

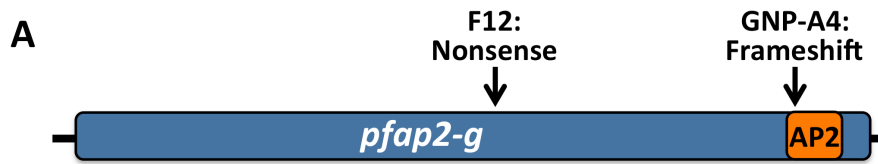
**Supplementary Figure 1: PfAP2-G clusters with early markers of gametocyte development.** Clustering analysis of the expression patterns of 176 clonally variant genes (rows) in 21 parasite lines (columns) (see<sup>1</sup> for details) revealed co-clustering of PfAP2-G (in red) with known markers of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

|           |   |      |
|-----------|---|------|
| TgAP2IV-3 | ELARI-----GGRLQEAGGT-ASFSSNS---RT---GEYRTSGASAASWASRASKQEFDAPKPKRPTGIASSETVA  | 506  |
| EtAP2-G   | LPFWPLLPOQQOPOQQQONAAPTAAAAQPAEATKQQQGSRNATAATVTSDAAAAASVKAASKKHE-----  | 1464 |
| PkAP2-G   | -----CSF-----NETHKDRCIENCKGRMDHG-----VE-----  | 2290 |
| PvAP2-G   | -----CSL-----NDIHAHIRIKCKSEVMMH-----AGRTKSEKSE  | 2584 |
| PbAP2-G   | -----DEN-----NSAINLNNKRGRRQKIMKRDNM-----DKRGRCK   | 2122 |
| PyAP2-G   | -----DEN-----NDAINLLNNKRGRRQKVMKRDNM-----DKRGRCK  | 2119 |
| PfAP2-G   | -----NM-----HLFHYGIQYKDLIPMNMFPPTAYLNLGRERR   | 2096 |
| TaAP2-G   | -----SMA-----D-----VQDLSETFYPPQYFNYPQANFAA  | 299  |
| BmAP2-G   | -----SVM-----GDYDTDNFMINTLSQMSTPSSPCYRPVYNGYCKFLVDSAP   | 250  |
| TgAP2IV-3 | SISITDROQFRPSTQGLLSDSAEADALGESDDSGAADGYLRGAKRORWDE-----EED-----EDEECSE  | 570  |
| EtAP2-G   | ---EHDNSSRN-----NATSSGS-----GSCRAMSGLC-DAGRGLLPPGYG-----ELPQIEATVD-----AAVLSR   | 1518 |
| PkAP2-G   | ---AKSEKVVKRRGRKKNNSDDF-K-YAHEKKE-----LLKKKYNVQ-----KDVVLTVEEGDLKLAEEITK  | 2349 |
| PvAP2-G   | NFENFEKSEKPRRRGRKKNSSDF-K-YAHEKKE-----LQKKYDVQ-----KEMILTVEEGDLKLADEIIR   | 2648 |
| PbAP2-G   | -----KEKKKNIVKE---GELKLYKE-RRYKKE-----MKRKKYE-----TOKSILSNLDDNIKEMVDEIVK  | 2178 |
| PyAP2-G   | -----KDKKKNITKE---GELKLYKE-RRYKKE-----MKRKKYE-----TOKSILSNLDDNIKEMVDEIVK  | 2175 |
| PfAP2-G   | -----SSKYEN-SSYNKKD-----LMKRRYEL-----OKATLLVVDENLKEVIDEIIIR   | 2139 |
| TaAP2-G   | ---QYDMSQFNNTMNG---YNTMPAYDM-SAYNTMNG-Y-DMSAAQGMVGS---NGMGYNGPVMDCINESITELANEIAN  | 369  |
| BmAP2-G   | ---QFTLQSA-----LSPNMNYTMDMGMNMNTNCCGYEELDDSFYAVDEIVE  | 295  |
| TgAP2IV-3 | PIDPMNPGCETRVRLIRARKEPFAAGVFDTRGNLSWRC@SWKV-HGKRS@S@SVMFCMERARAKALAE@M@HAP-   | 648  |
| EtAP2-G   | HFKPSPOPHTFAGAR---LDSRRPLHC@VWLDERR-OP@WR@WTASAG@R@FS@TF@S@N@L@FC@YEP@AKR@L@AW@SK@L@LV@STP  | 1593 |
| PkAP2-G   | N@S@L@L@P@E@R@G@P@Y@G@R@N@A---L@D@A@S@H@P@H@S@V@W@K@D@T@E@R@G@H@S@W@R@C@R@W@W@E---N@G@K@R@L@S@K@N@F@N@V@K@R@F@C@E@D@A@L@R@M@A@I@T@K@L@N@S@S@P       | 2424 |
| PvAP2-G   | N@I@L@L@P@E@R@G@P@Y@G@R@N@A---L@D@A@S@H@P@H@S@V@W@K@D@T@E@R@G@H@S@W@R@C@R@W@W@E---N@G@K@R@L@S@K@N@F@N@V@K@R@F@C@E@D@A@L@R@M@A@I@T@K@L@N@S@S@P       | 2723 |
| PbAP2-G   | T@S@L@L@P@E@K@L@K@R@Y@A---L@D@Y@N@H@P@H@S@V@W@K@D@T@E@R@G@H@S@W@R@C@R@W@W@E---N@G@R@L@S@K@N@F@N@V@K@R@F@C@E@C@A@L@R@I@A@V@A@M@K@L@H@K@S@T@P         | 2253 |
| PyAP2-G   | T@S@L@L@P@A@K@L@K@R@Y@A---L@D@Y@N@H@P@H@S@V@W@K@D@T@E@R@G@H@S@W@R@C@R@W@W@E---N@G@R@L@S@K@N@F@N@V@K@R@F@C@E@C@A@L@R@I@A@V@A@M@K@L@H@K@S@T@P         | 2250 |
| PfAP2-G   | N@S@L@L@P@O@K@I@R@G@R@N@T---L@D@C@N@H@P@H@S@V@W@K@D@T@E@R@G@H@S@W@R@C@R@W@W@E---N@G@R@L@S@K@N@F@N@V@K@R@F@C@E@D@A@L@R@M@A@I@T@K@L@K@S@N@P           | 2214 |
| TaAP2-G   | N@V@Q@L@P@D@K@D@S@N@G@K@Y@S---L@D@K@N@H@P@H@C@V@W@D@V@N@R@G@H@S@W@R@C@R@W@W@E---N@G@K@R@L@S@K@N@F@N@V@K@R@F@C@E@A@M@R@M@A@I@T@K@L@N@S@T@P           | 444  |
| BmAP2-G   | N@I@K@Y@L@P@K@N@S@E@V@K@C@S---L@D@A@N@H@P@H@C@V@W@K@D@N@R@G@H@S@W@R@C@R@W@W@E---N@G@K@R@L@S@K@N@F@N@V@K@R@F@C@E@M@E@A@M@A@I@T@K@L@N@S@T@P           | 370  |
| TgAP2IV-3 | H@S@Y@A---Q@S@P@S@M@N@S@Q@T@D@G@E@T@S@P@S@L@R@A@A@A@L@P@A@V@R@P@S@P@G@V@S@P@S@L@R@S@H@G@V@L@T@P@V@G@M@D@S@T@R@Y@H@G@L@P@G@G@Q@Y@L@Q@P@S@R@S@Q       | 725  |
| EtAP2-G   | H@O@R@R@A@C@T@O@E@L@W@D@S@V@Q@Q@V@R@H@P-----L@P@S@E-----V@A@-@A@L-----L-----K@H@G-----K-----V@K@P@E   | 1631 |
| PkAP2-G   | A@D@R@L@Y@L@N@H@O@R@E@F@L@N@C@Y@A@N@N-----W@I@Q@K-----R@E@-@S@D-----R@E@A@D@Q-----D@T@-@A-----K-----I@T@L@Q   | 2467 |
| PvAP2-G   | R@D@R@L@Y@L@K@H@O@R@E@F@L@K@C@Y@A@N@N-----W@I@P@K-----R@E@-@S@D-----S@A@E@E@G@V-----Q@A@D@S-----K-----K@A@T@E                                       | 2767 |
| PbAP2-G   | K@E@Q@E@H@L@L@K@O@R@E@F@L@K@C@Y@K@N@R-----W@I@N@N-----E@E@-@K-----C@I@D@N-----N@N@D@N-----N-----I@K@N@G   | 2294 |
| PyAP2-G   | K@E@Q@E@H@L@L@K@O@R@E@F@L@K@C@Y@K@N@R-----W@I@N@N-----E@D@-@K@C-----D@E@N@C@M@D@N-----N@N@D@N-----N-----I@Q@N@D                                     | 2295 |
| PfAP2-G   | K@E@Q@M@L@L@K@O@R@E@Y@L@K@C@Y@G@D@N-----W@E@K@K-----I@K@-@E@L-----Q@N@V@N@N-----N-----A   | 2249 |
| TaAP2-G   | V@E@R@I@O@L@L@K@Q@R@E@A@V@R@N@Q@L@N@L@Y-----G@S@F@P-----N@S@-@T@N-----Y@A@T@S@N@F-----N-----F@E@S@N   | 484  |
| BmAP2-G   | S@E@R@I@O@L@L@K@Q@R@E@V@V@R@V@N@M@E@K@R-----H@S@N@P-----P@I@-@D@L-----L@T@T@V@S@G-----Q-----F@P@T@P   | 410  |
| TgAP2IV-3 | V@C@A-----E@E@E@T@C@G@V@K@R@S@L@G@T@H@G@A@R@S@S@L@I@H@E@D@H@C@D---E@G@E@Q@L@T@A@V@L@R@R@V---L@S@M@K@E@P@P@R@C@S@I@O@R@S@H@S@K@A@S@K@L@S@C@Q@S@A@P@C | 798  |
| EtAP2-G   | V@C@C@F@E@L@R@P@A@L@L@R@H@Q@O@L@L@P@H@L@S@H@T@K@A@O@O@Q@O@Q@O@E@C@D@S-----V@T@S@T@P@S@S@P@V@I@G@P@S@S@K@E@K---Q@Q@-----                             | 1693 |
| PkAP2-G   | -----K@D@D@T@K@G@E@K@R@S@K@R@S-----T@T@G@D@M@A---H@E@T@N@K@E@C@N@Q@E-----K@D@S@K@T@I@H@L@S@V@P@S@I@T@Y@S@T@N@T@S@P@D@A-----                         | 2522 |
| PvAP2-G   | -----Q@D@E@V@N@G@E@K@R@G@K@G-----V@G@D@V@H@A@N@E@T@N@N@K@C@D@Q@N-----K@D@S@E@A@T@P@L@S@L@P@I@G@T@Y@N@I@S@C@F@S@N@D@A-----                           | 2821 |
| PbAP2-G   | V@E@T-----S@P@E@N@T@V@N@K@N@I@D@N@K@S@N@D@A@H---I@N@S@S@D@E@I---S@D@K@C@N@K@N@S@V@L@H@D@E@C-----S@S@E-----  | 2340 |
| PyAP2-G   | V@E@T-----S@S@D@N@T@A@N@N@I@D@N@K@S@N@D@A@H---I@N@S@N@D@E@I---S@D@Q@Y@N@K@N@S@V@L@S@D@E@C-----S@S@E@S-----  | 2342 |
| PfAP2-G   | Q@V@T@V@N@N@N@N@I@N@K@N@I@D@N@N@Q@N@D@N@N@S@K@S@N@S@N@S@N@N@N@S@N@O@H@Y@V@S@A@N@A@D@I@N@D@A@I@T@F@N@N@P@I@N@N@K@N@V                                 | 2317 |
| TaAP2-G   | F@N-----D@S@N@I@N@K-----V@M@N@D-----N@K@K@O@K@E-----N@W@E@R@L@S@W@Y@L@K@T@N@K@C---D@D@S@N@H@I@T@I@C@R@I@C@S@K@T@I@R@C@S@R@T@R@N-----                | 542  |
| BmAP2-G   | V@S@T@I@Q@P@P@S@P@N@I@S@D@Y@S@C@D@N@D@E@Y-----S@V@K@I@O@D@E-----N@F@K@R@L@S@W@A@Y@T@M@L@D@R@H---A@N@G@K@D@S@N@L@R@C@K@I@C@R@T@I@K@S@S@R@G@S@N-----  | 477  |

**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<sup>5</sup>. 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.



**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

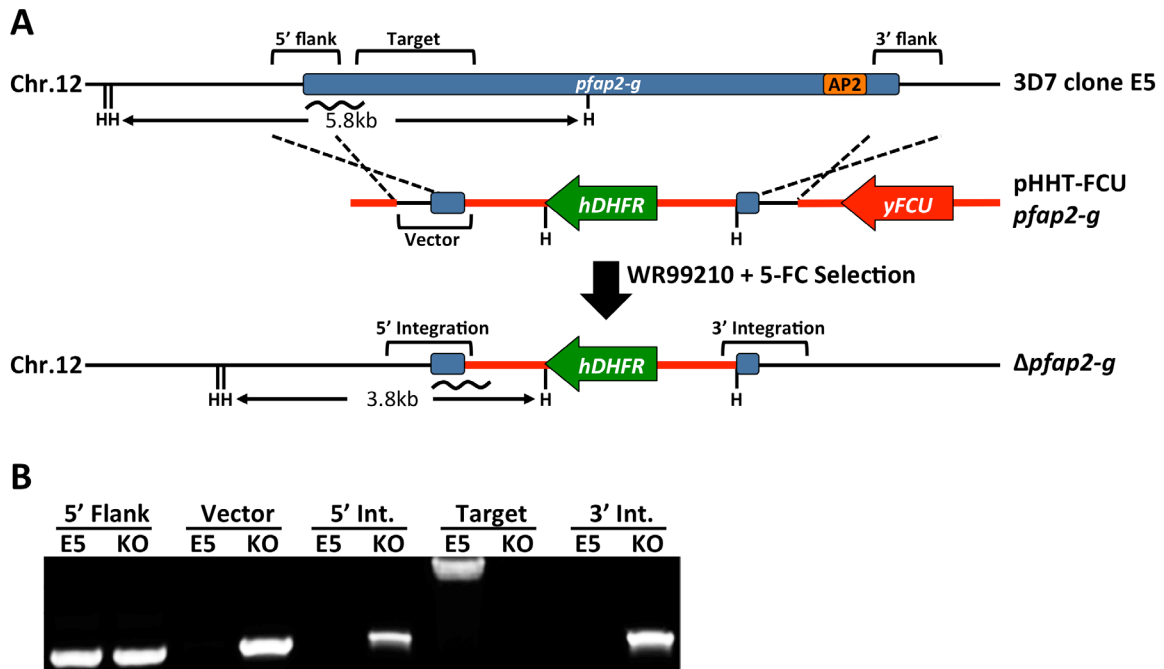
```
gaa tta aat aat tat cca ata aat taa
E L N N Y P I N stop1308
```

Parental 3D7 with a serine residue at position 1308

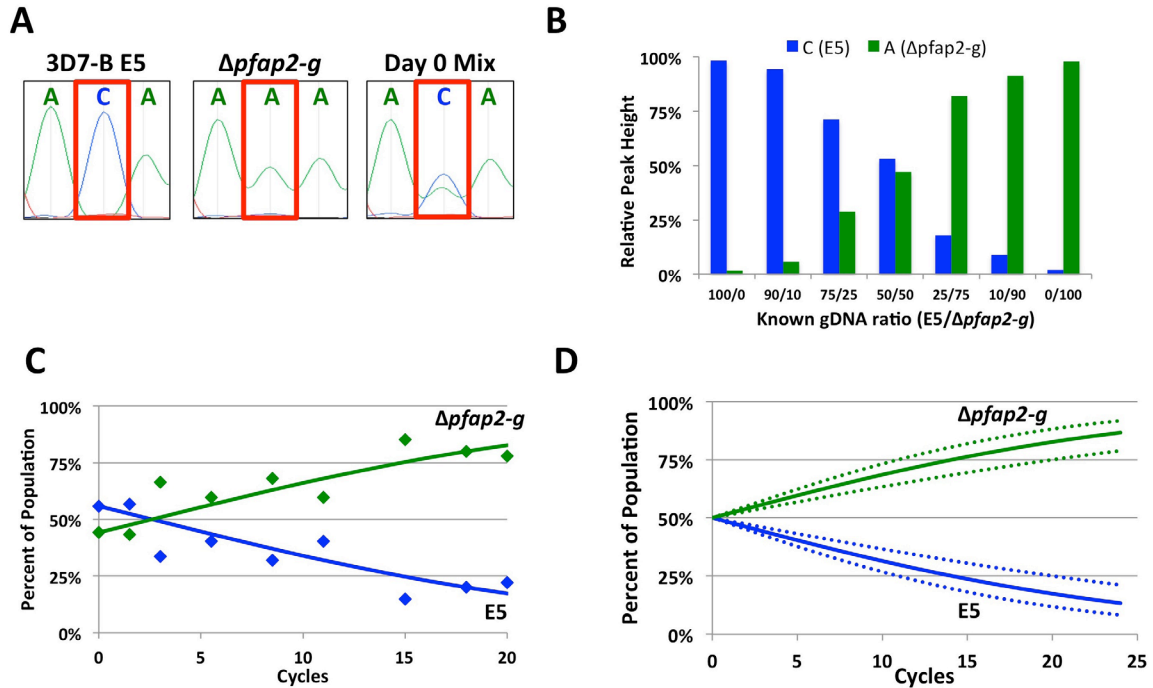
```
gaa tta aat aat tat cca ata aat tca aca caa gat tat ...
E L N N Y P I N S1308T Q D Y
```

**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.

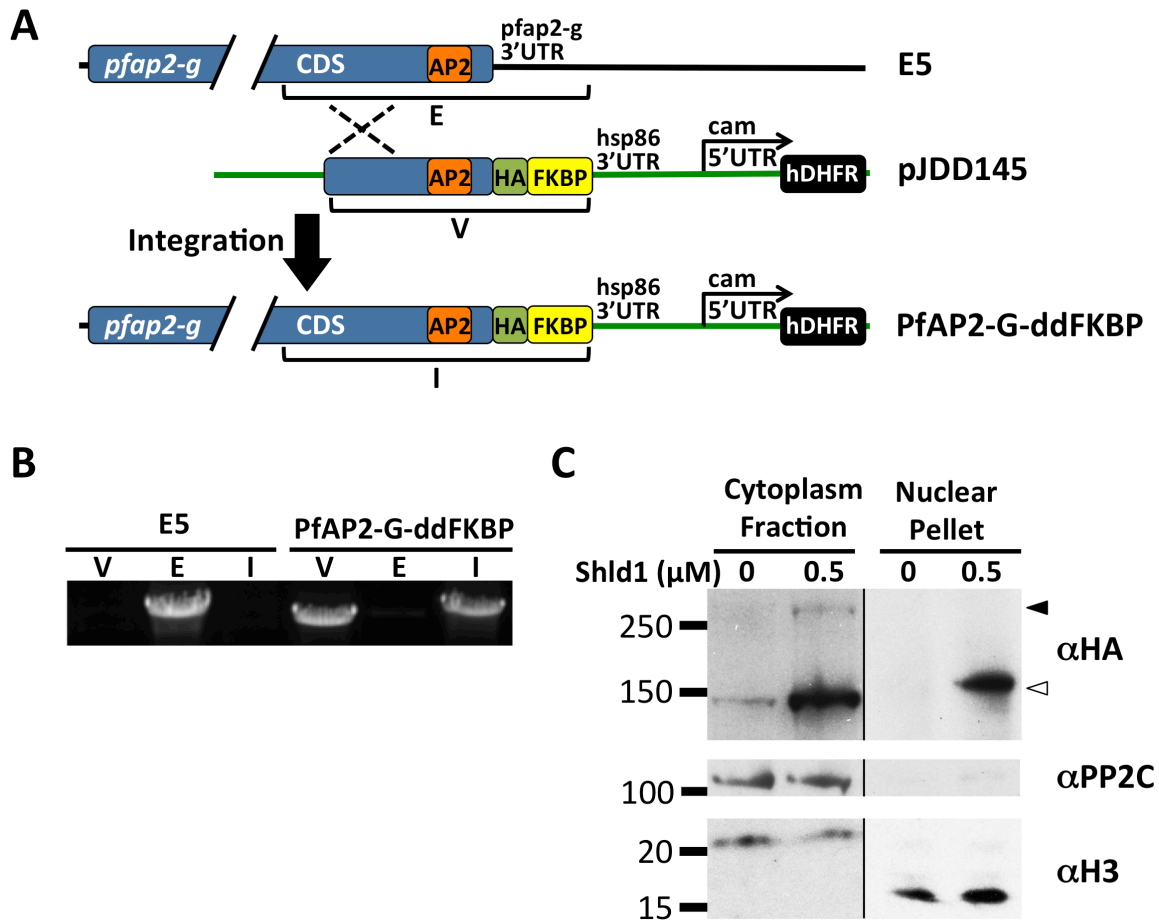


**Supplementary Figure 4: PfAP2-G knockout strategy and validation. A)** Overview of PfAP2-G knockout strategy using positive/negative selection for replacement by double homologous integration (dashed lines). Positions of HindIII restriction sites (H) and probe (undulating line) used for Southern blot 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  $\Delta$ *pfap2-g* while the knockout construct along with its integration at the 5' and 3' targeting regions can only be detected in  $\Delta$ *pfap2-g*. Conversely, the region targeted for deletion is only detectable in E5 but not  $\Delta$ *pfap2-g*. Representative of n=3.

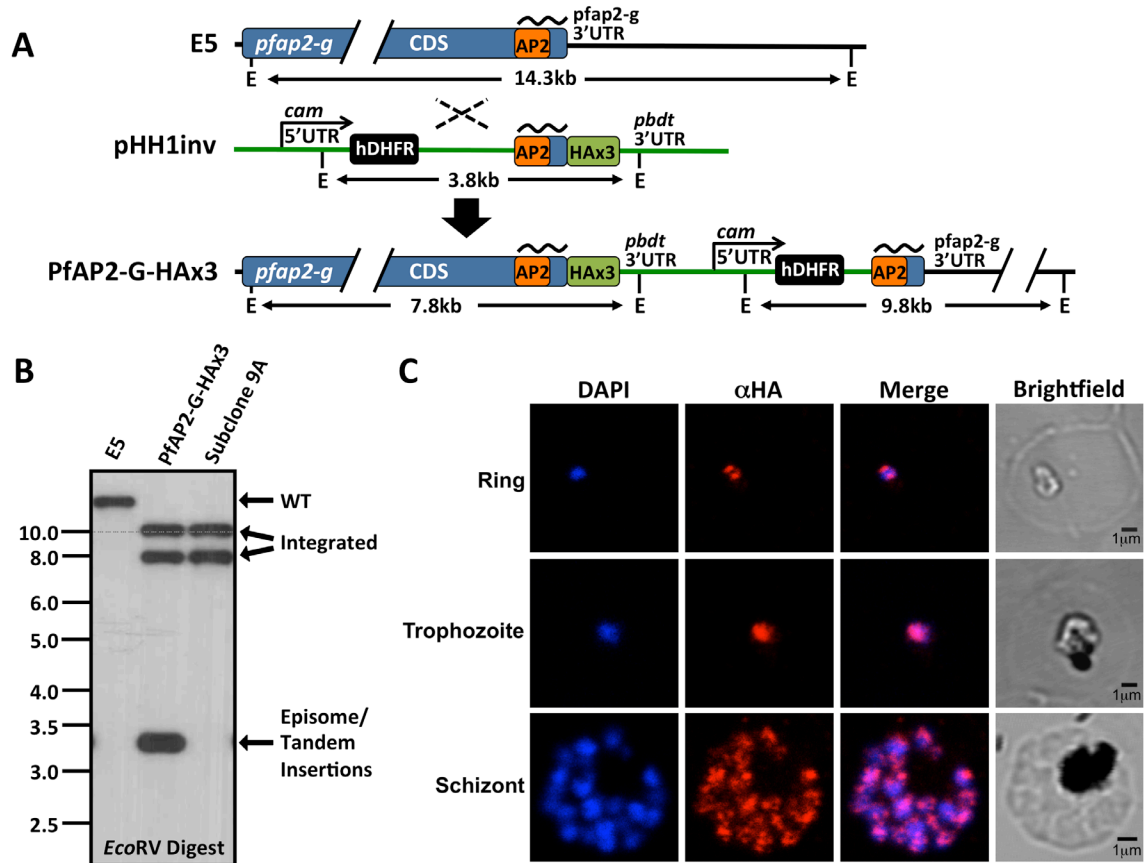


**Supplementary Figure 5: Growth competition between 3D7-B E5 and  $\Delta pfap2-g$ .**

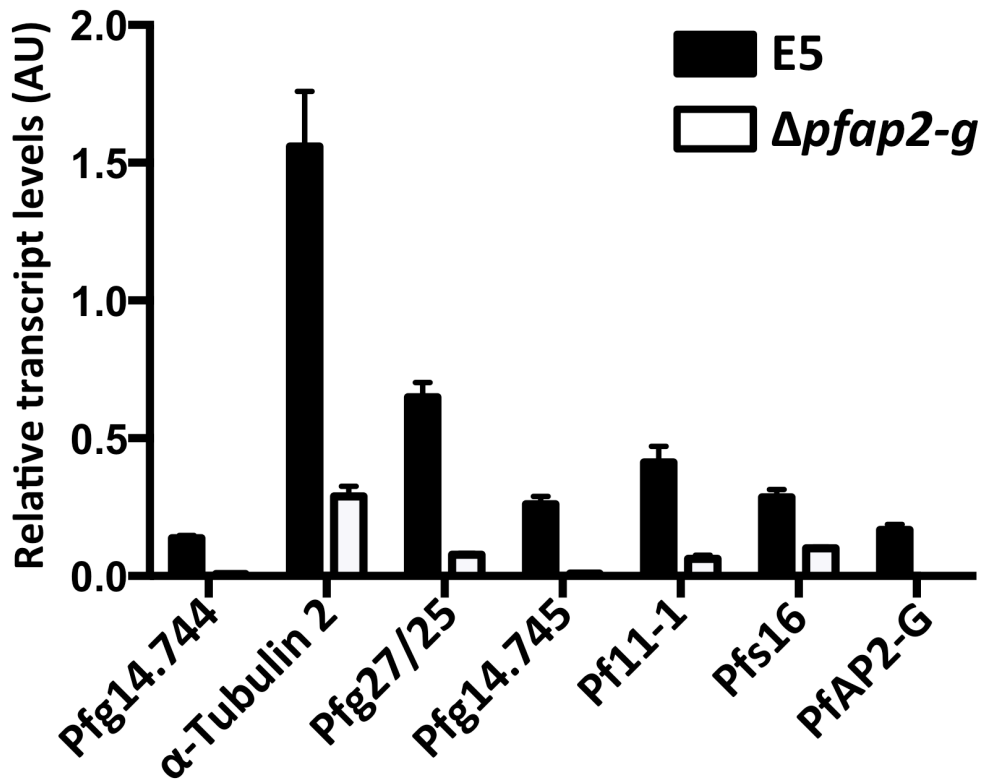
**A)** Example sequencing trace peaks for the PFF0275c SNP. **B)** Validation of peak heights as an accurate measure of E5 (blue) to  $\Delta pfap2-g$  (green) ratios (mean of n=2). **C)** Representative growth competition between 3D7-B E5 and  $\Delta pfap2-g$ . **D)** Projected growth of equal E5 (blue) to  $\Delta pfap2-g$  (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  $\mu\text{M}$  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  $\sim 150$  kDa (white arrow). Representative of  $n=3$ .

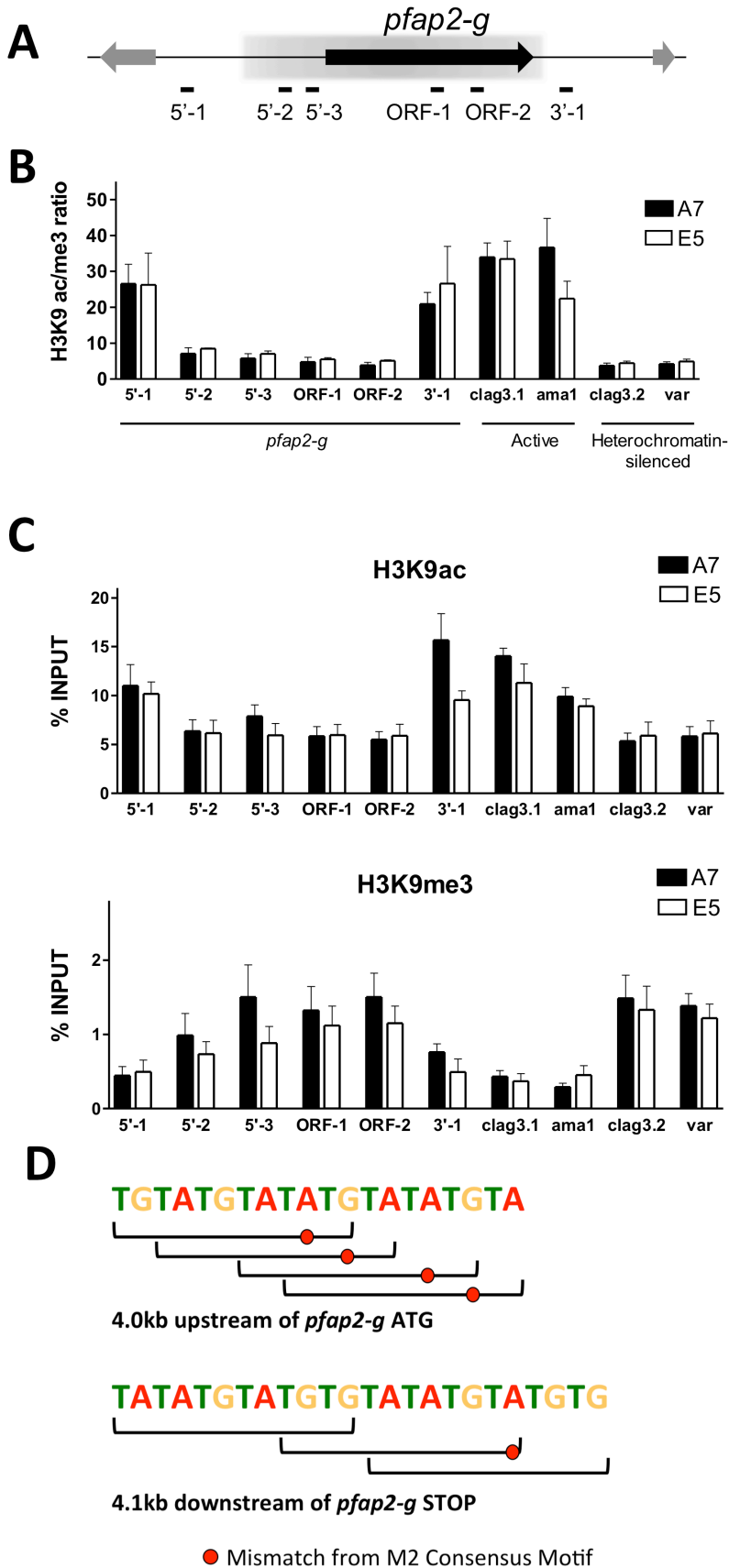


**Supplementary Figure 7: Nuclear localization of epitope-tagged PfAP2-G.** **A)** Overview of PfAP2-G HAx3-tagging strategy. Distances between EcoRV restriction sites (E) and the probe used for Southern blot analysis (undulating line) are shown. Nucleotide distances not drawn to scale. **B)** Southern 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  $\Delta pfap2-g$  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<sup>6,7</sup> (taken from [www.plasmodb.org](http://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<sup>8-9</sup>. **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.*<sup>10</sup>.



**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**

| <b>Name</b>  | <b>Sequence</b>          |
|--------------|--------------------------|
| PF14_0744_F2 | AATCCTTGTATCCAAAGAAATGT  |
| PF14_0744_R2 | TGAAGAAGAATAATTTGCAGAAG  |
| PFD1050w_F2  | GAAATTGTTGATGTATGTTTGG   |
| PFD1050w_R2  | TCTAATAATAACAACCAAGACC   |
| PF11_0037_F2 | TTAGAAATTTGTATGAAGATGAGT |
| PF11_0037_R2 | TAACATTGCTTCTACTCATACCA  |
| PF14_0745_F2 | GATGCTCGATATTTTATACTGA   |
| PF14_0745_R2 | GTATTTCACTTAAGGAACATAGAA |
| PF10_0374_F2 | CAAGAGTACATTGTATATTGTGG  |
| PF10_0374_R2 | CATTATCATCAATATTTACCATC  |
| PFD1035w_F2  | GAAAAGTTATGGAGATACAATCA  |
| PFD1035w_R2  | TCATATTCTTCAATTCTGATTCG  |
| PFL1085w_F2  | GCTACCTTACTAATGTATTAAGTC |
| PFL1085w_R2  | CATATTATTAAGAATTGGTTCACG |
| PF07_0073_F  | AAGTAGCAGGTCATCGTGGTT    |
| PF07_0073_R  | TTCGGCACATTCTTCCATAA     |