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.).

©2014 Macmillan Publishers Limited. All rights reserved
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 production13,14. 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 mutagenesis15. 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 isolates16, 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 (Δpfap2-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 Δpfap2-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 Δpfap2-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 development13–15 (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 Δpfap2-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 degradation17,18, 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 Δpfap2-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. 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 Δpfap2-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).
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 (P < 0.017). These results implicate PfAP2-G as a transcriptional switch that controls sexual differentiation byactivating 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 Δ pfap2-g locus shares many features that have been associated with the epigenetic silencing of multigene families in P. falciparum, such as high levels of the H3K9me3 histone modification, associated binding of heterochromatin protein 1 (PHH1), and perinuclear localization. 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 predominately 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 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 committed to the same developmental fate. Furthermore, although the PfAP2-G–positive fraction varied between experiments, it was highly predictive of subsequent gametocyte formation in commitment assays (R² = 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—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. 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. (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. Last, ligand-regulatable PfAP2-G is not only a powerful new tool for studying malaria biology but also a new avenue for the development of anti-malaria drugs.
Gametocyte induction was performed according to published methods. All parasites were grown in media containing AlbuMax II and synchronized by standard methods.

**METHODS SUMMARY**

**Parasites and strains.** Δpfap2-g knockout parasites were generated by transfection of 3D7-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (hidhf) / negative (fts) selection. pfap2-g-djkfbp parasites were generated by transfection of 3D7-B E5 with pDID145-pfap2-g (Supplementary Fig. 6). Parasites expressing PfAP2-G–HAx3 were generated by transfection of 3D7-B E5 with pH11uv–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. Microarray. Starting at 3 h post-invasion, tightly synchronized parasites were collected at eight time points with 6-h intervals. DNA 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 ranked ordered by their average relative transcript abundance differences across the eight time points between the wild type (E5) and mutant (F12 or Δpfap2-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 Δpfap2-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.


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

Acknowledgements We would like to thank C. Klein, T. Campbell and A. Schierer for technical assistance and are grateful to O. Billker, C. Flocke, 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. Gorvet 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-PF2c antibody and D. Goldberg for Std1. 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 SAF2010-201111. 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 (090851; 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 Δpap2-g knockout, PIAP2-G–ddFKBP and luciferase lines and designed, performed and analysed the microarray, gel shift, luciferase and ligand-regulatable gametogenesis experiments. V.M.C. performed qRT–PCR validation. A.E.W. prepared Δpap2-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 chimpanzee immunoprecipitation experiments. C.B. and A.C. generated the PIAP2-G–HA x 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 (ER5011445), 3D7A (ERS011446) and GNP-A4 (ERS011447) and study number SRP035432 for samples E5 (SR529791) and ptpap2-g (SR529811)). 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).

©2014 Macmillan Publishers Limited. All rights reserved
METHODS
Parasites and strains. Parasite lines 3DT-A\(^1\), 3DT-B\(^2\) and F12\(^2\) have been described previously. Note that 3DT-A is not the same line as the Plasmodium falciparum/Plasmodium knowlesi hybrid 3DTA\(^3\), which was used only as a reference genome for next-generation sequence analysis. 3DT-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 (PfPDE6, 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\(^4\). A subsequent successful knockout of PfPDE6 produced gametocytes at normal rates and the true phenotype was a significantly lower extraglandular rate than parental parasites owing to a reduced ability of gametocytes to egress from red blood cells\(^5\). Pfap2-g knockout parasites were generated by transfection of 3DT-B E5 with pHHT-FCU-pfap2-g (Supplementary Fig. 4) followed by positive (hlfr) negative (fcu) selection using WR99210 and 5-fluoro-cytosine as described previously\(^6\). Resistant parasites were subcloned and verified by PCR and Southern blot. pfap2-g-ddfkbp parasites were generated by transfection of 3DT-B E5 with pJDD145-pfap2-g and selected on WR99210. After subcloning, integration was verified by PCR using a forward primer at position +6,269 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 3DT-B E5 with the plasmid pHH1-inv-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\(^7\).

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/Sacl sites of pHTT-FCU\(^8\), respectively, to generate pHTT-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 (gift from M. Duraisingh). Luciferase expression constructs. The hlfr selectable marker of pVL1DB1\(^9\) was replaced with blasticidin-S deaminase using the SacI/NotI sites to generate pVL-BSD. The var76 promoter was excised with Hpal/KpnI, blunted and re-ligated, destroying the 1,445 bp, 1,226 bp and 1,159 bp upstream of the pf11-1, pfg27/25 and ppfpeg4 open reading frame (ORF) and introducing a pfap2-g tag, separated by the sequence YLQ.

Chromatin immunoprecipitation (ChIP). ChIP experiments were performed as described previously\(^10\). In brief, cultures were synchronized to late trophozoite/ 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 07-032) 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), −4,142 to −3,102 (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 Pf1L1950w (upstream region, presumably 5’ untranslated region) have been described before\(^11\). Replicates were biological not technical.

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\(^12\). Cy5-labelled CDNA was hybridized to the microarray slides.
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 (ES) and mutant (F12/Apfap2-2-g). Occurrences of the trimmed (6-m) PIAP2-G motif were mapped using ScanACE to intergenic regions up to 2,000 bp upstream of the start codon as previously described22 (see Supplementary Fig. 10 for motif). Significant enrichments of proteomic evidence and PIAP2-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 ES, *Afpfap2*-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 previously29. Genomic DNA for 3D7A32, F1212 and GNP-A433 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 previously30. In brief, the raw data for each isolate was mapped onto the 3D7 reference genome (version 3) using the SMALT short read alignment algorithm42. 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 data16,30,45 processed as described above. Experimental confirmation of informative genomic variants was performed using capillary sequencing methods.

42. SMALT - Wellcome Trust Sanger Institute. (http://www.sanger.ac.uk/resources/software/smaltnet/).
43. The Wellcome Trust Sanger Institute SRA Study ERP000190 (http://www.ebi.ac.uk/ena/data/view/ERP000190).