1 **PfAP2-MRP DNA-binding protein is a master regulator of parasite pathogenesis**

2 **during malaria parasite blood stages**

3

4 Amit Kumar Subudhi[1], Judith L. Green[2], Rohit Satyam[3†], Todd Lenz[4†], Rahul P.

5 Salunke[1], Muhammad Shuaib[1], Ioannis Isaioglou[5], Steven Abel[4], Mohit Gupta[4], Luke

6 Esau[6], Tobias Mourier[1], Raushan Nugmanova[1], Sara Mfarrej[1], Rupali Sivapurkar[1],

7 Zenaida Stead[1], Fathia Ben Rached[1], Yogesh Otswal[7], Rachid Sougrat[6], Ashraf Dada[8],

8 Abdullah Fuaad Kadamany[8], Wolfgang Fischle[7], Jasmeen Merzaban[5], Ellen Knuepfer[2#],

9 David J.P. Ferguson[9,10], Ishaan Gupta[11], Karine G. Le Roch[4], Anthony A. Holder[2*], Arnab

10 Pain[1,12*]

11

12 [1]Pathogen Genomics Group, Bioscience Program, Biological and Environmental Science

13 and Engineering (BESE) Division, King Abdullah University of Science and Technology

14 (KAUST), Thuwal, 23955-6900, Kingdom of Saudi Arabia

15 [2]Malaria Parasitology Laboratory, The Francis Crick Institute, London, NW1 1AT,

16 United Kingdom

17 [3]Department of Computer Science, Jamia Millia Islamia, Jamia Nagar, Okhla, New

18 Delhi, Delhi 110025, India

19 [4]Department of Molecular, Cell and Systems Biology, University of California Riverside,

20 Riverside, California, United States of America

21 [5]Cell Migration and Signaling Laboratory, Bioscience Program, BESE Division, King

22 Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900,

23 Kingdom of Saudi Arabia.

24 [6]KAUST Core Labs, KAUST, Thuwal, 23955-6900, Kingdom of Saudi Arabia

25 [7]Laboratory of Chromatin Biochemistry, Bioscience Program, BESE Division, King

26 Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900,

27 Kingdom of Saudi Arabia

28 [8]Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and

29 Research Center, Jeddah, Kingdom of Saudi Arabia

30 [9]Nuffield Department of Clinical Laboratory Science, University of Oxford, John

31 Radcliffe Hospital, Oxford OX1 2JD, United Kingdom

32   [10]Department of Biological and Medical Sciences, Faculty of Health and Life Sciences,

33   Oxford Brookes University, Oxford, United Kingdom

34   [11]Department of Biochemical Engineering and Biotechnology, Indian Institute of

35   Technology Delhi, New Delhi, India

36   [12]International Institute for Zoonosis Control; Hokkaido University, Sapporo, Japan

37   #Current Address: Molecular and Cellular Parasitology Laboratory, Department of
38   Pathobiology and Population Sciences, The Royal Veterinary College, Hawkshead Lane,
39   Hatfield, AL9 7TA, UK.
40
41   †These authors contributed equally to this work

42   *Corresponding Authors

43

44   Malaria pathogenicity results from the parasite's ability to invade, multiply within and

45   then egress from the host red blood cell (RBC). Infected RBCs are remodeled, expressing

46   antigenic variant proteins (such as PfEMP1, coded by the *var* gene family) for immune

47   evasion and survival. These processes require the concerted actions of many proteins, but

48   the molecular regulation is poorly understood. We have characterized an essential

49   *Plasmodium* specific Apicomplexan AP2 (ApiAP2) transcription factor in *Plasmodium*

50   *falciparum* (PfAP2-MRP; Master Regulator of Pathogenesis) during the intraerythrocytic

51   developmental cycle (IDC). An inducible gene knockout approach showed that PfAP2-

52   MRP is essential for development during the trophozoite stage, and critical for *var* gene

53   regulation, merozoite development and parasite egress. ChIP-seq experiments performed

54   at 16 hour post invasion (h.p.i.) and 40 h.p.i. matching the two peaks of PfAP2-MRP

55   expression, demonstrate binding of PfAP2-MRP to the promoters of genes controlling

56   trophozoite development and host cell remodeling at 16 h.p.i. and antigenic variation and

57   pathogenicity at 40 h.p.i. Using single-cell RNA-seq and fluorescence-activated cell

58   sorting, we show de-repression of most *var* genes in *Δpfap2-mrp* parasites that express

59   multiple PfEMP1 proteins on the surface of infected RBCs. In addition, the *Δpfap2-mrp*

60   parasites overexpress several early gametocyte marker genes at both 16 and 40 h.p.i.,

61   indicating a regulatory role in the sexual stage conversion. Using the Chromosomes

62   Conformation Capture experiment (Hi-C), we demonstrate that deletion of PfAP2-MRP

63   results in significant reduction of both intra-chromosomal and inter-chromosomal

2

64   interactions in heterochromatin clusters. We conclude that PfAP2-MRP is a vital

65   upstream transcriptional regulator controlling essential processes in two distinct

66   developmental stages during the IDC that include parasite growth, chromatin structure

67   and *var* gene expression.

68

69   **Main**

70   Malaria pathology is caused by asexual blood stage parasite growth. Merozoite invasion

71   of red blood cells (RBCs), intraerythrocytic replication and then egress from infected

72   RBC (iRBCs) drive proliferation[1]. The parasite modifies the iRBC surface, displaying

73   antigenically variant proteins encoded by multicopy gene families to escape immune

74   clearance[2]. These processes involve concerted expression of hundreds of proteins during

75   specific periods of the intraerythrocytic development cycle (IDC)[1-3], including proteins

76   encoded by genes that show clonal variant expression[4]. Regulation of gene expression

77   involves sequence-specific DNA binding proteins as transcriptional activators or

78   repressors and epigenetic modifiers[5]. Chromatin-mediated regulation of *P. falciparum*

79   gene expression controls invasion proteins and antigenically variant proteins[4].

80   Recruitment of epigenetic regulators is poorly understood. A likely possibility is that they

81   are recruited via DNA-binding regulatory proteins. In the genus *Plasmodium*, a family of

82   27 Apicomplexan-specific ApiAP2 DNA-binding proteins has been identified as the

83   major transcriptional regulators of various processes during development, differentiation,

84   and response to environmental changes[6-11]. However, few have been characterized in

85   detail to decipher their function and the role of the remainder is unknown.

86

87   Here, we use an inducible gene knockout approach[12,13] to show that PF3D7_1107800

88   codes for an ApiAP2 (PfAP2-MRP) essential for cell cycle, antigenic variation and

89   parasite egress and invasion. We establish that PfAP2-MRP directly or indirectly

90   regulates the expression of genes involved in host cell remodeling, antigenic variation,

91   egress, invasion and gametocytogenesis amongst the pathogenesis-related processes

92   corresponding to the two peaks of expression during the IDC. We also provide ChIP-seq

93   evidence that PfAP2-MRP binds to the promoter regions of genes predicted to be

94   involved in transcription, translation and nucleosome assembly. Chromosomes

95   Conformation Capture (Hi-C) analysis demonstrates that deletion of PfAP2-MRP results

96   in significant reduction of heterochromatin clusters leading to transcriptional activation of

97   gene involved in antigenic regulation and sexual differentiation. We conclude that

98   PfAP2-MRP is a master regulator of parasite growth, chromatin structure and *var* genes

99   expression.

100

101   **PfAP2-MRP is essential for parasite development and growth**

102   In a recent study[14], we identified 363 genes with transcripts displaying a 24 h (circadian-

103   like) rhythmic periodicity in the 48-hour *P. falciparum* IDC, one of which was an

104   ApiAP2 we now designate PfAP2-MRP (Extended Data Fig. 1a). The first peak of

105   expression at ~16 h.p.i. coincides with maximal *var* gene family expression. The second

106   peak at ~ 40 h.p.i., is just before maximal expression of known invasion- and egress-

107   associated genes, suggesting a role of PfAP2-MRP in regulating these genes (Extended

108   Data Fig. 1a).

109

110   PfAP2-MRP has one AP2 domain (residues 1487-1544, PFAM ID:PF00847), encoded in

111   the second exon (Extended Data Fig. 1b). PfAP2-MRP orthologs are found only in

112   *Plasmodium* spp. and in no other members of Apicomplexa (Extended Data Fig. 1c). The

113   PfAP2-MRP and its *P. berghei* ortholog (PBANKA_0939100) were identified as

114   essential during blood-stage development[15,16]. To investigate the function of PfAP2-

115   MRP, we generated a 3HA-tagged PfAP2-MRP inducible knockout *P. falciparum*

116   (PfAP2-MRP-3HA:loxP) using a parental line that expresses rapamycin (RAPA)-

117   inducible dimerizable Cre recombinase (DiCre)[13] (Fig. 1a, Methods).

118

119   Upon addition of RAPA to synchronized parasite culture, the *loxP*-flanked *pfap2-mrp*

120   second exon was efficiently excised (Fig. 1b). A time-series RNA-seq analysis revealed

121   that *loxP* flanked exon 2 deletion, and disruption of *pfap2-mrp* expression, occurred 14 to

122   16 hours after RAPA addition (Extended Data Fig. 1d), allowing us to determine when to

123   add RAPA to disrupt each peak of *pfap2-mrp* expression during the IDC. To ablate only

124   the second peak of functional *pfap2-mrp* expression at 40 h.p.i. in the cycle 0, RAPA was

125   added at ~16 h.p.i in the same cycle (i.e. cycle 0). To ablate only the first peak of

126   expression at 16 h.p.i. in cycle 1, RAPA was added at ~35 h.p.i. in cycle 0, with parasite

127   collection at 16 h.p.i. in cycle 1 (Extended Data Fig. 1e). From here onwards, *Δpfap2-*

128   *mrp* parasites (*Pfap2-mrp*-exon2 deleted) at 40 h.p.i. refers to parasites with disrupted

129   second peak of expression that were collected at 40 h.p.i. and *Δpfap2-mrp* parasites at 16

130   h.p.i. refers to parasites with disrupted first peak of expression that were collected at 16

131   h.p.i.

132

133   RAPA addition at ~16 h.p.i in cycle 0 ablated the second peak in the same cycle as

134   demonstrated by RNA-seq (Fig. 1c) and western blot (Fig. 1d), and these *Δpfap2-mrp*

135   parasites progressed through the IDC, forming morphologically normal segmented

136   schizonts (49 h.p.i [cycle 0], Fig. 1e), with no significant difference in the number of

137   nuclei in late schizonts of *Δpfap2-mrp* and DMSO (mock)-treated PfAP2-MRP-

138   3HA:loxP parasites (Fig. 1f). However, *Δpfap2-mrp* parasites failed to egress (Fig. 1e),

139   with a resultant dramatic reduction in parasitemia in subsequent cycles compared to

140   controls (Fig. 1g), indicating an essential role of PfAP2-MRP in parasite proliferation. In

141   egress assays with highly synchronized segmented mature schizonts under either static or

142   vigorous shaking (shear) conditions, no egress or increased parasitemia was observed for

143   *Δpfap2-mrp* parasites, in contrast to the increased parasitemia in control cultures,

144   irrespective of conditions (Fig. 1h); therefore, the blocked egress phenotype was not

145   corrected by mechanical disruption of the iRBC.

146

147   To look for morphological differences, we compared *Δpfap2-mrp* and control parasites

148   by indirect immunofluorescence with antibodies specific for proteins of the parasite

149   surface pellicle: GAP45, a protein of the glideosome/inner membrane complex (IMC),

150   and merozoite surface protein 7 (MSP7) (Fig. 1i and 1j). In *Δpfap2-mrp* compared with

151   control parasites, the distribution of these proteins was disordered. Electron microscopy

152   analysis of *Δpfap2-mrp* and parasites treated with compound 2 (an inhibitor of egress[17])

153   at 49 h.p.i. revealed that most mature schizonts contained fully formed merozoites in the

154   control sample (Fig. 1k and Fig. 1k, insert). In contrast, most *Δpfap2-mrp* schizonts were

155   at an earlier segmented stage (Fig. 1l and Fig. 1l, insert). In samples of 50 iRBCs, the

156   control group contained 83% mature schizonts compared to the *Δpfap2-mrp* group that

157 contained only 2% mature schizonts (Extended Data Fig. 1f). These data suggest that the

158 function of the second peak of *pfap2-mrp* expression is manifest in the final IDC stages

159 after nuclear division.

160

161 RAPA treatment from ~35 h.p.i. had little effect on parasite egress or invasion at the end

162 of cycle 0, but functional deletion of *pfap2-mrp* occurred well before the first expression

163 peak in the next cycle. These parasites with the first peak of *pfap2-mrp* expression

164 ablated were collected at 16 h.p.i. in cycle 1 (Extended Data Fig. 1e), and loss of

165 functional *pfap2-mrp* expression was confirmed by RNA-seq (Fig. 1m). Parasite

166 development stalled at the late trophozoite/early schizont stage in cycle 1 (Fig. 1n),

167 suggesting that the first peak of *pfap2-mrp* expression during the IDC plays a critical role

168 in parasite development immediately after its expression at ~16 h.p.i.

169

170 **PfAP2-MRP is a crucial regulator of malaria pathogenesis associated genes**

171 A comparative RNA-seq analysis of *Δpfap2-mrp* and control parasite populations at 16

172 and 40 h.p.i. identified 793 and 1,389 differentially expressed genes (FDR ≤ 0.05),

173 respectively (Extended Data Fig. 2a,b and Supplementary Data 1). Because disruption of

174 the second peak of *pfap2-mrp* expression caused parasites to stall at the late segmented

175 stage, we examined whether the observed transcriptional differences at 40 h.p.i. were due

176 to the knockout or reduced viability. We focused on the expression of 1,042 genes that

177 are known to have peak expression after 35 h.p.i.[5], and expected that death or a delay in

178 development of *Δpfap2-mrp* parasites would result in most of these genes being identified

179 as down-regulated. However, 658 (63%) of them showed no significant change in

180 expression level (Extended Data Fig. 2c), indicating that observed differences in gene

181 expression were due to *pfap2-mrp* deletion and not reduced viability.

182

183 The *P. falciparum* genome contains 59 intact *var* genes and their expression is mutually

184 exclusive: a single parasite expresses one *var* gene at a time, with the remainder of the

185 family remaining transcriptionally silent by heterochromatin formation[18]. Gene ontology

186 (GO) enrichment analysis of up-regulated genes in 16 and 40 h.p.i *Δpfap2-mrp* parasites

187 identified pathogenesis as the most enriched biological term (adjusted P = 1.23e-07 and

188 6.8e-10, respectively; Supplementary Data 2). The *var* gene family has the largest

189 number of members assigned to this GO term, with significant up-regulation of 24 and 29

190 *var* genes (*Padj* < 0.05) in *Δpfap2-mrp* compared to control parasites at 16 h.p.i. and 40

191 h.p.i., respectively (Fig. 2a,b). Data was validated by quantitative real time PCR (qRT-

192 PCR) (Fig. 2c and 2d). These results suggest that PfAP2-MRP may acts a repressor of

193 most *var* genes.

194

195 We also observed up-regulation of 4 of 8 *surfins* at 16 h.p.i. (Supplementary Data 1).

196 Several gene families coding for other variant proteins such as *rifins* (n=78/132), *stevors*

197 (n=13/30), and *Pfmc-2ms* (n=13/13) were significantly down-regulated (FDR ≤ 0.05) at

198 16 h.p.i., suggesting that these genes are positively regulated by PfAP2-MRP (Extended

199 Data Fig. 2d, Supplementary Data 1). While some members of the *hyp* and *phist* gene

200 families, encoding exportome-associated proteins showed up-regulation, others showed

201 down-regulation at 16 h.p.i., consistent with a functional diversification (Supplementary

202 Data 1). Together, these results suggest that *pfap2-mrp* expression at 16 h.p.i. has a major

203 role in regulating the expression of genes coding for proteins important in antigenic

204 variation and host cell remodeling.

205

206 Gene ontology (GO) enrichment analysis of the down-regulated genes in *Δpfap2-mrp*

207 parasites at 40 h.p.i. identified entry into and egress from the host cell as the two most

208 enriched biological process terms (Extended Data Fig. 2e, Supplementary Data 2),

209 consistent with our finding that deletion of *pfap2-mrp* critically affects late merozoite

210 development and egress. Of the top twenty down-regulated genes, fifteen function in

211 parasite egress or invasion (Fig. 2e). The expression of all top down-regulated genes

212 peaks just after the second peak of *pfap2-mrp* expression (Extended Data Fig. 1a),

213 consistent with the proposed role of PfAP2-MRP in controlling their expression. For

214 example, known egress-associated genes coding for SUB1 -2 and -3, SERA5 and -6,

215 Plasmepsin X, CDPK1 and -5, MSA 180, phospholipase, PKG, and MSP1 were down-

216 regulated (Supplementary Data 1). Strikingly, 36 out of 72 known invasion-associated

217 genes, including those coding for micronemal, rhoptry, merozoite surface, IMC and

218 motor proteins were also down-regulated (Fig. 2f, Supplementary Data 1). Western blots

219 and immunofluorescence assays were used to confirm the reduced abundance for a few

220 selected proteins (Fig. 2g). Altogether, the data suggest that PfAP2-MRP is essential to

221 the expression of egress and invasion-associated genes.

222

223 Other biological processes enriched in the group of down-regulated genes in *Δpfap2-mrp*

224 parasites at 40 h.p.i. include protein phosphorylation, actin cytoskeleton organization, and

225 fatty acid elongation. Protein phosphorylation has a significant role in control of both

226 egress and invasion in addition to other cell cycle events[19,20]. We identified 30 out of 105

227 genes annotated as involved in protein phosphorylation to be significantly down-

228 regulated (Extended Data Fig. 2f, Supplementary Data 2), including 10 out of 21 FIKK

229 kinase genes, the majority of which (17) are exported[21,22]. This suggests a broad

230 regulatory role for PfAP2-MRP in signaling processes during egress and invasion

231 (Supplementary Discussion).

232

233 **PfAP2-MRP is essential for trophozoite stage development**

234 We tested whether the gene expression pattern observed in the bulk RNA-seq data is also

235 observed at single cell resolution. Single-cell RNA sequencing (scRNA-seq) was

236 performed using parasites from 16 h.p.i. (control parasites and parasites with the first

237 peak of expression disrupted) and 40 h.p.i. (control parasites and parasites with only the

238 second peak of expression disrupted) stages. Based on the stage-specific annotation of

239 cells in the malaria cell atlas (MCA)[23], 16 h.p.i. cells are ring/early trophozoite parasites

240 and 40 h.p.i. cells are late trophozoite and schizont stage parasites (Fig. 2h). We observed

241 that more single *Δpfap2-mrp* parasites with disrupted first peak of expression at 16 h.p.i.

242 were classified as ring rather than trophozoite compared to controls (Fig. 2i). However,

243 no difference in stage was observed at 40 h.p.i. for parasites with disrupted second peak

244 of expression suggesting that it is not RAPA toxicity that is responsible for the observed

245 delay of *Δpfap2-mrp* parasite development at 16 h.p.i. The delay observed in the parasite

246 development may be an early phenotype resulting from the loss of the first *pfap2-mrp*

247 expression peak (Fig. 1n), which suggest that the first peak of expression is critical for

248 trophozoite development. No difference in developmental stages between control and

249 *Δpfap2-mrp* parasites with only disrupted second peak of expression at 40 h.p.i. further

250   suggests that differentially expressed genes at 40 h.p.i. (Extended Data Fig. 2b) were due

251   to *pfap2-mrp* deletion and not due to difference in developmental stage or reduced

252   viability.

253

254   The role of PfAP2-MRP in repressing *var* genes confirmed with single cells

255   From the scRNA-seq analysis, we observed significantly higher expression of *var*s (Fig.

256   2j) and *surfins* (Extended Data Fig. 2g) and down-regulation of *rifins* and *pfmc-2tms* in

257   *Δpfap2-mrp* parasites, supporting the bulk RNA-seq data (Extended Data Fig. 2g). We

258   further examined whether *pfap2-mrp* deletion activated expression of multiple *var* genes

259   in a single cell. Most control parasites expressed only one *var* gene following the

260   mutually exclusive expression pattern reported for *var* genes[18], but a few expressed two

261   or more *var* genes (Fig. 2k). A significantly higher number of *Δpfap2-mrp* parasites

262   expressed two or more *var* genes ($P= 6.23e^{-13}$, Fisher's exact test, Fig. 2k), indicating

263   apparent disruption of mutually exclusive *var* gene expression in the *Δpfap2-mrp*

264   parasites.

265

266   *Δpfap2-mrp* parasites expresses multiple surface PfEMP1

267   To determine whether in *Δpfap2-mrp* parasites expressing multiple *var* genes, EMP1s are

268   translated and exported to the iRBC surface, a FACS-based assay was developed with

269   pooled serum from patients infected with *P. falciparum*[24]. Amongst many exported

270   malaria proteins, PfEMP1 is a major target of naturally acquired antibodies[25], and

271   therefore we hypothesized that if *Δpfap2-mrp* parasites express multiple PfEMP1s, then

272   they will bind more antibodies in the serum against different PfEMP1s, compared to the

273   control parasites. We observed significantly more *Δpfap2-mrp* iRBCs than control

274   parasites with bound antibody ($P< 0.0001$, two-tailed t-test, Fig. 2l, Extended Data Fig.

275   2h). This result indicates that activation of multiple *var* genes and the translation and

276   transport of PfEMP1 to the iRBC surface occurs in *Δpfap2-mrp* parasites, and suggests

277   that PfAP2-MRP is involved broadly in silencing *var* gene expression.

278

279   PfAP2-MRP is a repressor of early gametocyte-associated marker genes

280 GO enrichment analysis of 526 up-regulated genes at 40 h.p.i. in *Δpfap2-mrp* schizonts

281 identified significant enrichment of many biological processes, including lipid and fatty

282 acid metabolism (Extended Data Fig. 3a and Supplementary Data 2). Genes such as

283 *elongation of fatty acids protein 3* (*elo3*) and *acyl-CoA synthetase 9* (*acs9*) have essential

284 roles in gametocyte development[16,26] and were up-regulated in 16 and 40 h.p.i. *Δpfap2-*

285 *mrp* parasites (Supplementary Data 1).   We examined the list of up-regulated genes for

286 other known or putative early gametocyte marker genes[27,28] and found 18 out of 28 genes

287 to be up-regulated in both 16 and 40 h.p.i. *Δpfap2-mrp* parasites (Extended Data Fig. 3b,

288 Supplementary Data 1), suggesting that PfAP2-MRP is also a repressor of commitment to

289 sexual stage development (Supplementary Discussion). We validated the expression of

290 some differentially regulated genes by qRT-PCR, and found strong agreement between

291 RNA-seq and qRT-PCR data (Extended Data Fig. 3c).

292

293 **PfAP2-MRP is a direct regulator of heterochromatin-associated genes**

294 To distinguish between direct and indirect targets of PfAP2-MRP, we performed

295 chromatin immunoprecipitation with the 3HA-epitope tagged PfAP2-MRP protein

296 followed by sequencing (ChIP-seq). We identified 1,081 and 640 ChIP-seq peaks at 16

297 and 40 h.p.i., respectively, of which 78% and 68% were in intergenic/promoter regions

298 upstream of at least one gene (with most located less than 2 kb upstream of an ATG

299 translational start site) (Extended Data Fig. 4a and 4b, Supplementary Data 3). We

300 validated a few PfAP2-MRP binding regions identified by ChIP-seq at 40 h.p.i. using

301 ChIP-qPCR, with complete agreement (Extended Data Fig. 4c). Consistent with previous

302 finding[29], we observed binding of PfAP2-MRP to both central and sub-telomeric

303 heterochromatin regions at both 16 and 40 h.p.i. stages in addition to its binding to the

304 promoter of euchromatic genes (Extended Data Fig. 4d). GO analysis identified encoded

305 proteins enriched in several processes, the most significant of which is antigenic variation

306 (*Padj*=8.36e-32 at 16 h.p.i and 6.42e-38 at 40 h.p.i.; Supplementary Data 4). We

307 observed that PfAP2-MRP binds significantly ($q < 0.05$) to the promoters of at least 37

308 and 33 *var* genes at 16 h.p.i. and 40 h.p.i, respectively; a total of 45 *var* genes, including

309 both sub-telomeric and internal *var* genes of all the upstream sequence (ups) types (Fig.

310 3a and Extended Data Fig. 4e).

311

312 PfAP2-MRP binding at both 16 and 40 h.p.i. was also enriched at the promoter of genes

313 encoding exportome proteins with relatively strong signals at 16 h.p.i stage and at the

314 promoter or intragenic regions of many other gene families for antigenic variant proteins

315 (*rifins*, *stevor*, *surfins* and *pfmc-2tms*) with relatively weaker yet statistically significant

316 signals at 16 h.p.i. stage (Fig. 3b and 3c, Supplementary Data 3). Most of these genes

317 were also differentially expressed in *Δpfap2-mrp* parasites suggesting that they are under

318 direct PfAP2-MRP control (Supplementary Data 1). These gene families for antigenic

319 variant and other exported proteins are located within the chromosomal central and

320 subtelomeric heterochromatin regions.

321

322 The promoters of genes associated with critical processes in early-stage development

323 such as translation (*Padj*=1.3e-10), glycolytic process (*Padj*=0.0005) and nucleosome

324 assembly (*Padj*= 0.02) were also strongly bound by PfAP2-MRP at 16 h.p.i. (Fig. 3c,

325 Supplementary Data 3 and 4) suggesting these processes are regulated by PfAP2-MRP,

326 and to some extent this may explain why development of parasites with disrupted first

327 peak of expression stalled at late trophozoite stage.

328

329 **PfAP2-MRP binds to its own promoter and those of another 13 *ApiAP2*s**

330 PfAP2-MRP binds to the promoter of a total of 14 *Pfapiap2* genes including its own

331 promoter at both 16h.p.i. and 40 h.p.i. (Fig. 3d, Extended Data Fig. 5 and Supplementary

332 Data 3). Nine of these 13 *apiap2s* were differentially expressed in *Δpfap2-mrp* parasites

333 at one or both of the 16 and 40 h.p.i. time points, suggesting that they are under the direct

334 control of PfAP2-MRP.

335

336 We observed binding of PfAP2-MRP to the *pfap2-i* promoter (Extended Data Fig. 6a),

337 and down-regulation of *pfap2-i* in *Δpfap2-mrp* parasites at 16 h.p.i. (Fig. 3d). It has been

338 reported that PfAP2-I binds the *pfap2-mrp* and its own promoter at 40 h.p.i.[30], and in this

339 study, we observed that PfAP2-MRP binds to its own promoter at both 16 h.p.i. and 40

340 h.p.i. (Extended Data Fig. 6b). Together, this information suggests that PfAP2-MRP

341 autoregulates its expression at both 16 h.p.i. and 40 h.p.i., positively regulates the

11

342 expression of *pfap2-i* at16 h.p.i.; PfAP2-I autoregulates its expression at 40 h.p.i. and

343 *pfap2-mrp* expression might be additionally controlled by PfAP2-I at ~ 40 h.p.i

344 (combinatorial regulation; Extended Data Fig. 6c).

345

346 We found that PfAP2-MRP bound to the same promoter region of many genes as PfAP2-

347 I and PfAP2-G, two proteins implicated in invasion[30] and gametocytogenesis[31],

348 respectively (Extended Data Fig. 6d and 6e, Supplementary Data 3), suggesting a

349 complex interplay of ApiAP2 DNA-binding proteins in control of gene expression

350 (Supplementary discussion). Genes with promoters bound by PfAP2-MRP, PfAP2-I, and

351 PfAP2-G include *gap45*; *msp1* and *-9*; *dblmsp*; e*tramp 4 -10.2* and *Pfap2-g* (Extended

352 Data Fig. 6e).

353

354 PfAP2-MRP is an indirect regulator of most invasion- and egress-associated genes

355 PfAP2-MRP binds to the promoters of only few invasion- and egress- associated genes at

356 40 h.p.i. (Extended Data Fig. 7 and Supplementary Data 3). These genes were

357 significantly (FDR ≤ 0.05; $\log_2$ fold change ≤ -1) or weakly ($\log_2$ fold change between -

358 0.5 to -1) down-regulated in *Δpfap2-mrp* parasites at 40 h.p.i. compared to controls

359 (Supplementary Data 1), suggesting that PfAP2-MRP directly regulates these genes.

360 However, PfAP2-MRP was not bound to the promoters of most top down-regulated

361 invasion- and egress-associated genes (Fig. 2b, Supplementary Data 1) except for

362 *msa180*, zinc finger protein (PF3D7_0818100), *sub1* and *msp6* (Supplementary Data 3).

363 These results suggest that the consequences of PfAP2-MRP knockout are both direct and

364 indirect on gene expression.

365

366 PfAP2-MRP binds to conserved sequence motifs

367 Analysis of the sequences bound by PfAP2-MRP at 16 h.p.i. and at 40 h.p.i. identified

368 RCATGCR ($6.1e^{-36}$, Fig. 4a); and GTGCR ($8.7e^{-28}$, Fig. 4b) and RCATGCA ($8.9e^{-12}$, Fig.

369 4c) as the most significantly enriched motifs respectively, excluding highly degenerate

370 motifs. GTGCR is very similar to the motif bound by PfAP2-I *in vitro*[30] ($P= 4.7e^{-03}$, Fig.

371 4d), consistent with the observation that PfAP2-MRP binds many sequences bound by

372 PfAP2-I (Extended Data Fig. 7d,e). RCATGCR (identified at 16 h.p.i.) is very similar to

373   RCATGCA (identified at 40 h.p.i., *P*=6.97e-5, Fig. 4c) suggesting that PfAP2-MRP

374   binding is sequence-specific with dynamic binding to similar sequence motifs in different

375   genomic regions at different IDC stages (Fig. 4e-g and Supplementary Data 3). Such

376   differential binding might be due to differences in PfAP2-MRP accessibility to different

377   genomic regions, such as that conferred by differential chromatin structure[5].

378

PfAP2-MRP associates with known and putative epigenetic regulators

380   Since PfAP2-MRP activates or represses the expression of most gene families under

381   epigenetic control and coding for clonally variant antigenic proteins, we proposed that

382   PfAP2-MRP recruits epigenetic regulator(s) to modulate target genes. In support of this

383   hypothesis, immunoprecipitation and mass spectrometry of PfAP2-MRP-3HA protein

384   complexes identified several known or putative histone modifiers and chromatin

385   remodelers, together with five other PfAP2 DNA binding proteins (Fig. 4h,i and

386   Supplementary Data 5). While PfAP2MRP-associated histone modifiers such as

387   PfSET10, PfSET6, and chromatin remodeler PfISWI proteins are transcriptional

388   activators[32-34], PfMORC-a putative chromatin remodeler along with PfHDAC1-a histone

389   deacetylase are associated with repressed genes and are regarded as transcriptional

390   repressors[35,36]. Many of the proteins associated with PfAP2-MRP were present at both 16

391   and 40 h.p.i., but others were unique to one time point. The association of PfAP2-I,

392   PfAP2-MRP and ISWI[30] is consistent with the observed overlap between PfAP2-MRP

393   and PfAP2-I binding sites and the significant similarity between the enriched motifs

394   (Extended Data Fig. 7d,e).

395

Depletion of PfAP2-MRP increases chromatin accessibility

397   While the relationship between chromatin organization and gene regulation remains

398   unclear, parasite chromatin is organized into euchromatin and heterochromatin clusters

399   and can act as a scaffold facilitating gene expression[37-39]. Furthermore, considering the

400   strong effect of disruption of *pfap2-mrp* on genes associated with heterochromatin

401   regions, we explored the effect of *pfap2-mrp* deletion on chromatin structure. We

402   therefore performed comparative analysis of the chromatin conformation landscape using

403   whole-genome chromosome conformation capture (Hi-C) followed by deep sequencing

13

404   from either the tagged or the *Δpfap2-mrp* strains at 16 and 40 h.p.i. Correlation analysis

405   of Hi-C replicates indicated a clear reproducibility at 16 and 40 h.p.i. where overall intra-

406   chromosomal and interchromosomal interaction matrices appeared largely unchanged for

407   each strain (Fig. 5a, and Extended Data Fig. 8). However, comparative analysis between

408   the tagged and the *Δpfap2-mrp* lines revealed a slight reduction compared to background

409   in long distance interactions and heterochromatin clusters in *Δpfap2-mrp* (Fig. 5b, and

410   Extended Data Fig. 8) at 16 h.p.i. The effect was detected more globally at 40 h.p.i.

411   Genome-wide mapping of interaction changes revealed a reduction in interaction

412   frequency between telomere ends including internal *var* genes (Fig. 5c), consistent with a

413   reduced clustering of the heterochromatin. Further support was provided by genome-wide

414   3D modeling on these two strains wherein we see clear separation of telomeric clustering

415   in *Δpfap2-mrp* and overall expansion of the chromatin structure (Fig. 5d). We concluded

416   that chromatin compaction is impaired in *pfap2-mrp* deficient parasites resulting in

417   increased chromatin accessibility and up-regulation of *var* genes and genes involved in

418   sexual differentiation. Such an effect could explain the partial disconnect between RNA-

419   seq and ChIP-seq data observed in our study at both 16 and 40 h.p.i.

420

421   **DISCUSSION**

422   We provide multiple lines of evidence that PfAP2-MRP is a master regulator that

423   controls key processes of malaria development and pathogenesis during the IDC.

424   Functional deletion of *PfAP2-MRP* at time points corresponding to its two peaks of

425   expression enabled us to show that each peak has separate essential roles. The first peak

426   at 16 h.p.i. is critical for parasite development beyond the trophozoite stage while the

427   second peak is indispensable for merozoite development and egress (Extended Data Fig.

428   9). Through the use of RNA-seq, ChIP-seq and Hi-C, we show that PfAP2-MRP

429   represses *vars* at both 40 h.p.i. and 16 h.p.i. by either interacting with their promoter or

430   by regulating the chromosome accessibility that might contribute to the maintenance of

431   chromatin compaction. scRNA-seq experiments further revealed the activation of most

432   silenced *var* genes at 16 h.p.i. in *ΔPfAP2-MRP* parasites at a single nucleus level. The

433   FACS-based experiment with pooled immune sera from a malaria-endemic region

434   indicated greater immune recognition of the translated transcripts of multiple *var* genes as

14

435   PfEMP1 proteins on the surface of iRBCs in the Δ*PfAP2-MRP* parasites. Taken together,

436   these results suggest that PfAP2-MRP plays a crucial role in maintaining the mutually

437   exclusive expression pattern of the *var* gene family by silencing most *var* genes.

438   Regulation of *var* gene expression involves many chromatin-associated proteins[34], and

439   many of them were found to be associated with PfAP2-MRP, suggesting that it is a

440   recruiter of these proteins (Supplementary discussion).

441

442   Disruption of PfAP2-MRP expression at 40 h.p.i. was sufficient to down-regulate genes

443   associated with merozoite development, egress and invasion processes. Using mass

444   spectrometry, we and other have demonstrated that PfAP2-MRP forms complexes with

445   other Api-AP2s and chromatin-associated proteins, indicating that PfAP2-MRP-regulated

446   genes may be under the combinatorial control of multiple regulatory factors

447   (Supplementary Discussion). We showed that PfAP2-MRP and PfAP2-I[30] bind to the

448   same promoter regions of many invasion and egress-associated genes. Our combined

449   results from RNA-seq and ChIP-seq experiments as well as IP-mass spectrometry

450   indicate that a positive feedback loop-based transcriptional regulatory network exists

451   between these two transcription factors. We observed over-expression of many early

452   gametocyte marker genes in *Δpfap2-mrp* parasites in both 16 and 40 h.p.i. stages. This

453   up-regulation of sexual genes may be explained by the decrease in heterochromatin

454   clustering as well as a direct interaction with critical AP2 transcription factors. Overall,

455   our result indicates that as an essential positive regulator of asexual growth, PfAP2-MRP

456   has evolved to prevent sexual commitment by repressing early gametocyte marker genes

457   critical for sexual conversion (Supplementary discussion).

458

459   In addition to enrichment of egress and invasion-associated genes, we observed many

460   genes encoding hypothetical proteins of unknown function in the list of most down-

461   regulated genes at 40 h.p.i. (n=50). Based on the principle of 'guilt-by-association'[40], we

462   suggest that most of the top down-regulated genes encode hypothetical proteins that are

463   previously unidentified components of the egress and invasion pathways. For example,

464   we note that four of these genes encode hypothetical proteins (Pf3D7_1014100,

465   Pf3D7_0210600, PF3D7_0308300, PF3D7_0507400) that have recently been shown to

466 be important in invasion and egress pathways[41-44] are also down regulated in our study.

467 Future studies will be required to identify downstream regulator(s) of many other

468 invasion and egress-associated genes. It is most likely that the functional characterization

469 of PfAP2-MRP-associated proteins will elucidate their specific roles in chromatin

470 remodeling, histone modification, and transcription.

471

472 In conclusion, this study establishes that several important processes in malaria

473 development and pathogenicity are controlled by a master regulator, PfAP2-MRP. This

474 paves the way for further understanding of the regulatory mechanisms controlling gene

475 expression. It may also facilitate novel therapeutic strategies, such as the use of *Δpfap2-*

476 *mrp* parasites as a live anti-disease vaccine that expresses most of the PfEMP1 repertoire

477 to elicit antibody responses that will decrease severe malaria-associated disease.

478

479 **METHODS**

480 **Parasite culture, maintenance, synchronization, and transfection**

481 The DiCre-expressing *P. falciparum* clone II3[13] was maintained in human A+

482 erythrocytes at 37 °C in RPMI 1640 medium containing AlbumaxII (Invitrogen)

483 supplemented with 2 mM L-glutamine. Parasites were either synchronized by sorbitol

484 treatment or by purifying mature schizont stages using 70% (v/v) isotonic Percoll (GE

485 Healthcare Life Sciences) before allowing reinvasion to occur, followed by sorbitol

486 treatment. For transfection of plasmid constructs, purified Percoll-enriched mature

487 schizonts (~20 µl packed volume) were suspended in 100 µl of P3 primary cell solution

488 (Lonza) containing 60 µg of linearized repair plasmid DNA (repair plasmid 1 and 2

489 separately in two separate transfection experiments) and 20 µg of pDC2 plasmid with the

490 required cloned guide DNA and electroporated using an Amaxa 4D-Nucleofactor

491 (Lonza), using program FP158 as previously described[45]. Drug selection was applied

492 ~20 h after transfection with 5 nM WR99210 (a kind gift of Jacobus Pharmaceuticals) for

493 four days. Once sustained growth of drug-resistant transgenic parasites was observed, the

494 cultures were treated with 1 µM 5-fluorocytosine (5-FC) provided as clinical Ancotil

495 (MEDA) for four days. Transgenic parasite clones PfAP2-MRP:loxP and PfAP2-MRP-

496 3HA:loxP were obtained by limiting dilution cloning in microplates. DiCre recombinase-

16

497   mediated excision was induced by adding rapamycin to the culture at a final

498   concentration of 10 nM.

499

500   **Strategy to perform conditional knockout of PfAP2-MRP**

501   To excise exon 2 of *pfap2-mrp,* which contains the DNA-binding AP2 domain and

502   nuclear localization signal, the endogenous exon 2 of *pfap2-mrp* in the II3 DiCre-

503   expressing *P. falciparum* clone was replaced with transgenic, "floxed" and HA-tagged

504   and non-tagged forms of the exon 2 using two sequential Cas9-mediated genome editing

505   procedures. In the first step, the single intron of *pfap2-mrp* was replaced with *sera2-*

506   loxPint[46]. The repair plasmid (repair plasmid 1) was synthesized commercially (GeneArt;

507   Life Technologies) with recodonised sequences in the 2986–3024 bp and 3273–3305 bp

508   regions of the *pfap2-mrp* and 400 bp homology arms flanking the 5' and 3' regions of the

509   intron. In the second step, cloned parasites with the integrated *sera2-loxPint* were used to

510   introduce a further loxP immediately after the stop codon of the *pfap2-mrp* gene. The

511   repair plasmid for this (repair plasmid 2) contained recodonised sequences in the gene's

512   6261–6330 bp region, a *XmaI* restriction enzyme site followed by a loxP sequence just

513   after the TAA stop codon, and 400 bp homology regions.

514

515   A triple HA-tagged version of the PfAP2-MRP was also prepared. For this, the stop

516   codon was removed from the donor sequence (repair plasmid 2) by PCR amplification

517   using primer pairs (Oligos P1and P2) and digested with *NotI* and *XmaI* restriction

518   enzymes. To remove the sequence with the stop codon from the plasmid backbone, repair

519   plasmid 2 was also digested with *Not1* and *XmaI* restriction enzymes. Digested, amplified

520   sequence without stop codon and digested plasmid backbone were ligated to get the

521   repair plasmid 2 without stop codon after the PfAP2-MRP coding sequence. A triple HA

522   tag sequence was PCR amplified from pFCSS plasmid with *XmaI* sites on both the sides

523   of the HA tag sequence using oligos P5 and P6. *XmaI* digested repair plasmid 2 without

524   stop codon was ligated with *XmaI* digested triple HA tag fragment to generate a triple

525   HA-tagged encoding repair plasmid 2.

526

527   pDC2-Cas9-U6-hdhfr plasmid expressing spCas9 was used in this study. Guide RNA

528   sequences were identified using Benchling. The TATTTATATTCTCAATTGAA and

529   TTATATTCTCAATTGAATGG sequences targeting the 3' end of the first exon,

530   upstream of the TGG protospacer adjacent motif and 5' end of the second exon upstream

531   of the TGG protospacer, respectively, were cloned into the pDC2-Cas9-U6-hdhfr (pDC2

532   plasmid 1) and used to target Cas9 in the first Cas9-mediated editing step (**Fig. 1a**). The

533   CCCTTCAATAGATTCGCACA sequence towards the 3' end of the second exon,

534   upstream of the CGG protospacer adjacent motif, was cloned into the pDC2-Cas9-U6-

535   hdhfr (pDC2 plasmid 1) and used to target Cas9 in the second Cas9-mediated editing

536   step. Sequences of all the oligos and primers used are listed in Supplementary Data 6.

537

538   **Flow cytometry-based quantification of parasitemia**

539   Parasites were fixed in 0.1% glutaraldehyde, incubated for 30 minutes at 37 °C, then

540   stored at 4 °C until further use. Glutaraldehyde fixed parasites were stained with SYBR

541   green (diluted 1: 10,000) for 30 min at 37 °C, then parasitemia was determined using an

542   BD LSR Fortessa flow cytometer (BD Biosciences, San Jose, CA, USA).

543

544   **Invasion and growth assays**

545   For invasion assay, the parasites were treated with DMSO or rapamycin at ~16 h.p.i. in

546   cycle 0 and mature schizonts in the same cycle were isolated using Percoll as described

547   earlier and added to the culture at 5% parasitemia. Merozoites were allowed to invade for

548   4 h under either static or vigorous shaking (250 rpm) conditions. Four hours after adding

549   mature schizonts to the culture, the parasitemia was measured by flow cytometry. Three

550   biological replicates per condition were used.

551

552   **Electron Microscopy**

553   Compound 2 treated and RAPA treated (parasites with disrupted second peak of *pfap2-*

554   *mrp* expression) highly synchronized parasites were allowed to grow until they reached

555   the segmented schizont stage. Then the infected RBCs were fixed with 2.5%

556   glutaraldehyde in cacodylate buffer (0.1 M, pH 7.4) for 48 h. A first osmication was

557   performed using reduced osmium (1:1 mixture of 2% osmium tetroxide and 3%

18

558   potassium ferrocyanide) for 1 hour. After a quick wash with distilled water, a second

559   osmication was performed with 2% osmium tetroxide for 30 min. Samples were washed

560   3x 5min in water and placed in 1% Uranyl acetate for 12 hours at 4°C. Before

561   dehydration, samples were washed 3x 15 min in water. After pre-embedding in 1% agar,

562   samples were dehydrated in an ethanol series (40% to 100%) and embedded in epoxy

563   resin. Thin sections (100 to 150 nm thick) were collected on copper grids and contrasted

564   with lead citrate. Imaging was performed using a transmission electron microscope

565   (TEM) operating at 300 kV (Titan Cryo Twin, Thermo Fisher Scientific). Images were

566   recorded on a 4k x 4k CCD camera (Gatan Inc.).

567

568   **Nucleic acid extraction and polymerase chain reaction**

569   For DNA isolation, cells were pelleted and treated with 0.15% saponin in PBS for 10 min

570   on ice, then washed twice with PBS. DNA was extracted from parasite pellets using

571   DNeasy blood and tissue kit (Qiagen) following the manufacturer's instructions. For

572   diagnostic PCR to check clones, GoTaq (Promega) DNA master mix was used; for

573   amplification of fragments (3xHA tag) used in construct design, Phusion high fidelity

574   DNA polymerase (NEB) was used, and for amplification of fragments longer than > 3 kb

575   Platinum taq HiFi polymerase (Invitrogen) was used.

576

577   **PAGE, immunoblotting and immunofluorescence**

578   Ring-stage parasites were treated with DMSO or rapamycin, and subsequently, mature

579   schizonts (>45 h.p.i) were purified using 70% Percoll, then treated with 0.15% saponin in

580   PBS and washed twice with PBS. Schizonts pellets were lysed by adding sample lysis

581   buffer (1% NP40, 0.1% SDS, 150 mM NaCl), and 5 μg protein of each sample were

582   separated under reducing conditions on Bis-Tris NuPAGE polyacrylamide gels and

583   transferred to nitrocellulose membranes by electroblotting. Blots were blocked overnight

584   in 5% milk powder (w/v) in phosphate-buffered saline (PBS) containing 0.2% Tween-20.

585   To detect 3xHA tagged PfAP2-MRP, the rat anti-HA mAb 3F10 (Sigma) was used at a

586   1:1000 dilution, followed by horseradish conjugated secondary antibody (1:2500). For

587   the proteins tested, relevant primary antibodies were used (see below), then secondary

588   HRP conjugated antibodies specific for mouse, rabbit, or rat IgG (Biorad) were used at a

589   dilution of 1:2500. The signal was developed using Immobilon Western

590   Chemoluminescent HRP Substrate (Merck Millipore) and detected using Hyperfilm ECL

591   film (GE Healthcare).

592

593   For immunofluorescence, thin parasite films were prepared and fixed in 4%

594   paraformaldehyde in PBS for 20 min at room temperature. The cells were permeabilized

595   with 0.1% (v/v) Triton X100 in PBS for 5 min, then blocked with 3% bovine serum

596   albumin (BSA) in PBS overnight at 4 °C before being probed with relevant primary

597   antibodies. Secondary Alexa Fluor 488- or 594- conjugated antibodies specific for

598   mouse, rabbit, or rat IgG (Invitrogen) were used at a 1:5000 dilution. Slides were

599   examined using a Nikon Ni microscope with a 100x Plan Apo NA 1.45 objective; images

600   were captured with an Orca Flash 4 digital camera and prepared with Nikon NIS

601   Elements and Adobe Photoshop software.

602

603   **Antibodies**

604   The following antibodies and dilutions were used for Western blots in these studies.

605   Rabbit anti-EBA175 (1:10,000), rat anti-MyoA (1:1,000), rat anti-BiP (1:1,000), rabbit

606   anti-PTRAMP (1:4,000), rabbit anti-ARO (1:1,000), rabbit anti-AMA1 (1:10,000), and

607   rabbit anti-SUB1 (1:1,000). Antibodies used in immunofluorescence were rabbit anti-

608   GAP45 (1:1,000) and rabbit anti-MSP7 (1:1,000). Anti-EBA175 was obtained from MR4

609   (beiresources.org/MR4Home), anti-BiP was provided by Dr. Ellen Knuepfer, anti-SUB1

610   was a generous gift from Prof. Mike Blackman (Francis Crick Institute), and anti-AMA1

611   was a generous gift from Bart Faber and Clemens Kocken from the Primate Research

612   centre in Rijswijk. All other antibodies were generated in the Holder laboratory and are

613   now held and freely available at NIBSC-CFAR (please contact cfar@nibsc.org with any

614   inquiries).

615

616   **RNA extraction and strand-specific RNA-seq library preparation**

617   Parasite cultures (~0.5-2 ml depending on the asexual developmental stage) were pelleted

618   at 2400 rpm for 3 min, lysed by adding 1ml of Trizol (Sigma), then immediately stored at

619   -80 °C until further use. Total RNA was isolated from Trizol lysed parasites following

620 the manufacturer's instructions (Life Technologies). Strand-specific mRNA libraries were

621 prepared from total RNA using TruSeq Stranded mRNA Sample Prep Kit LS (Illumina)

622 according to the manufacturer's instructions. Briefly, for each sample, 100-300 ng of

623 total RNA was used to prepare the libraries. PolyA+ mRNA was captured from total

624 RNA using oligo-T attached to magnetic beads. First-strand synthesis was performed

625 using random primers followed by second-strand synthesis where dUTP was incorporated

626 in place of dTTP to achieve strand-specificity. Double-stranded cDNA ends were ligated

627 with adaptors, and the libraries were amplified by PCR for 15 cycles before sequencing

628 the libraries on Illumina HiSeq-4000 platform with paired-end 150 bp read chemistry

629 according to manufacturer's instructions (Illumina).

630

631 **RNA-seq data processing and analysis**

632 The quality of the raw reads was assessed using FASTQC

633 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Low-quality reads and

634 Illumina adaptor sequences from the read ends were removed using TrimmomaticR[47].

635 Processed reads were mapped to the *P. falciparum* 3D7 reference genome (release 40 in

636 PlasmoDB- http://www.plasmoddb.org) using Hisat2[48] (V 2.1.0) with parameter "—rna-

637 strandness FR". Counts per feature were estimated using FeatureCounts[49]. Raw read

638 counts data were converted to counts per million (cpm), and genes were excluded if they

639 failed to achieve a cpm value of 1 in at least one of the three replicates performed.

640 Library sizes were scale-normalized by the TMM method using EdgeR software[50] and

641 further subjected to linear model analysis using the voom function in the limma

642 package[51]. Differential expression analysis was performed using DeSeq2[52]. Genes with a

643 false discovery rate corrected *P* value (Benjamini-Hochberg procedure) < 0.05 and $\log_2$

644 fold change ≥ 1 or ≤ -1 were considered as up-regulated or down regulated, respectively.

645

646 **Real-time quantitative PCR**

647 Total RNA was treated with DNAase (TURBO DNase, Cat No. AM2238, Invitrogen)

648 following the manufacturer's instructions. The removal of DNA was confirmed by

649 performing PCR using housekeeping genes. cDNA (1 μg) was prepared from total RNA

650 using LunaScript cDNA synthesis mix (NEB), diluted 5 times before using it for qRT-

21

651 PCR. mRNA expression levels were estimated on a Quant Studio 3 real-time qPCR

652 machine (Applied Biosystems) using Fast SYBR green master mix (Applied Biosystems,

653 Cat. No. 4385612). Seryl-tRNA ligase (PF3D7_0717700) was used as the internal

654 control to normalize mRNA levels. Specific amplification of the PCR product was

655 verified by dissociation curve analysis and relative quantities of mRNA calculated using

656 the ΔΔCt Method[53]. PCR primers used in the real-time qRT-PCR experiment are listed in

657 **Supplementary Data 6.** For the *var* gene qRT-PCR experiment, primer sets targeting

658 individual *var* genes described elsewhere[54] were used, including the primers targeting the

659 seryl tRNA ligase gene (PF3D7_0717700) used as the internal control.

660

661 **PfAP2-MRP chromatin immunoprecipitation (AP2-MRP ChIP)**

662 The ChIP assay was performed as described [55] with a few modifications. Parasite culture

663 (50 ml) containing synchronized late-stage schizonts (~ 40 h.p.i.) at ~5% parasitemia was

664 centrifuged at 900 g for 4 min, and the cells were washed once with PBS. To the cell

665 pellet, 25 ml of 0.15% saponin in PBS was added and incubated on ice for 10 min,

666 followed by washing twice with cold PBS. Parasites were crosslinked for 10 min by

667 adding methanol-free formaldehyde at 1% final concentration and incubated for 10 min at

668 37 °C with occasional shaking. The crosslinking reaction was quenched by adding 1.25

669 M glycine to achieve a final concentration of 0.125 M and incubated at 37 °C for another

670 5 min. Parasites were centrifuged for 10 min at 3250 g at 4 °C, washed three times with

671 DPBS, snap-frozen in liquid nitrogen, and stored at -80 °C until further use. Frozen

672 formaldehyde-fixed parasites were thawed on ice for chromatin immunoprecipitation.

673 One ml of nuclear extraction buffer (10 mM HEPES, 10 mM KCl, 0.1 mM EDTA, 0.1

674 mM EGTA, 1mM DTT, 1x EDTA-free protease inhibitor cocktail [Roche]) was added to

675 the tubes containing thawed parasites and incubated on ice for 30 min. After the

676 incubation, 10% NP-40 was added to reach a final concentration of 0.25%, and the

677 parasites were lysed by passing through a $26^{1/2}$ G needle seven times. Parasite nuclei were

678 collected by centrifuging at 2500*g* for 10 min at 4 °C. Shearing of chromatin was carried

679 out using the Covaris Ultra Sonicator (E220) for 14 min with the following settings; 5%

680 duty cycle, 140 intensity peak incident power, 200 cycles per burst to obtain fragment

681 size of 200 to 600 bp. Insoluble materials were removed by centrifuging the sheared

682   chromatin for 10 min at 13500 g at 4 °C. 30 µl of fragmented chromatin were stored as

683   input at -80°C.

684

685   Fragmented chromatin was diluted 1:1 in ChIP dilution buffer (30 mM Tris-HCl pH 8.0,

686   0.1% SDS, 3 mM EDTA, 300 mM NaCl, 1% Triton X-100, EDTA-free protease inhibitor

687   cocktail). Chromatin was incubated overnight with 6 µg of rabbit polyclonal anti-HA

688   (Abcam no. ab9110) or, as control, the same amount of rabbit IgG isotype control (Cat.

689   No. 10500c, Invitrogen). The antibody-protein complex was recovered with protein A

690   coupled to magnetic beads (Dynabeads, Invitrogen, Cat No. 10002D), followed by

691   extensive washes with low salt immune complex wash buffer, high salt immune complex

692   wash buffer (washes done at 4 °C), and TE buffer (washes done at RT). Chromatin was

693   eluted with elution buffer (1% SDS, 0.1 M NaHCO$_3$) at 45 °C for 30 min with shaking.

694   Immunoprecipitated chromatin and input were reverse crosslinked overnight at 45 °C by

695   adding 5 M NaCl to a final concentration of 0.5 M. Samples were treated with RNaseA

696   for 30 min at 37 °C followed by a 2 h incubation at 45 °C with proteinase K (final

697   concentration 66 µg/ml). DNA was purified using ChIP DNA clean & concentrator

698   (Zymo Research, Cat. No. D5205).

699

700   **PfAP2-MRP ChIP-sequencing and analysis**

701   Libraries were prepared using NEBNext Ultra II DNA library kit following the

702   manufacturer's instructions until the step of adapter ligation (Adapters were diluted at 1:

703   20 ratio). Adapter ligated libraries were purified using AmpureXP beads. The libraries

704   were amplified for a total of 6 PCR cycles (2 min at 98 °C initial denaturation; 6 cycles

705   of 30 s at 98 °C, 50 s at 62 °C, final extension 5 min at 62 °C) using the KAPA HiFi

706   HotStart Ready Mix (Kapa Biosystems). Amplified libraries were purified, and size

707   selected for 350 bp inserts using AmpureXP beads and sequenced on the Illumina

708   HiSeqX platform with 150 bp paired-end read layouts. Low-quality reads and Illumina

709   adaptor sequences from the read ends were removed using Trimmomatic[47].   Quality

710   trimmed reads were aligned to the *P. falciparum* genome (plasmodb.org, v3, release v32)

711   using HiSat2. Duplicate reads were removed using samtools (markdup) [56]. GC bias was

712   corrected using deeptool's correctGCBias tool[57]. For coverage plots of AP2-MRP 40

713   h.p.i. and 16 h.p.i. ChIP-seq experiments, deeptool's bamCompare tool was used to

714   normalize the read coverage per base of the genome position (option '-bs 1') in the

715   respective input and ChIP samples or IgG and ChIP samples to the total number of reads

716   in each library (-- normalizeUsing RPKM). Normalized input coverage or IgG coverage

717   per bin was subtracted from the ChIP values (option – operation subtract). Coverage plots

718   were visualized using IGV genome browser[58].

719

720   ChIP-Peaks (q-value cutoff < 0.05) were identified using macs2[59] by comparing the input

721   with ChIP or IgG with ChIP with default settings but without prior peak modeling (option

722   '-nomodel'), the fragment size set to 200 bp (option '-extsize 200') and the genome size

723   (option '-g') set to 233332839. Robust common peaks between replicates were identified

724   using bedtools 'intersect' (option –f 0.30 –r)[60]. Peak annotation was carried out using

725   Homer's annotatePeaks.pl that assigned each peak with the nearest downstream gene.

726   After intersecting, common peaks with peak score > 50 were kept for further analysis.

727   Enrichment heatmaps and profile plots were generated using the deepTools

728   computeMatrix and plotHeatmap tools.

729

730   **Processing of published PfAP2-I and PfAP2-G ChIP-seq data**

731   PfAP2-I and PfAP2-G ChIP-seq published raw data[30,31] were downloaded from ENA

732   and processed exactly as the ChIP-seq data for PfAP2-MRP. ChIP-Peaks (q-value cutoff

733   < 0.05) were identified using macs2[59] by comparing the input with ChIP for both the

734   replicates of PfAP2-G and PfAP2-I. Robust common peaks between PfAP2-MRP and

735   PfAP2-I; PfAP2-MRP and PfAP2-MRP and between all three were identified using

736   bedtools 'intersect' (option –f 0.30 –r) to find peaks that overlapped at least 30%.

737

738   **Parasite sample preparation for single-cell RNA-seq**

739   For 40 h.p.i. time-point, tightly synchronous parasites were enriched using 63% Percol,

740   washed twice with incomplete RPMI medium, and processed immediately on the 10X

741   Chromium controller (10X Genomics, Pleasanton, CA). For 16 h.p.i. time point parasites

742   were stained with Mitotracker Deep Red FM (Life Technologies, #M22426) for FACS

743   analysis and flow sorting, respectively. Briefly, 50 µl of SYBR Green I stained RBCs were

24

744   analyzed on BD LSR Fortessa Flow Cytometer with High Throughput sampler (BD

745   Biosciences, San Jose, CA, USA) using BD FACS Diva Software v6.2 and 488 laser

746   excitation / 530 emission filter to determine the concentration of SYBR Green I positive

747   cells per microliter. A BD Influx Cell Sorter (BD Biosciences, San Jose, CA, USA) with

748   BD FACS Sortware v1.0.01 software was used to sort ~40,000 MitoTracker Deep Red FM

749   - positive RBC's using a 70 µm nozzle, a 640 nm laser excitation / 670 nm emission filter

750   and a pressure setting of 30 psi. Post-sorted cell concentration and quality were checked

751   using a Countess® II Automated Cell Counter (Invitrogen) and FLoid™ Cell Imaging

752   Station (ThermoFisher). Finally, labeled cells (i.e., SYBR Green I or MitoTracker Deep

753   Red FM positive cells) were then loaded onto a 10X chip (Chip G) and processed

754   immediately on the 10X Chromium controller (10X Genomics, Pleasanton, CA).

755

756   **Single-cell RNA-seq library preparation**

757   Single-cell libraries were constructed using the 10X Genomics Chromium Next GEM

758   Single Cell 3′ Reagent Kits v3.1 with Single Index Kit T Set A. Due to the extremely low

759   RNA content of single-cell malaria parasite and the AT-rich genome (~70% AT),

760   modifications to the cDNA amplification and library preparation workflow were made

761   accordingly. These modifications included - 30x cDNA amplification cycles; taking 50%

762   cDNA as input into library generation; reducing fragmentation time to 2 minutes; and

763   changing the extension time to 65 ºC during index PCR. Individual library Qc was

764   performed using the BioAnalyzer HS DNA Assay kit (Agilent).

765

766   **Sequencing of single-cell RNA-seq library**

767   Library concentration was determined with the KAPA Library Quantification Kit

768   (ROCHE) using the QuantStudio3 Real-Time PCR systems (ThermoFisher) and assessed

769   for fragment size using the BioAnalyzer HS DNA Assay kit (Agilent). Following library

770   pooling in equimolar concentrations, a total of 1.2 nM library was sequenced on the

771   Illumina NovaSeq 6000 with SP flow cell using version 1.5 chemistry as follows: Read 1

772   – 28bp; Index read i7 – 8bp; Index read i5 – 0bp; and Read 2 – 91bp.

773

774   **Single-Cell Transcriptome alignment and read count estimation**

25

775 The Droplet-based sequencing reads were aligned to the Hybrid genome of Human hg38

776 and *P. falciparum* pd37 (PlasmoDB-46_Pfalciparum3D7_Genome.fasta) to remove any

777 human transcript contamination. This was achieved using CellRanger standard pipeline

778 using --nosecondary flag. The raw gene count matrix was subjected to various single-cell

779 preprocessing steps separately.

780

781 **Preprocessing and Normalization**

782 Primarily, the reads mapping to human genes were removed, followed by the

783 identification and removal of empty droplets using the emptyDrops() function from R

784 package dropletUtils v1.12.3[61]. This function determines whether the RNA content of a

785 cell barcode differs considerably from the ambient background RNA present in each

786 sample. Cells with FDR <= 0.001 (Benjamini–Hochberg-corrected) were examined for

787 subsequent investigation. The per-cell quality metrics were computed by the

788 addPerCellQC function of the scuttle package v1.2.1[62]. The deconvolution approach in

789 the computeSumFactors function of the Scran R package v1.20.1[63] was utilized to

790 normalize cell-specific biases. We kept the mitochondrial genes of *P falciparum* since the

791 proportion of UMIs allocated to mitochondrial genes in both controls and knockouts were

792 similar.

793

794 Further, the doublets identification was performed using the computeDoubletDensity

795 function of the scDblFinder Bioconductor package v1.6.0 **[Germain P (2021).**

796 **scDblFinder: scDblFinder. R package version 1.6.0,**

797 **https://github.com/plger/scDblFinder.]**. This was achieved in three steps. a) The log

798 normalization of counts was achieved using the logNormCounts function of scuttle

799 package. b) The modelGeneVarByPoisson function of scran[63] was then used to model the

800 per-gene variance followed by c) doublet score calculation using the top 10% of highly

801 variable genes (HVGs). We then cleaned our data based on a 95% quantile cut-off.

802 Additionally, the standard functions of the Seurat package[64] were also used to generate

803 intuitive QC plots.

804 **Cell type and infection stage recognition:**

805 The transcriptomic data at single-cell resolution about the malaria life cycle was obtained

806 from the Malaria Cell Atlas. The phenotypic data for MCA was obtained from

807 https://github.com/vhowick/MalariaCellAtlas/blob/master/Expression_Matrices/10X/pf1

808 0xIDC/pf10xIDC_pheno.csv[65]. We used the SingleR package v1.6.1 to transfer labels

809 from the reference atlas to each cell in our data. It identifies marker genes for each stage

810 in the reference atlas and uses them to compute assignment scores (based on the

811 Spearman correlation across markers) for each cell in the test dataset against each label in

812 the reference. The top 20 marker genes were identified using the Wilcoxon ranked sum

813 test[66]

814 The time-series data of the intra-erythrocytic development cycle (IDC) was obtained

815 from[14]. To transfer the time-series labels from Bulk RNASeq, we used a subset of

816 samples for both 16h.p.i and 40h.p.i to prevent label misassignments. Each biological

817 replicate was considered a single cell and markers were identified using the "classic"

818 method of SingleR.

819 The correlation between cells of different stages was carried out using the

820 CorrelateReference() function of the CHETAH R package v1.8.0[67]. Similarly, we

821 performed an inter-sample correlation between the transcriptional profiles of samples

822 from different time points.

823 **Integration of in-house data with MCA**

824 The cleaned data was then integrated with Malaria Cell Atlas Single Cell data using

825 standard scRNAseq integration workflow as described elsewhere[64]. Briefly, we create an

826 "integrated" data assay but first identify the pairwise anchors using

827 FindIntegrationAnchors followed by IntegrateData that exploit this anchor-set to combine

828 the MCA and In-house data. Next, we ran the standard workflow on integrated assay

829 including scaling, dimension reduction and clustering using parameters described in

830 https://satijalab.org/seurat/articles/integration_introduction.html#integration-goals-1. The

831 clusters were then manually annotated using the RenameIdents function of Seurat [64].

832

833 *Var* gene expression per cell

834   A set of 61 *var* genes were used for their average expression calculation. The *var* gene

835   expression was calculated using Raw counts (RNA assay data slot). *var* gene Expression

836   calculation=sum of *var* gene expression/sum of Expression of all genes.

837

838   **PfAP2-MRP motif identification**

839   Sequences from the commonly identified peaks from two replicates were extracted from

840   the *P. falciparum* 3D7 genome using the bedtools 'getfasta'. These retrieved sequences

841   were then uploaded to the DREME server[68] to identify significantly enriched motifs in

842   the peak region. Tomtom[69] was used to compare the de novo identified motifs to

843   previously in silico discovered motifs[70].

844

845   **Immunoprecipitation of PfAP2-MRP and identification of proteins by mass**

846   **spectrometry**

847   To identify the AP2 interacting proteins, the on beads digestion with trypsin approach

848   was used after immunoprecipitation of AP2 complex. Before performing the on beads

849   digestion, after TE buffer washes the immunoprecipitated complex was washed two times

850   with exchange buffer (100 mM NaCl, Tris 50 mM pH 7.5) for 10 minutes at 4°C.

851   Subsequently, the beads were resuspended in 100 µl of 100 mM Triethylammonium

852   bicarbonate (TEAB), and reduction of bound proteins was done with 1mM DTT at 37°C

853   for 30 minutes on thermo mixture with constant shaking at 750 rpm. The sample was

854   brought to room temperature and alkylated in the dark with 3mM iodoacetamide for 45

855   minutes, the excess iodoacetamide was quenched with 3mM DDT for 10 minutes (Sigma

856   Aldrich). Afterward, the sample was digested with 2.5µg trypsin (Promega) overnight on

857   a thermo mixture with constant shaking at 1000 rpm. The resulting digested peptides

858   were purified from beads, and trypsin digestion was stopped by adding TFA to a 2% final

859   concentration. The acidified peptide was desalted using Sep-Pak C18 1 cc Vac Cartridge,

860   50 mg Sorbent per Cartridge (Waters). In brief, the Sep-Pak column was conditioned

861   with 1 ml 100% methanol twice and equilibrated with 1 ml of 0.1% trifluoroacetic acid

862   (TFA) twice. After that, the digested acidified peptides were loaded. The bound peptides

863   were washed with 1 ml of 0.1% TFA twice and eluted with 300 µl of elution buffer

864 (Acetonitrile 75% with 0.1% TFA in water) twice. The eluted peptides were dried in

865 SpeedVac and kept at -80°C till further use.

866

867 **LC-MS analysis of peptides**

868 The LC-MS analysis was performed on Q-Exactive HF mass spectrometer coupled with

869 an UltiMate™ 3000 UHPLC (Thermo Fisher Scientific). The peptides were dissolvedin

870 0.1% formic acid (FA; Sigma Aldrich), and approximately 1µg of peptides was separated

871 on an Acclaim PepMap^{TM} C18 column (75 µm I.D. X 250 mm, 2 µm particle sizes, 100

872 Å pore sizes) with a gradient of 5-35% mobile phase A and B respectively for 55

873 minutes, ramping up to 90% phase B for 5 minutes, 90% phase B for 5 minutes and the

874 column was conditioned to 2% phase B for 10 minutes with the flow rate of 300 nl/min^{-1}

875 (Phase A 0.1% FA, Phase B 99.9 % ACN with 0.1%FA). The peptides were introduced

876 into the mass spectrometer through Nanospray Flex with an electrospray potential of 2.5

877 kV. Data were acquired in the Orbitrap at the resolution of 60,000 in the mass range of

878 350-16000 m/z with a maximum ion accumulation time set to 50 ms. The 20 most intense

879 ions with a threshold of more than $1x\ e^6$ and having the multiple charges were further

880 fragmented by using higher-energy collision dissociation (HCD) at 15000 resolution. The

881 Dynamic exclusion for HCD fragmentation was 30 seconds. The maximum time for

882 fragmented ion accumulation was 30 ms, a target value of $2.50xe^3$, the normalized

883 collision energy at 28%, and the isolation width of 1.6. During the acquisition, the ion

884 transfer tube temperature was set at 160°C, data was acquired in data-dependent

885 acquisition (DDA) mode, and the total run time was 75 minutes.

886

887 **Identification, quantification, and statistical analysis of LC-MS data**

888 Raw LC-MS data files from Q-Exactive HF were converted to .mgf files using Proteo

889 Wizard MS covertgui 64 bit and analyzed using Mascot v2.4. The annotated protein

890 sequence for *Plasmodium falciparum* was downloaded from https://plasmodb.org

891 (Release 51). Trypsin was set as the enzyme of choice with maximum missed cleavage 1,

892 fixed modification carbamidomethyl (K), variable modification deamidated and oxidation

893 of N, Q, and M, respectively, with the peptide and fragment mass tolerance at 0.6 Da.

894

29

ChIP qPCR

After de-crosslinking, ChIP DNA was purified using the Zymo ChIP DNA kit and quantified by Qubit HS DNA assay. The purified ChIP DNA was first diluted 20-fold in elution buffer and then analyzed by qPCR using the CFX-96 Biorad system. All ChIP primers used (Supplementary Data 6) were first checked using genomic DNA to determine specificity (based on a single peak in the melting curve) and efficiency. ChIP qPCR data were analyzed using the ΔΔCt method. *Pfap2-mrp* ChIP-qPCR results are expressed as a percentage of input. Three biological replicates of both samples and negative control (mock IP using IgG) were used for ChIP-qPCR experiments.

Flow cytometry using pooled human serum.

iRBCs with trophozoite stage parasites from cycle 1, treated with DMSO or rapamycin in cyle 0 to disrupt the first peak of *PfAP2-MRP* expression, were washed thrice with PBS supplemented with 0.1% BSA. iRBCS were either untreated or treated with Hiserum. When untreated, the same volume of 0.1% BSA in PBS was added and incubated for 30 mins at room temperature. Cells were washed thrice with 0.1% BSA in PBS, and all the samples were treated with sybr green (1x) and mouse anti-human IgG conjugated with Alexflour 647 (1:100 dilution, from BioLegend) for 30 mins at room temperature. After incubation, samples were washed thrice again with 0.1% BSA in PBS and analyzed on an BD LSR Fortessa flow cytometer (BD). Data were analyzed using FlowJo v9 software.

In situ Hi-C

Parasite were crosslinked using 1.25% formaldehyde in warm 1X PBS for 25 min at 37 °C with rotation. Glycine was then added to a final concentration of 150 mM to quench the formaldehyde and incubated for 15 min at 37 °C and 15 min at 4 °C, both with rotation. Following centrifugation at 660 x g for 15 min at 4 °C, the pellet was resuspended in 5 volumes of ice-cold 1X PBS and incubated for 10 min at 4°C with rotation. After another centrifugation at 660 x g for 15 min at 4 °C the pellet was resuspended in 20 ml ice-cold 1X PBS. Several more washes in cold 1X PBS were used to clear cellular debris before resuspending in 1 ml 1X PBS and separated into multiple 1.5 ml tubes at a concentration of 1×10^8 parasites per tube. The tubes were flash frozen

926 in liquid nitrogen and stored at 80 °C before continuing with the rest of the in-situ Hi-C

927 protocol[71] using MboI restriction enzyme, with modifications to the standard protocol[72].

928 The final Hi-C libraries were sequenced using the Illumina NovaSeq 6000 using the S4

929 300 cycle flow cell for paired-end read libraries.

930

931 **Hi-C data processing and analysis**

932 Paired-end HiC library reads were processed using HiC-Pro (Servant et al., 2015) with

933 default parameters and mapping at 10kb resolution to the *P. falciparum* genome (release-

934 50, plasmodb.org). The ICED-normalized interaction matrices output by HiC-Pro were

935 interaction counts-per-million normalized before generating interaction heatmaps. All

936 intra-bin contacts and contacts within a two-bin distance were set to 0 to enhance

937 visualization and the color map per chromosome was scaled based on the minimum

938 number of interactions in the highest 10% of interacting bins to aid in comparison

939 between samples. Interaction matrices for the replicates were merged and differential

940 interactions were identified by calculating the $\log_2$ fold change between the wild type and

941 AP2-KO at each time point. Coordinate matrices generated by PASTIS[73] were visualized

942 as 3D chromatin models using ChimeraX[74].

943

944 **Acknowledgements**

966

Data availability

968 The data sets generated in this study are available in the following databases:

969 - RNA-seq data: NCBI BioProject accession # GSE190342

970 - ScRNA-seq: NCBI BioProject accession # GSE191025

971 - ChIP-seq data: NCBI BioProject accession # GSE 190497

972 - Proteomics data: Pride accession number # PXD030308.

973 - Hi-C data: ENA BioProject accession number #PRJNA847684

974 The bulk RNA-seq, ScRNA-seq and ChIP-Seq datasets have been added under the super
975 series GSE190519

976

Author contributions

978 Conceptualization, A.K.S., and A.P.; Methodology, A.K.S., J.L.G., A.A.H., and A.P.;

979 Investigation, A.K.S., J.L.G., M.G., R.P.S., M.S., I.I., T.M., R.N., Z.S., S.M., L.E., R.S.,

980 Y.O., R.S., F.B.R., S.A., A.D., A.F.K., J.M., W.F., E.K., I.G., D.J.P.F.; Analysis, A.K.S.,

981 J.L.G., R.S., S.A., T.L., I.G.; Writing the original draft, A.K.S.; Writing-Review and

982 Editing, A.K.S., J.L.G., K.G.L.R., A.A.H and A.P.; Resources, A.A.H., and A.P.;

983 Funding acquisition, A.A.H., and A.P.; Supervision, A.P. All authors read and approved

984 the manuscript.

985

References

987 1   Cowman, A. F., Tonkin, C. J., Tham, W. H. & Duraisingh, M. T. The Molecular Basis of
988   Erythrocyte Invasion by Malaria Parasites. *Cell Host Microbe* **22**, 232-245,
989   doi:10.1016/j.chom.2017.07.003 (2017).

990 2   Boddey, J. A. & Cowman, A. F. Plasmodium nesting: remaking the erythrocyte from the inside
991   out. *Annu Rev Microbiol* **67**, 243-269, doi:10.1146/annurev-micro-092412-155730 (2013).
992 3   Tan, M. S. Y. & Blackman, M. J. Malaria parasite egress at a glance. *J Cell Sci* **134**,
993   doi:10.1242/jcs.257345 (2021).
994 4   Cortes, A. & Deitsch, K. W. Malaria Epigenetics. *Cold Spring Harb Perspect Med* **7**,
995   doi:10.1101/cshperspect.a025528 (2017).
996 5   Toenhake, C. G. *et al.* Chromatin Accessibility-Based Characterization of the Gene Regulatory
997   Network Underlying Plasmodium falciparum Blood-Stage Development. *Cell Host Microbe* **23**,
998   557-569 e559, doi:10.1016/j.chom.2018.03.007 (2018).
999 6   Iwanaga, S., Kaneko, I., Kato, T. & Yuda, M. Identification of an AP2-family protein that is
1000   critical for malaria liver stage development. *PLoS One* **7**, e47557,
1001   doi:10.1371/journal.pone.0047557 (2012).
1002 7   Kafsack, B. F. *et al.* A transcriptional switch underlies commitment to sexual development in
1003   malaria parasites. *Nature* **507**, 248-252, doi:10.1038/nature12920 (2014).
1004 8   Painter, H. J., Campbell, T. L. & Llinas, M. The Apicomplexan AP2 family: integral factors
1005   regulating Plasmodium development. *Mol Biochem Parasitol* **176**, 1-7,
1006   doi:10.1016/j.molbiopara.2010.11.014 (2011).
1007 9   Tinto-Font, E. *et al.* A heat-shock response regulated by the PfAP2-HS transcription factor
1008   protects human malaria parasites from febrile temperatures. *Nat Microbiol* **6**, 1163-1174,
1009   doi:10.1038/s41564-021-00940-w (2021).
1010 10   Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. & Kaneko, I. Transcription factor AP2-Sp and its
1011   target genes in malarial sporozoites. *Mol Microbiol* **75**, 854-863, doi:10.1111/j.1365-
1012   2958.2009.07005.x (2010).
1013 11   Yuda, M. *et al.* Identification of a transcription factor in the mosquito-invasive stage of malaria
1014   parasites. *Mol Microbiol* **71**, 1402-1414, doi:10.1111/j.1365-2958.2009.06609.x (2009).
1015 12   Collins, C. R. *et al.* Malaria parasite cGMP-dependent protein kinase regulates blood stage
1016   merozoite secretory organelle discharge and egress. *PLoS Pathog* **9**, e1003344,
1017   doi:10.1371/journal.ppat.1003344 (2013).
1018 13   Knuepfer, E., Napiorkowska, M., van Ooij, C. & Holder, A. A. Generating conditional gene
1019   knockouts in Plasmodium - a toolkit to produce stable DiCre recombinase-expressing parasite
1020   lines using CRISPR/Cas9. *Sci Rep* **7**, 3881, doi:10.1038/s41598-017-03984-3 (2017).
1021 14   Subudhi, A. K. *et al.* Malaria parasites regulate intra-erythrocytic development duration via
1022   serpentine receptor 10 to coordinate with host rhythms. *Nat Commun* **11**, 2763,
1023   doi:10.1038/s41467-020-16593-y (2020).
1024 15   Gomes, A. R. *et al.* A genome-scale vector resource enables high-throughput reverse genetic
1025   screening in a malaria parasite. *Cell Host Microbe* **17**, 404-413, doi:10.1016/j.chom.2015.01.014
1026   (2015).
1027 16   Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite Plasmodium
1028   falciparum by saturation mutagenesis. *Science* **360**, doi:10.1126/science.aap7847 (2018).
1029 17   Thomas, J. A. *et al.* A protease cascade regulates release of the human malaria parasite
1030   Plasmodium falciparum from host red blood cells. *Nat Microbiol* **3**, 447-455, doi:10.1038/s41564-
1031   018-0111-0 (2018).
1032 18   Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in Plasmodium falciparum. *Annu
1033   Rev Microbiol* **62**, 445-470, doi:10.1146/annurev.micro.61.080706.093134 (2008).
1034 19   Baker, D. A. *et al.* Cyclic nucleotide signalling in malaria parasites. *Open Biol* **7**,
1035   doi:10.1098/rsob.170213 (2017).
1036 20   Singh, S. & Chitnis, C. E. Molecular Signaling Involved in Entry and Exit of Malaria Parasites
1037   from Host Erythrocytes. *Cold Spring Harb Perspect Med* **7**, doi:10.1101/cshperspect.a026815
1038   (2017).
1039 21   Sargeant, T. J. *et al.* Lineage-specific expansion of proteins exported to erythrocytes in malaria
1040   parasites. *Genome Biol* **7**, R12, doi:10.1186/gb-2006-7-2-r12 (2006).
1041 22   Siddiqui, G., Proellochs, N. I. & Cooke, B. M. Identification of essential exported Plasmodium
1042   falciparum protein kinases in malaria-infected red blood cells. *Br J Haematol* **188**, 774-783,
1043   doi:10.1111/bjh.16219 (2020).
1044 23   Howick, V. M. *et al.* The Malaria Cell Atlas: Single parasite transcriptomes across the complete
1045   Plasmodium life cycle. *Science* **365**, doi:10.1126/science.aaw2619 (2019).

33

1046   24   Taylor, T. E. *et al.* Intravenous immunoglobulin in the treatment of paediatric cerebral malaria.
1047   *Clin Exp Immunol* **90**, 357-362, doi:10.1111/j.1365-2249.1992.tb05851.x (1992).
1048   25   Chan, J. A., Fowkes, F. J. & Beeson, J. G. Surface antigens of Plasmodium falciparum-infected
1049   erythrocytes as immune targets and malaria vaccine candidates. *Cell Mol Life Sci* **71**, 3633-3657,
1050   doi:10.1007/s00018-014-1614-3 (2014).
1051   26   Gulati, S. *et al.* Profiling the Essential Nature of Lipid Metabolism in Asexual Blood and
1052   Gametocyte Stages of Plasmodium falciparum. *Cell Host Microbe* **18**, 371-381,
1053   doi:10.1016/j.chom.2015.08.003 (2015).
1054   27   Josling, G. A. & Llinas, M. Sexual development in Plasmodium parasites: knowing when it's time
1055   to commit. *Nat Rev Microbiol* **13**, 573-587, doi:10.1038/nrmicro3519 (2015).
1056   28   Poran, A. *et al.* Single-cell RNA sequencing reveals a signature of sexual commitment in malaria
1057   parasites. *Nature* **551**, 95-99, doi:10.1038/nature24280 (2017).
1058   29   Shang, X. *et al.* Genome-wide landscape of ApiAP2 transcription factors reveals a
1059   heterochromatin-associated regulatory network during Plasmodium falciparum blood-stage
1060   development. *Nucleic Acids Res*, doi:10.1093/nar/gkac176 (2022).
1061   30   Santos, J. M. *et al.* Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I
1062   Transcription Factor. *Cell Host Microbe* **21**, 731-741 e710, doi:10.1016/j.chom.2017.05.006
1063   (2017).
1064   31   Josling, G. A. *et al.* Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat Commun*
1065   **11**, 1503, doi:10.1038/s41467-020-15026-0 (2020).
1066   32   Volz, J. C. *et al.* PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var
1067   gene in a poised state during parasite division. *Cell Host Microbe* **11**, 7-18,
1068   doi:10.1016/j.chom.2011.11.011 (2012).
1069   33   Volz, J. *et al.* Potential epigenetic regulatory proteins localise to distinct nuclear sub-
1070   compartments in Plasmodium falciparum. *Int J Parasitol* **40**, 109-121,
1071   doi:10.1016/j.ijpara.2009.09.002 (2010).
1072   34   Bryant, J. M. *et al.* Exploring the virulence gene interactome with CRISPR/dCas9 in the human
1073   malaria parasite. *Mol Syst Biol* **16**, e9569, doi:10.15252/msb.20209569 (2020).
1074   35   Farhat, D. C. *et al.* A MORC-driven transcriptional switch controls Toxoplasma developmental
1075   trajectories and sexual commitment. *Nat Microbiol* **5**, 570-583, doi:10.1038/s41564-020-0674-4
1076   (2020).
1077   36   Andrews, K. T., Haque, A. & Jones, M. K. HDAC inhibitors in parasitic diseases. *Immunol Cell*
1078   *Biol* **90**, 66-77, doi:10.1038/icb.2011.97 (2012).
1079   37   Ay, F. *et al.* Three-dimensional modeling of the P. falciparum genome during the erythrocytic
1080   cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*
1081   **24**, 974-988, doi:10.1101/gr.169417.113 (2014).
1082   38   Bunnik, E. M. *et al.* Changes in genome organization of parasite-specific gene families during the
1083   Plasmodium transmission stages. *Nat Commun* **9**, 1910, doi:10.1038/s41467-018-04295-5 (2018).
1084   39   Bunnik, E. M. *et al.* Comparative 3D genome organization in apicomplexan parasites. *Proc Natl*
1085   *Acad Sci U S A* **116**, 3183-3192, doi:10.1073/pnas.1810815116 (2019).
1086   40   Oliver, S. Guilt-by-association goes global. *Nature* **403**, 601-603, doi:10.1038/35001165 (2000).
1087   41   Nagaoka, H. *et al.* PfMSA180 is a novel Plasmodium falciparum vaccine antigen that interacts
1088   with human erythrocyte integrin associated protein (CD47). *Sci Rep* **9**, 5923, doi:10.1038/s41598-
1089   019-42366-9 (2019).
1090   42   Liffner, B. *et al.* PfCERLI1 is a conserved rhoptry associated protein essential for Plasmodium
1091   falciparum merozoite invasion of erythrocytes. *Nat Commun* **11**, 1411, doi:10.1038/s41467-020-
1092   15127-w (2020).
1093   43   Wichers, J. S. *et al.* Identification of novel inner membrane complex and apical annuli proteins of
1094   the malaria parasite Plasmodium falciparum. *Cell Microbiol* **23**, e13341, doi:10.1111/cmi.13341
1095   (2021).
1096   44   Tarr, S. J. *et al.* A malaria parasite subtilisin propeptide-like protein is a potent inhibitor of the
1097   egress protease SUB1. *Biochem J* **477**, 525-540, doi:10.1042/BCJ20190918 (2020).
1098   45   Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi
1099   to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A* **110**, 531-536,
1100   doi:10.1073/pnas.1216457110 (2013).

1101 46   Jones, M. L. *et al.* A versatile strategy for rapid conditional genome engineering using loxP sites
1102   in a small synthetic intron in Plasmodium falciparum. *Sci Rep* **6**, 21800, doi:10.1038/srep21800
1103   (2016).
1104 47   Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
1105   data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
1106 48   Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory
1107   requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
1108 49   Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1109   assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
1110   doi:10.1093/bioinformatics/btt656 (2014).
1111 50   McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-
1112   Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297,
1113   doi:10.1093/nar/gks042 (2012).
1114 51   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
1115   microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
1116 52   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
1117   RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
1118 53   Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic
1119   Acids Res* **29**, e45, doi:10.1093/nar/29.9.e45 (2001).
1120 54   Jiang, L. *et al.* PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium
1121   falciparum. *Nature* **499**, 223-227, doi:10.1038/nature12361 (2013).
1122 55   Zeeshan, M. *et al.* Real-time dynamics of Plasmodium NDC80 reveals unusual modes of
1123   chromosome segregation during parasite proliferation. *J Cell Sci* **134**, doi:10.1242/jcs.245753
1124   (2020).
1125 56   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,
1126   doi:10.1093/bioinformatics/btp352 (2009).
1127 57   Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for
1128   exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:10.1093/nar/gku365
1129   (2014).
1130 58   Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-
1131   performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192,
1132   doi:10.1093/bib/bbs017 (2013).
1133 59   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
1134   doi:10.1186/gb-2008-9-9-r137 (2008).
1135 60   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
1136   features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
1137 61   Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-
1138   cell RNA sequencing data. *Genome Biol* **20**, 63, doi:10.1186/s13059-019-1662-y (2019).
1139 62   McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality
1140   control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**,
1141   1179-1186, doi:10.1093/bioinformatics/btw777 (2017).
1142 63   Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of
1143   single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122,
1144   doi:10.12688/f1000research.9501.2 (2016).
1145 64   Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529,
1146   doi:10.1016/j.cell.2021.04.048 (2021).
1147 65   Real, E. *et al.* A single-cell atlas of Plasmodium falciparum transmission through the mosquito.
1148   *Nat Commun* **12**, 3196, doi:10.1038/s41467-021-23434-z (2021).
1149 66   Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional
1150   profibrotic macrophage. *Nat Immunol* **20**, 163-172, doi:10.1038/s41590-018-0276-y (2019).
1151 67   de Kanter, J. K., Lijnzaad, P., Candelli, T., Margaritis, T. & Holstege, F. C. P. CHETAH: a
1152   selective, hierarchical cell type identification method for single-cell RNA sequencing. *Nucleic
1153   Acids Res* **47**, e95, doi:10.1093/nar/gkz543 (2019).
1154 68   Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**,
1155   1653-1659, doi:10.1093/bioinformatics/btr261 (2011).

1156   69   Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity
1157   between motifs. *Genome Biol* **8**, R24, doi:10.1186/gb-2007-8-2-r24 (2007).
1158   70   Campbell, T. L., De Silva, E. K., Olszewski, K. L., Elemento, O. & Llinas, M. Identification and
1159   genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from
1160   the malaria parasite. *PLoS Pathog* **6**, e1001165, doi:10.1371/journal.ppat.1001165 (2010).
1161   71   Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of
1162   chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).
1163   72   Gupta, M. K., Lenz, T. & Le Roch, K. G. Chromosomes Conformation Capture Coupled with
1164   Next-Generation Sequencing (Hi-C) in Plasmodium falciparum. *Methods Mol Biol* **2369**, 15-25,
1165   doi:10.1007/978-1-0716-1681-9_2 (2021).
1166   73   Varoquaux, N., Ay, F., Noble, W. S. & Vert, J. P. A statistical approach for inferring the 3D
1167   structure of the genome. *Bioinformatics* **30**, i26-33, doi:10.1093/bioinformatics/btu268 (2014).
1168   74   Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis.
1169   *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).

1170
1171
1172
1173
1174 **Figures**
1175
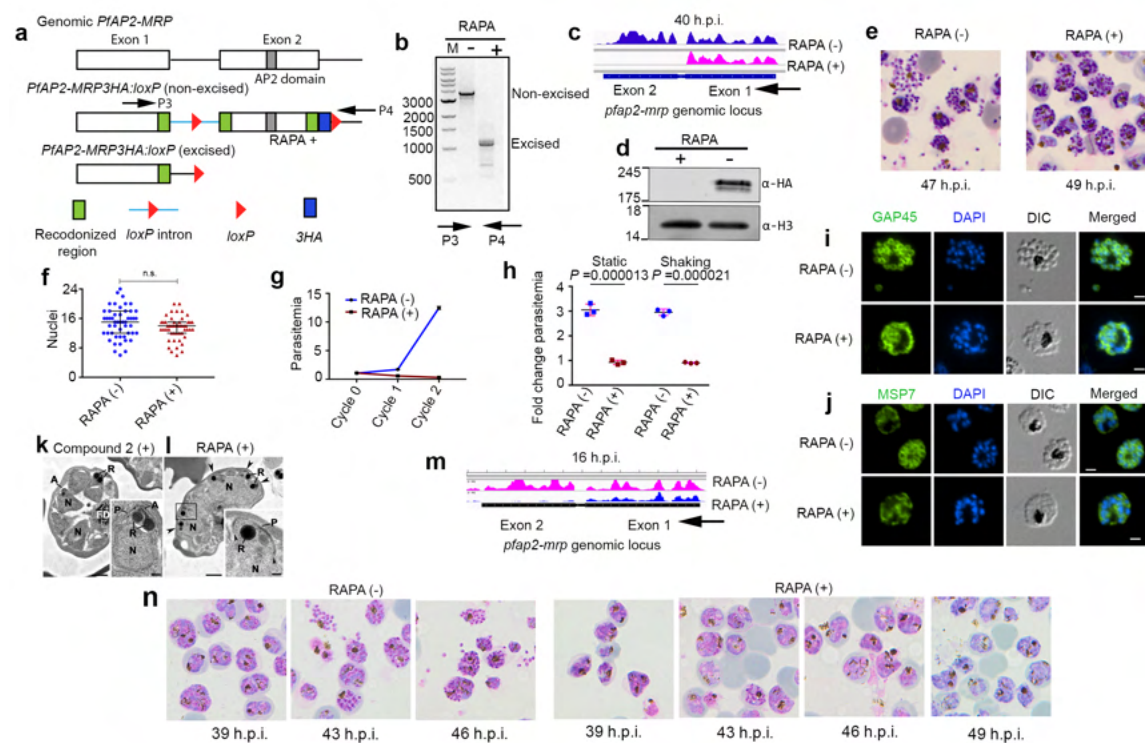
1176
1177
1178
1179 **Fig. 1| PfAP2-MRP is essential for parasite growth and development**

1180 **a,** Schematic of strategy to tag *pfap2-mrp* with 3HA sequence and its conditional knockout by excision of the *loxP*-flanked second
1181 exon. **b**, Diagnostic PCR confirms exon2 excision in RAPA-treated parasites. **c,** RNA-seq reads coverage from 40 h.p.i. parasites of
1182 *pf-ap2-mrp* locus. **d,** Western blot of control (RAPA -) and rapamycin (RAPA +) treated PfAP2-MRP-3HA:*loxP* schizont extracts
1183 probed with anti-HA and -histone H3 antibodies. Molecular mass (kDa) of standards on left side of each panel. **e,** Images of Giemsa-
1184 stained (control [RAPA (-)] and treated [RAPA (+)]; treatment at 16 h.p.i) schizonts at end of cycle 0 (representative of 4 independent
1185 experiments). **f**, Nuclei in control [RAPA (-)] and treated [RAPA (+)] schizonts. The mean number was 14.7 (control) and 13.4
1186 (treated); *P*-value of 0.178, one-way ANOVA using Sidak's multiple comparisons test, assumed Gaussian distribution. g, Replication

36

1187 of control [RAPA (-)] and treated [RAPA (+)] parasites over two growth cycles (three biological replicates, average parasitemia ±
1188 s.d.). **h**, Erythrocyte invasion by control [RAPA (-)] and treated [RAPA (+)] PfAP2-MRP-3HA:*loxP* parasites under static and shaking
1189 conditions. Statistical significance: two-tailed t-test; RAPA (–) versus RAPA (+) parasites in static conditions (n=3, t = 14.17, d.f. = 4,
1190 P = 0.000013, 95% CI 1.606 to 2.389) and RAPA (–) versus RAPA (+) parasites in shaking conditions (n=3, t = 12.41, d.f. = 4, P =
1191 0.000021, 95% CI 1.578 to 2.487). **i-j**, Immunofluorescence microscopy of control [RAPA (–)] and treated [RAPA (+)] parasites
1192 incubated with anti-GAP45 IgG (**i**) and anti-MSP7 serum (**j**); scale bar is 2 μm. **k-l**, Electron micrographs of iRBCs treated with
1193 compound 2(**k**) or rapamycin [RAPA(+)] (**l**); scale bar is 1 μm or 100 nm (inserts). **m**, RNA-seq reads coverage from 16 h.p.i.
1194 parasites (cycle 1) of *pf-ap2-mrp* locus. **n**, Images of Giemsa-stained parasites at 39, 43, 46 and 49 h.p.i. stages in cycle 1 following
1195 parasite treatment with DMSO [RAPA (-)] or rapamycin [RAPA(+)] at 35 h.p.i. at cycle 0.
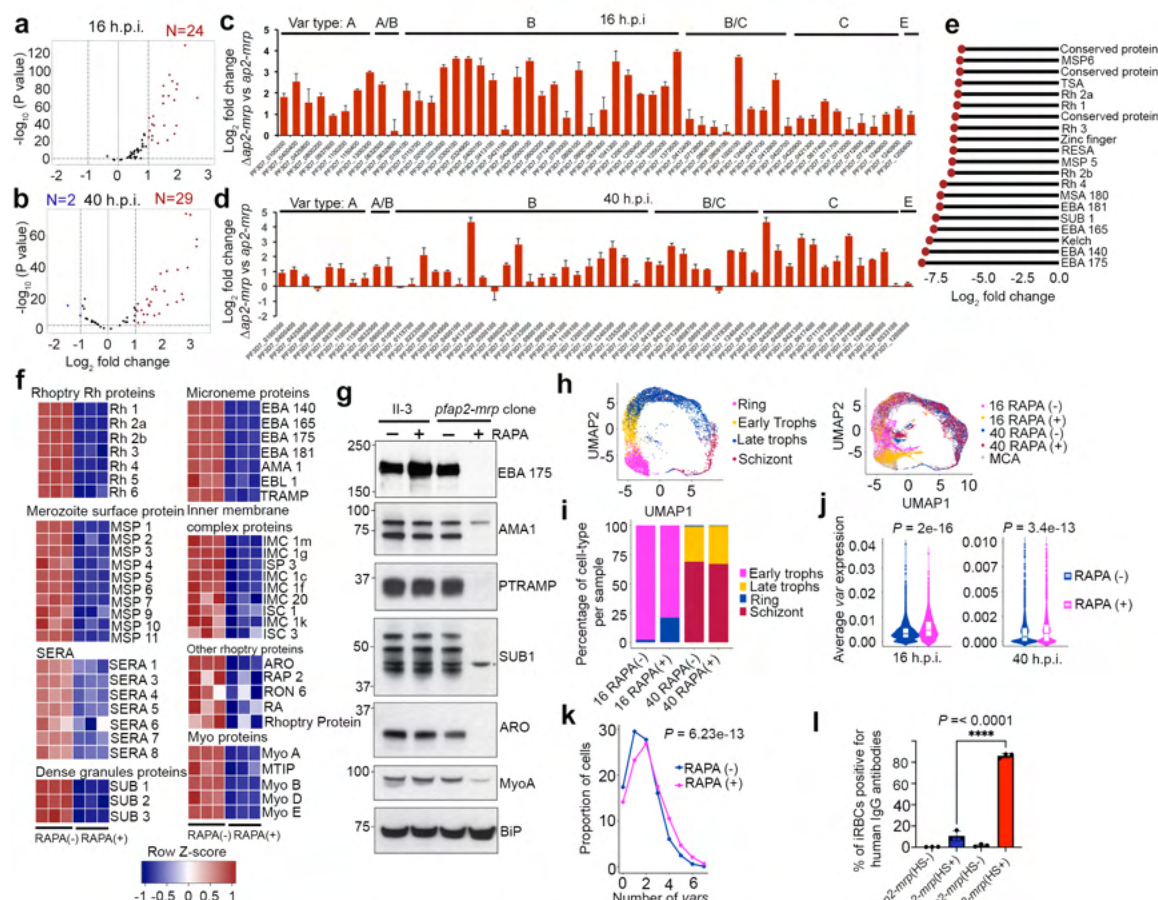
1207
1208 **Fig. 2| PfAP2-MRP regulates most of the malaria pathogenesis-associated genes**

1209 **a-b**, Volcano plots showing differentially expressed *var* genes in 16 h.p.i (cycle 1) (**a**) and 40 h.p.i. (cycle 0) (**b**) Δ*pfap2-mrp* parasites.
1210 **c-d**, Differential expression of *var* genes (log₂ ratio of Δ*pfap2-mrp* to mock-treated control parasites) measured by qRTPCR at 16
1211 h.p.i. (**c**) and 40 h.p.i. (**d**); error bar is s.e.m. **e**, Lollipop plot of expression level of top 20 significantly down-regulated genes in
1212 Δ*pfap2-mrp* compared to control parasites at 40 h.p.i. **f,** Heatmaps for most down-regulated genes known to be involved in parasite

1213 egress and invasion in treated [RAPA (+)] or control [RAPA (-)] parasites, grouped based on the sub-cellular location of their
1214 products. **g**, Western blots of schizont extract from parental II-3 and PfAP2-MRP-3HA parasites in the absence (-) or presence (+) of
1215 rapamycin, probed with antibodies specific for invasion proteins. BiP was detected as a loading control. Molecular mass (kDa) of
1216 standards on left side of each panel. A non-specific cross-reacting protein on the SUB1 blot is marked with an asterisk. **h**, Left:
1217 uniform manifold approximation and projection (UMAP) of scRNA-seq data from Malaria Cell Atlas (MCA), with annotated
1218 developmental stages. Right: UMAP projections of scRNA-seq in-house data; each dot represents gene expression data from a single
1219 parasite (colours corresponding to h.p.i. and *pfap2-mrp* knockout status) plotted over MCA data. **i**, Distribution of developmental
1220 stages of treated [RAPA(+)] or control [RAPA(-)] parasites at 16 and 40 h.p.i. **j**, Violin plots of average *var* gene expression in
1221 treated [RAPA (+)] or control [RAPA (-)] parasites at both 16 h.p.i. and 40 h.p.i. **k**, Proportion of cells expressing one or more *var*
1222 genes from treated [RAPA(+)] or control [RAPA (-)] cultures (P= 6.23e-13, Fisher's exact test, odds ratio=1.53). **l**, Percentage of
1223 iRBCs containing control or Δ*pfap2-mrp* parasites bound by IgG from serum of malaria-infected (HS+) patients, or untreated samples
1224 (HS-) . Significance determined using a two-tailed t-test (t = 6.687, d.f. = 4, P < 0.0001, 95% CI 0.9844 to 2.382, n=3).
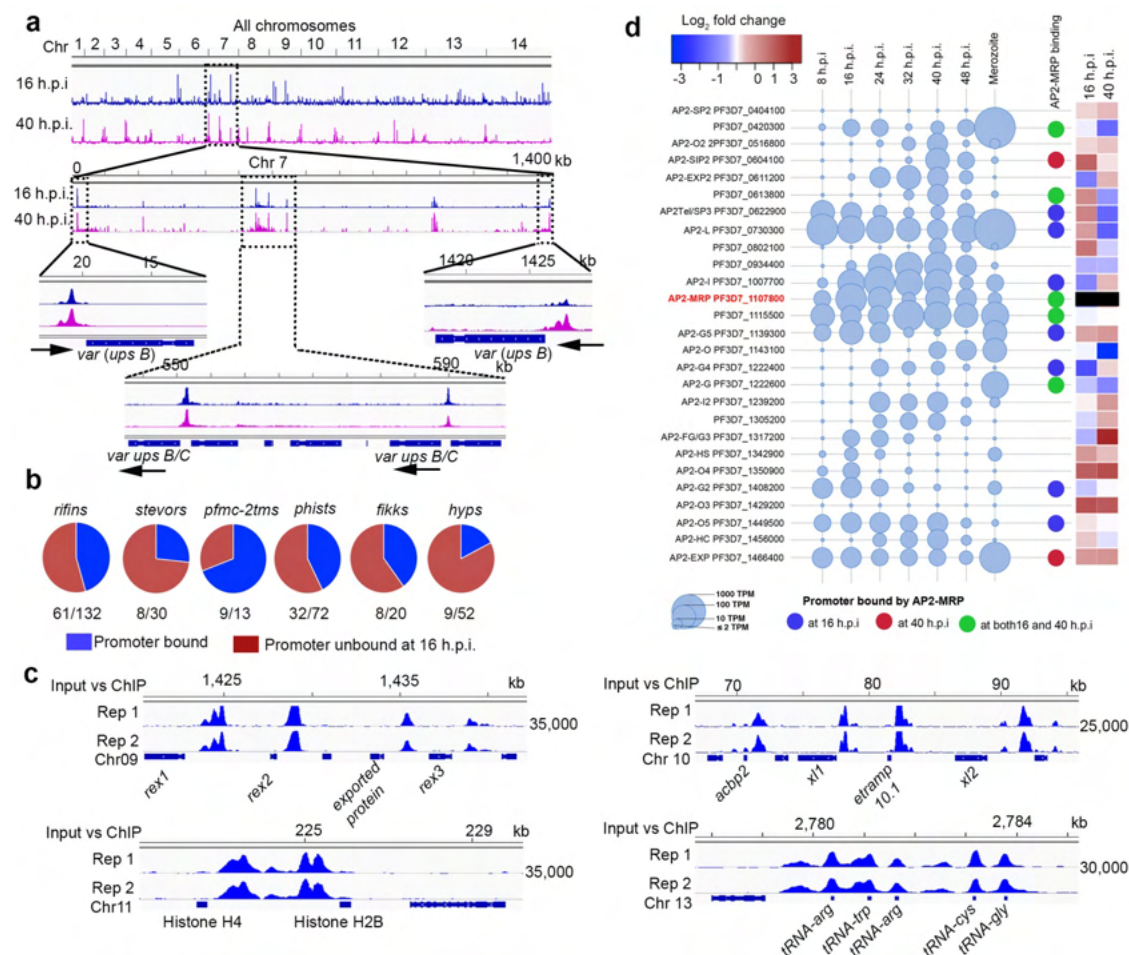
Fig. 3| PfAP2-MRP regulates pathogenesis-associated genes via promoter binding

1236 **a**, Genome-wide occupancy of PfAP2-MRP at 16 h.p.i. and 40 h.p.i., determined by ChIP-seq. Sub-telomeric and internal regions of
1237 chromosome 7 (~1450 kb) containing *var* genes with PfAP2-MRP bound in their promoter regions are shown as an example.

38

1238 Chromosomal positions are indicated. Results are representative of 2 independent replicates. **b**, Pie charts showing the proportion of
1239 each family of genes with PfAP2-MRP bound to the promoter region (in blue) at 16 h.p.i. **c**, PfAP2-MRP occupancy at 16 h.p.i. in
1240 putative promoter regions of genes implicated in iRBC remodeling and parasite development. Two biological replicate ChIP vs. Input
1241 tracks are shown, (input subtracted PfAP2-MRP-ChIP). Positions on Chr 09, 10, 11 and 13 are indicated. X-axis shows the genomic
1242 position and numbers on the right, show the enrichment score (see Methods). **d**, Expression levels of 27 *P. falciparum* AP2 genes
1243 during different IDC stages and in merozoites are depicted by the diameter of the circles. On the right, blue, brown and green circles
1244 indicate the binding of PfAP2-MRP to the promoter at either or both 16 h.p.i. and 40 h.p.i.; the heatmap displays the expression status
1245 of all 26 *api-ap2s* in Δ*pfap2-mrp* parasites compared to controls at 16 and 40 h.p.i. The heatmap for *pfap2-mrp* is black because
1246 Δ*pfap2-mrp* parasites only express RNA from the first exon and have no functional AP2-MRP protein.

1247
1248
1249
1250
1251
1252
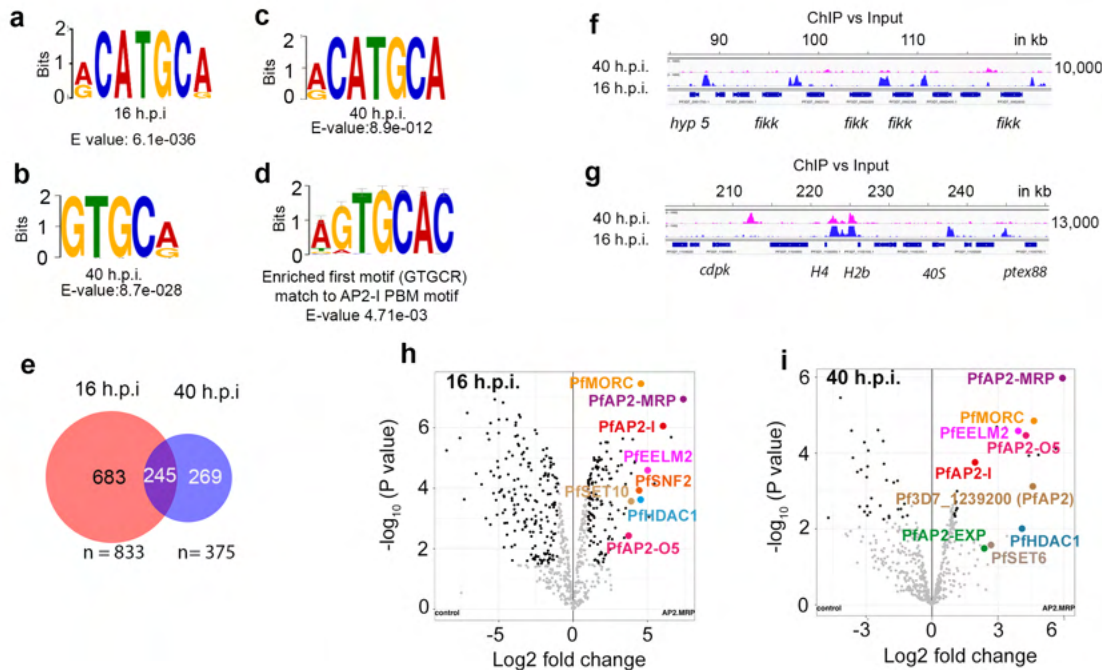
1253
1254
1255
1256 **Fig. 4| PfAP2-MRP binds to specific DNA motifs and associates with chromatin**
1257 **remodelers**
1258 **a,** Most significantly enriched motif bound by PfAP2-MRP at 16 h.p.i. **b-c**, The two most significantly enriched motifs bound by
1259 PfAP2-MRP at 40 h.p.i. **d**, The most enriched motif (panel **c**) at 40 h.p.i. is similar to the PfAP2-I binding motif **e**, The numbers of
1260 genes with promoters bound by PfAP2-MRP at either 16 h.p.i., 40 h.p.i. or both. **f-g**, Genomic regions uniquely bound by PfAP2-
1261 MRP at 16 h.p.i. (**f**) or at both 16 and 40 h.p.i. (**g**). X-axis shows the genomic position and numbers on the right, show the enrichment
1262 score. **h-i**, Label-free quantitative proteomic analysis of *P. falciparum* proteins enriched in PfAP2-MRP immunoprecipitates at
1263 16h.p.i. and 40 h.p.i.

1264
1265
1266
1267
1268
1269
1270

1271
1272
1273
1274
1275
1276
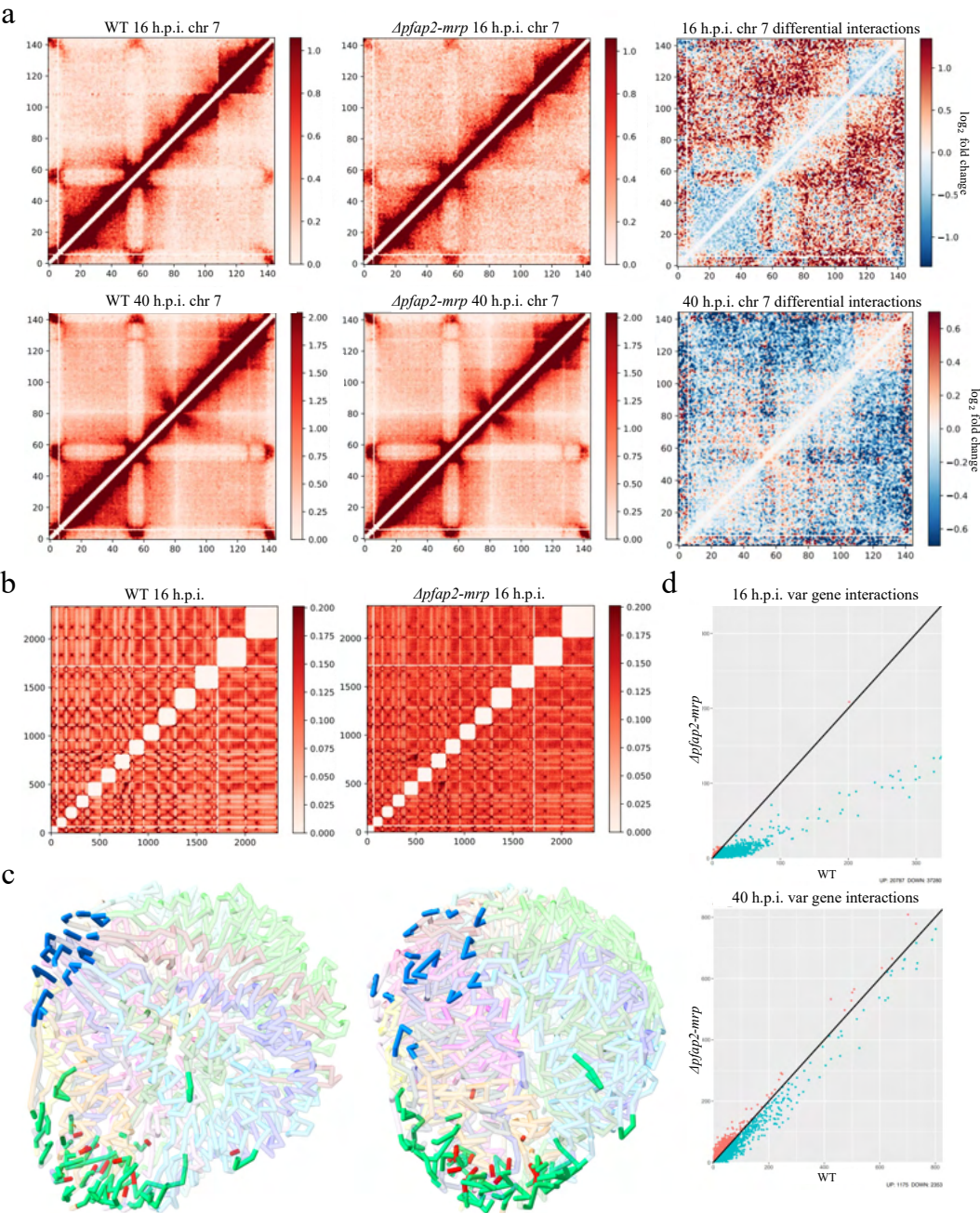
1277
1278 **Fig. 5| Depletion of PfAP2-MRP increases chromatin accessibility**
1279 **a,** Chromatin contact count heatmap of chromosome 7 at 16 h.p.i. (top) and 40 h.p.i (bottom) for the WT
1280 (left) and *Δpfap2-mrp* (middle), as well as the log_2 fold change in interactions (right) between the WT and

1281 *Δpfap2-mrp*. Blue indicates a loss of interactions and red indicates an increase of interactions *Δpfap2-mrp*
1282 over WT. **b,** Whole-genome interchromosomal contact count heatmaps at 16 h.p.i. for the WT (left) and
1283 *Δpfap2-mrp* (right). Chromosomes are sorted from left to right and bottom to top. Intrachromosomal
1284 interactions are removed. **c,** Whole-genome 3D chromatin models at 16 h.p.i. for the WT (left) and *Δpfap2-*
1285 *mrp* (right). Centromeres (blue), telomeres (green), and *var* genes (red) are enhanced to display differences
1286 between the two samples. **d,** Number of interactions between *var* gene containing bins.
1287
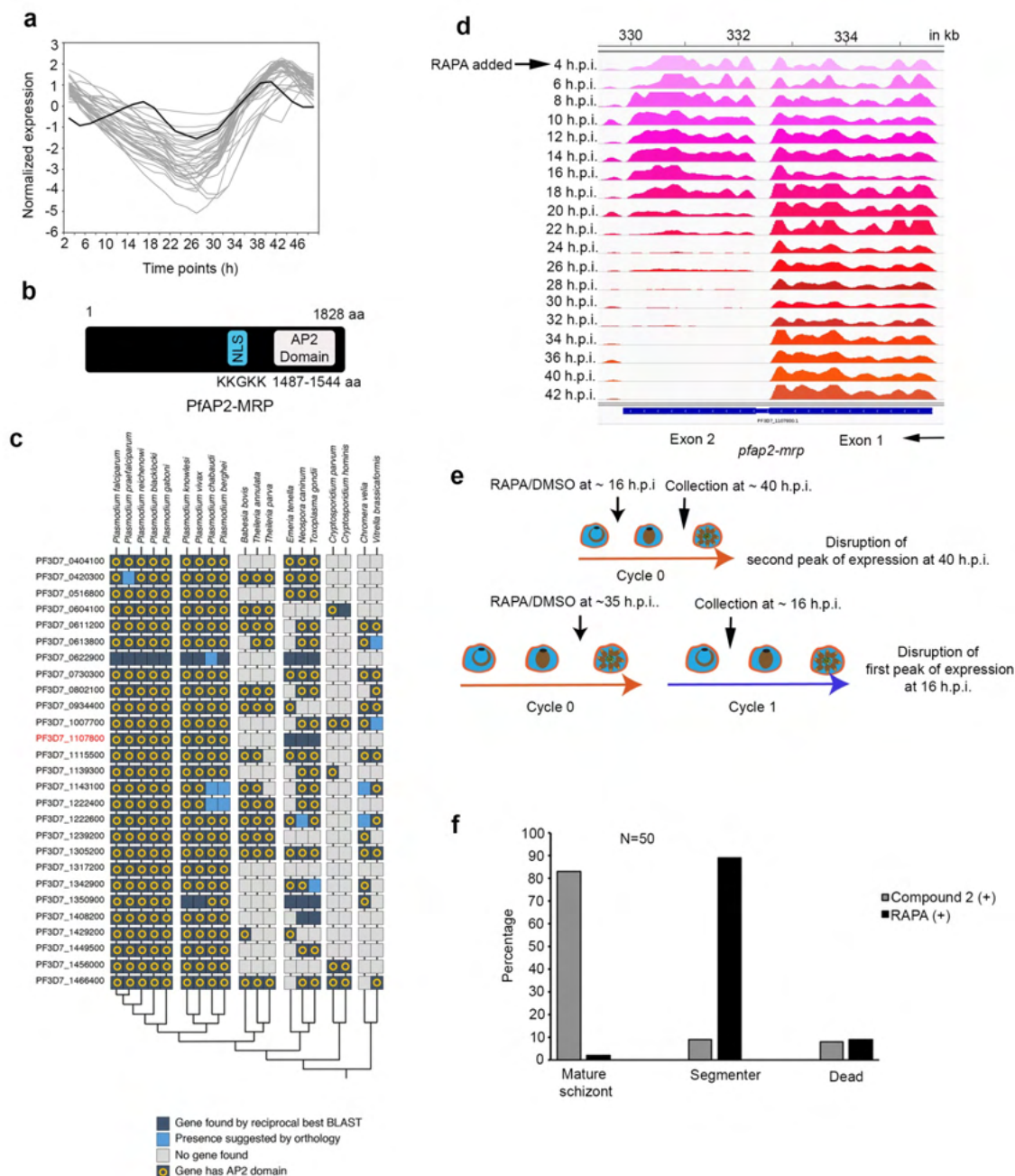
1288
1289
Extended Data Fig. 1| *pfap2-mrp* knockout affects merozoite development
1291 **a,** Expression of *pfap2-mrp* (thick black line) and the top 40 down-regulated genes (grey lines) in *P. falciparum* line II3 over a 48-hour
1292 IDC[14]. **b,** The primary structure of PfAP2-MRP, with a nuclear localization signal (NLS) and a single AP2 DNA binding domain.
1293 Both NLS and DNA binding domain are encoded by exon 2 of the gene. **c,** A phylogenetic distribution of AP2 genes in the genus

1294 *Plasmodium* and related alveolates. **d**, RNA-seq data from different IDC time-points mapped to the *pfap2-mrp* locus; there is a drastic
1295 reduction in RNA-seq reads mapping to the second exon 16 hours after the addition of rapamycin i.e. at 20 h.p.i. **e**, Schematic showing
1296 rapamycin treatment schedule to disrupt either first or second peak of *pfap2-mrp* expression. **f**, The different developmental stages of
1297 the Compound 2- and RAPA-treated parasites (at 49 h.p.i.); 50 randomly selected iRBCs from each group were inspected and the
1298 parasites were categorized as either mature schizont, segmenter or dead.

1299
1300
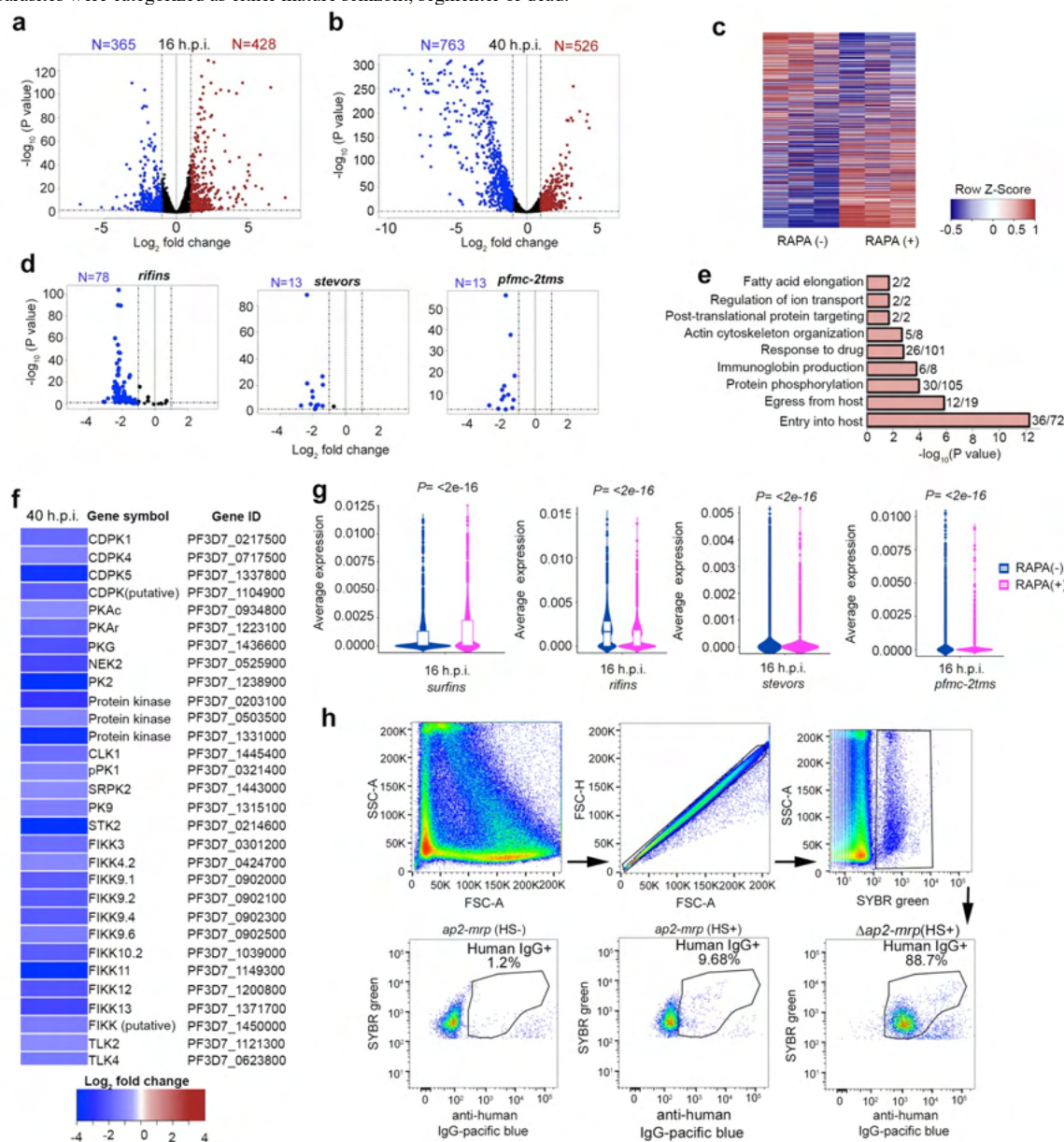
Extended Data Fig. 2| Disruption of *pfap2-mrp* deregulates malaria pathogenesis-associated genes.

a-b, Volcano plots showing significantly differentially expressed genes in rapamycin-treated compared to control parasites at 16 h.p.i.
(**a**) and 40 h.p.i. (**b**). **c**, Non-differentially expressed genes in treated [RAPA (+)] compared to control [RAPA (-)] parasites at 40 h.p.i.,
which had been reported to express at least 4-fold higher in schizont-stage parasites (> 35 h.p.i.) compared to early-stage parasites (<
35 h.p.i.)[5]. **d**, Expression of members of the gene families: *rifin*, *stevor* and *Pfmc-2tm* at 16 h.p.i.that encode antigenically variant
proteins. **e**, Gene-ontology (GO) enrichment analysis of all genes down-regulated in rapamycin-treated compared with control
parasites at 40 h.p.i. Shown is the number of genes down-regulated in treated parasites out of the total number of genes assigned to
that specific GO term. **f**, Heatmap of the expression of all known and putative *P. falciparum* kinases downregulated following
rapamycin treatment compared to controls, at 40 h.p.i. **g**, Violin plots of all *surfin*, *rifin, stevor* and *pfmc-2tm* expression per cell in
treated [RAPA (+)] and control [RAPA (-)] parasites at 16 h.p.i.. **h**, FACS gating strategy for surface PfEMP1 expression detected by
IgG binding from pooled serum of malaria-infected individuals. Cells were incubated either without (HS-) or with (HS+) pooled
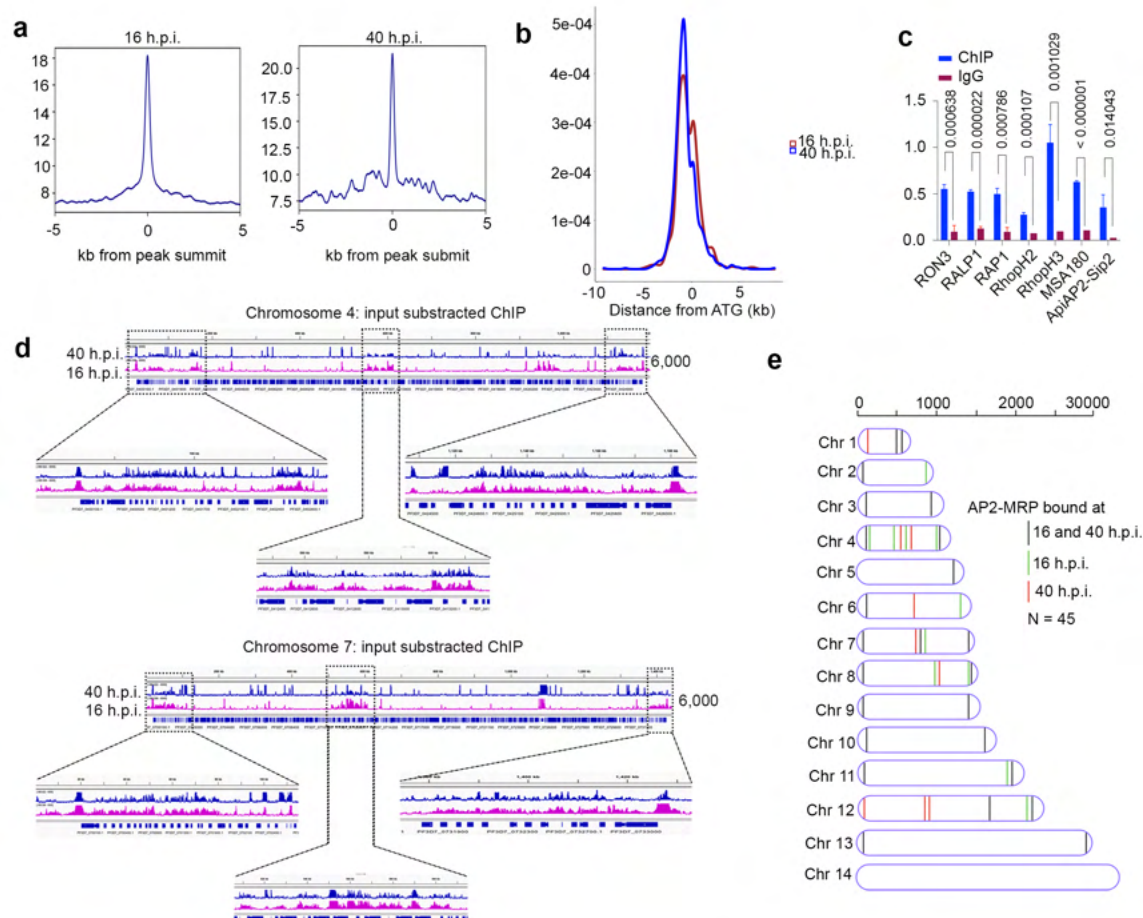
42

1313 serum from malaria-infected individuals. The percentage of Human IgG+ iRBCs is the mean of three biological replicates. A total of
1314 7,500 Sybr green positive events (iRBCs) per sample were analyzed.
1315
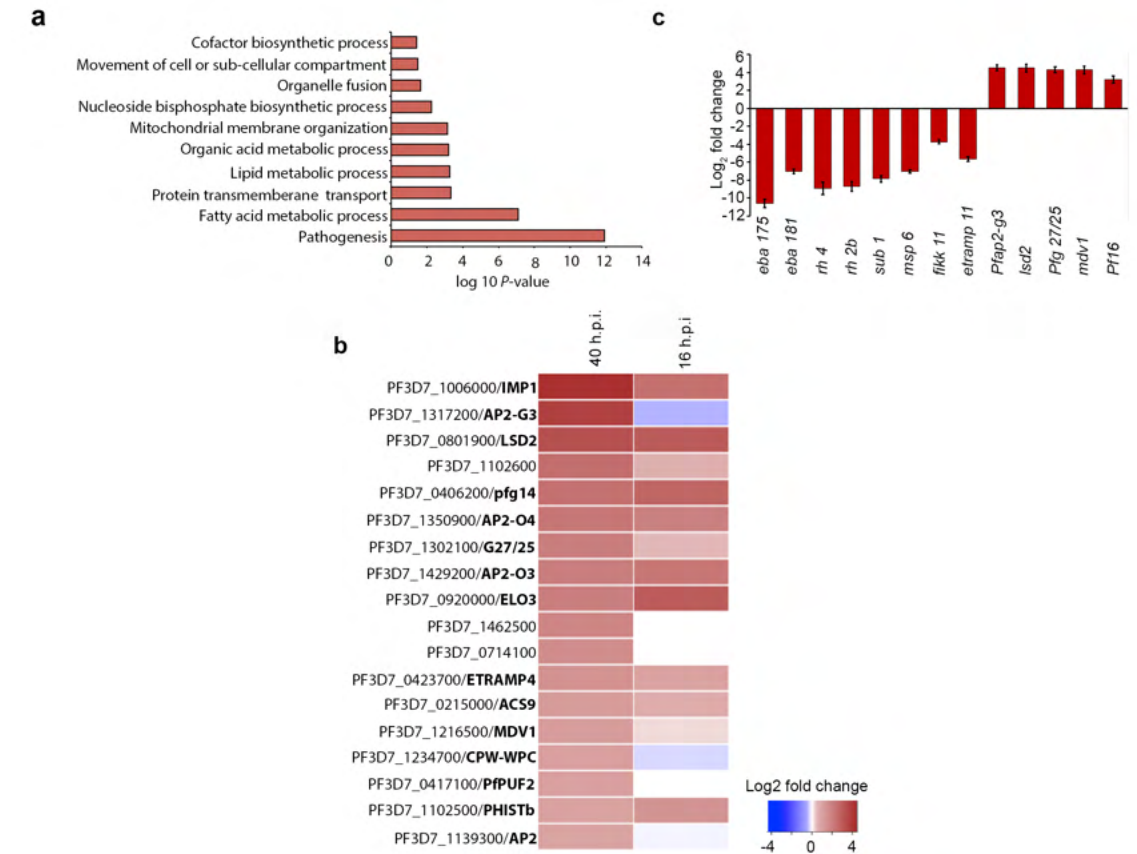
1316
1317
1318
1319 **Extended Data Fig. 3| PfAP2-MRP is a repressor of early gametocyte marker genes.**
1320 **a**, Gene ontology enrichment analysis of up-regulated genes in Δ*pfap2-mrp* parasites at 40 h.p.i. **b**, Differential expression in Δ*pfap2-*
1321 *mrp* parasites at 40 h.p.i. and 16 h.p.i. of genes that are known or putative early gametocyte markers. **c**, Differential expression of
1322 selected genes in RAPA-treated and control parasites measured using qRT-PCR, to validate RNA-seq data.
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338

43

1339
1340
1341
1342
1343

1344
1345 **Extended Data Fig. 4| PfAP2-MRP binds to the putative promoter regions of *var* genes.**
1346
1347 **a**, Enrichment of PfAP2-MRP bound reads (from replicate 1) around the ± 5 kb region of peak summits, from 16 h.p.i. (left panel) and
1348 40 h.p.i. (right panel) parasite samples. **b**, The position of ChIP-seq peak summits (common between two biological replicates)
1349 relative to the predicted ATG translational start codon, in 16 h.p.i. (red) and 40 h.p.i. (blue) parasites. **c**, Three independent PfAP2-
1350 MRP ChIP experiments followed by qPCR, were performed to validate ChIP-seq data, using selected PfAP2-MRP-bound promoter
1351 regions of genes from samples at 40 h.p.i. The bar-plot shows percent input (% Input) enrichment of PfAP2-MRP on target genes
1352 (mean ± SD of three independent experiments). IgG was used as the mock-treated control. *P*-values were calculated using a two-tailed
1353 t-test. **d**, Input subtracted ChIP peaks of PfAP2-MRP in chromosome 4 and 7 as representatives in both 16 and 40 h.p.i. stages. Also,
1354 zoomed in PfAP2-MRP bound central chromosomal and sub-telomeric heterochromatin regions are shown. X-axis shows the genomic
1355 position and numbers on the right show the enrichment score. **e**, Schematic diagram of the chromosomal position of *var* genes with
1356 promoters bound by PfAP2-MRP at either 16 h.p.i. (green), 40 h.p.i. (red) or at both stages (black).
1357
1358
1359
1360
1361
1362
1363
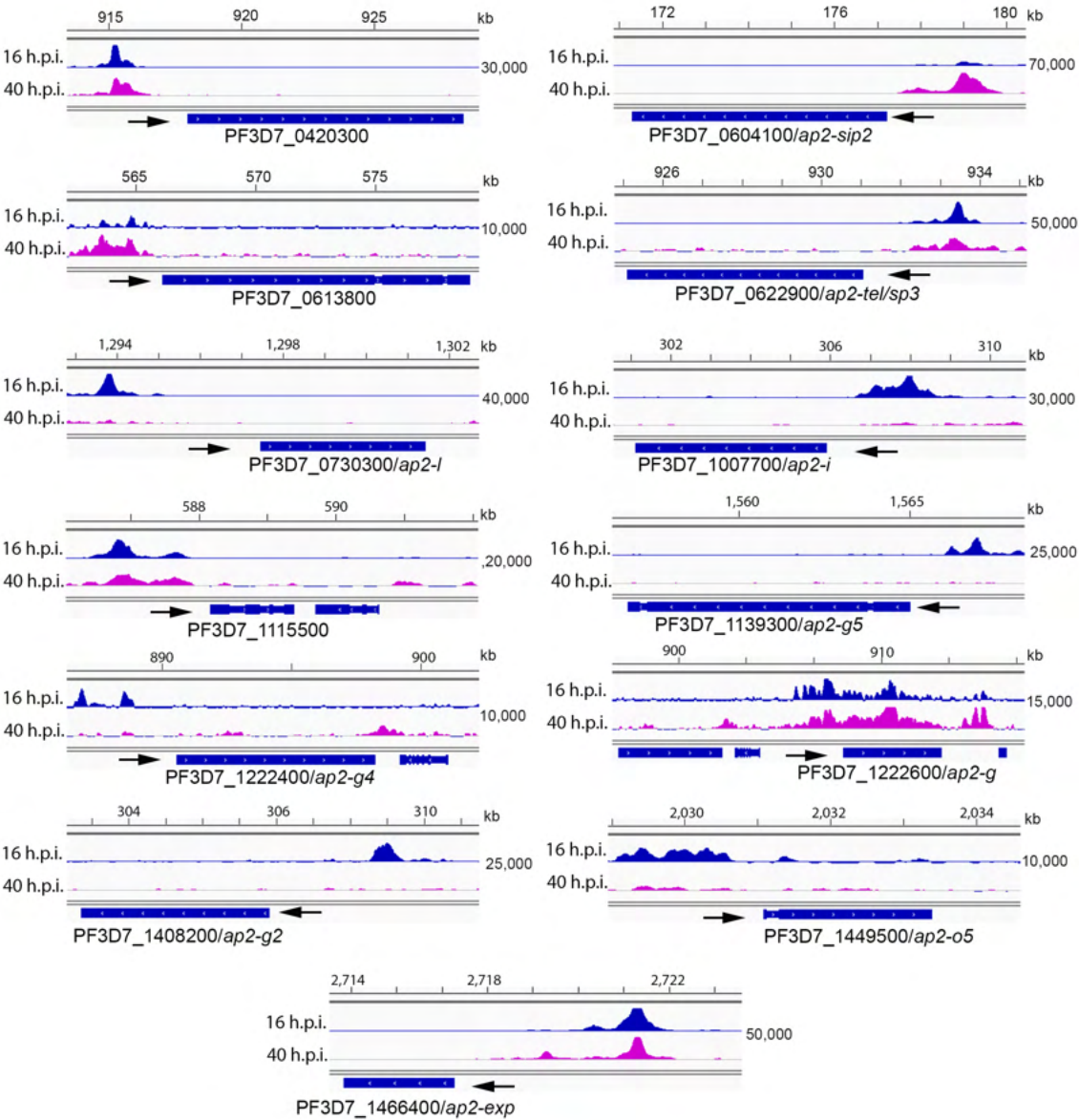
Extended Data Fig. 5| PfAP2-MRP binds to the putative promoter regions of multiple *apiap2* genes.

Occupancy of PfAP2-MRP in the promoter region of *apiap2* genes. ChIP tracks show input subtracted PfAP2-MRP-ChIP from 16 and 40 h.p.i. parasites. Arrow marks show the direction of gene transcription X-axis shows the genomic position, and numbers on the right show the enrichment score.
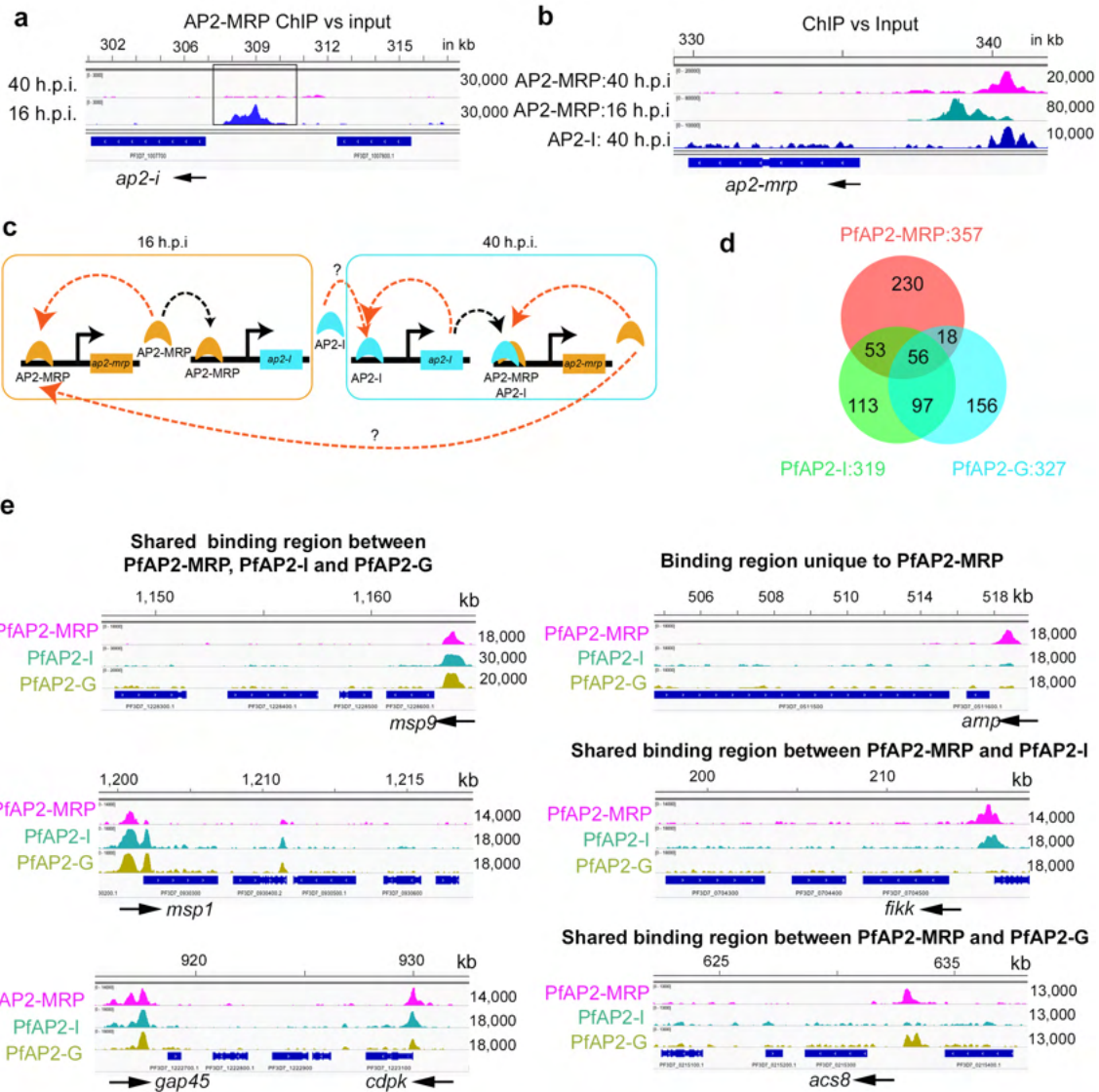
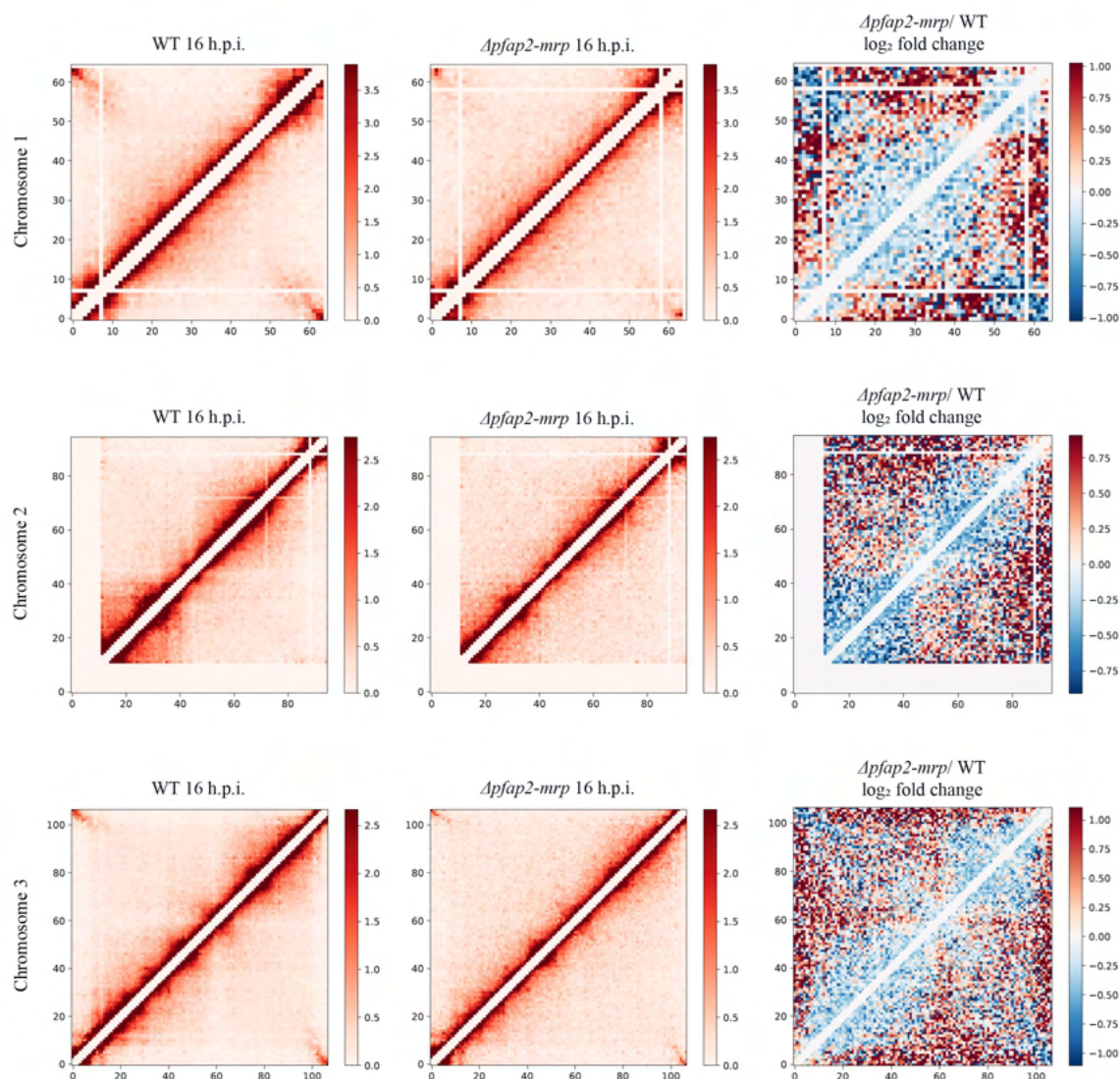**Extended Data Fig. 6| PfAP2-MRP, PfAP2-I and PfAP2-G binds to many common genomic regions**

a, Occupancy of PfAP2-MRP in the promoter region of *pfap2-I* at 16 and 40 h.p.i. **b**, Occupancy of PfAP2-MRP and PfAP2-I in the promoter region of *pfap2-mrp* at 16 and 40 h.p.i. **c**, Schematic showing probable gene regulatory network between PfAP2-MRP and PfAP2-I. Orange arrows indicate the binding of protein to its gene promoter. **d**, Comparison of genes with promoters bound by PfAP2-MRP (red), PfAP2-I (green) and PfAP2-G (light blue) at 40 h.p.i. **e**, Input-subtracted ChIP-seq read coverage for exemplar genes with promoters that are either bound by all three AP2 proteins (left panel), uniquely by PfAP2-MRP, by both PfAP2-MRP and PfAP2-I, or by both PfAP2-MRP and PfAP2-G (right panel). Arrows show the direction of gene transcription. X-axes show the chromosomal position, and the numbers on the side show the peak enrichment score. X-axis shows the genomic position and numbers on the right show the enrichment score.
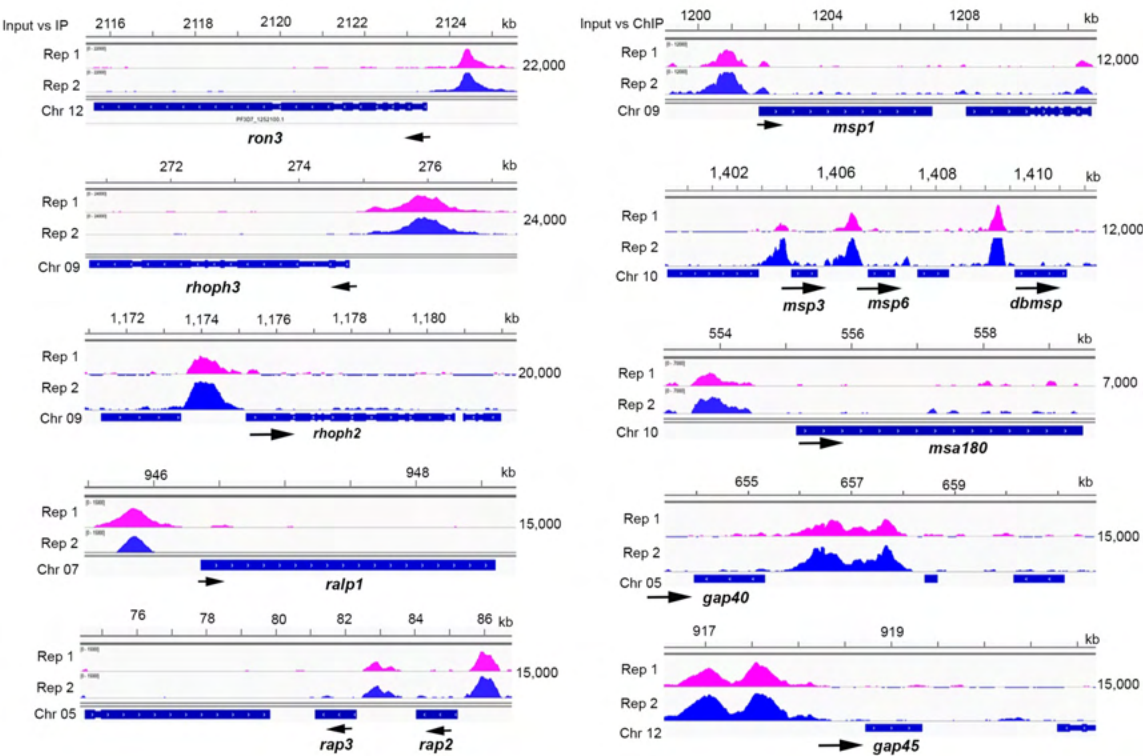
Extended Data Fig. 7| PfAP2-MRP binds to the putative promoter region of invasion-associated genes.

Occupancy of PfAP2-MRP in the promoter region of different invasion associated genes. The ChIP tracks show two replicates with input subtracted from the PfAP2-MRP-ChIP data. Arrows indicate direction of transcription. X-axis shows the chromosomal position, and numbers on the right, show the enrichment score.
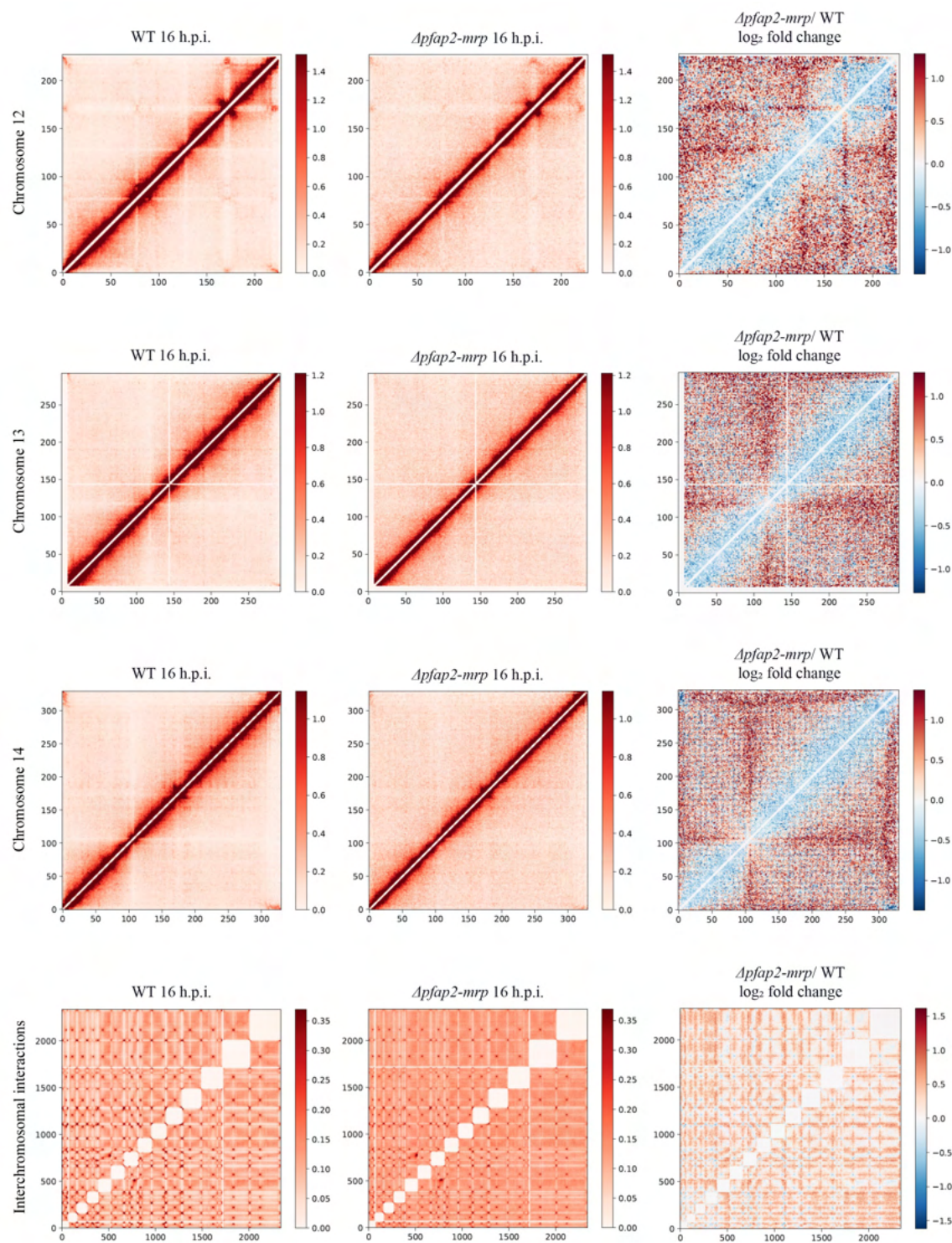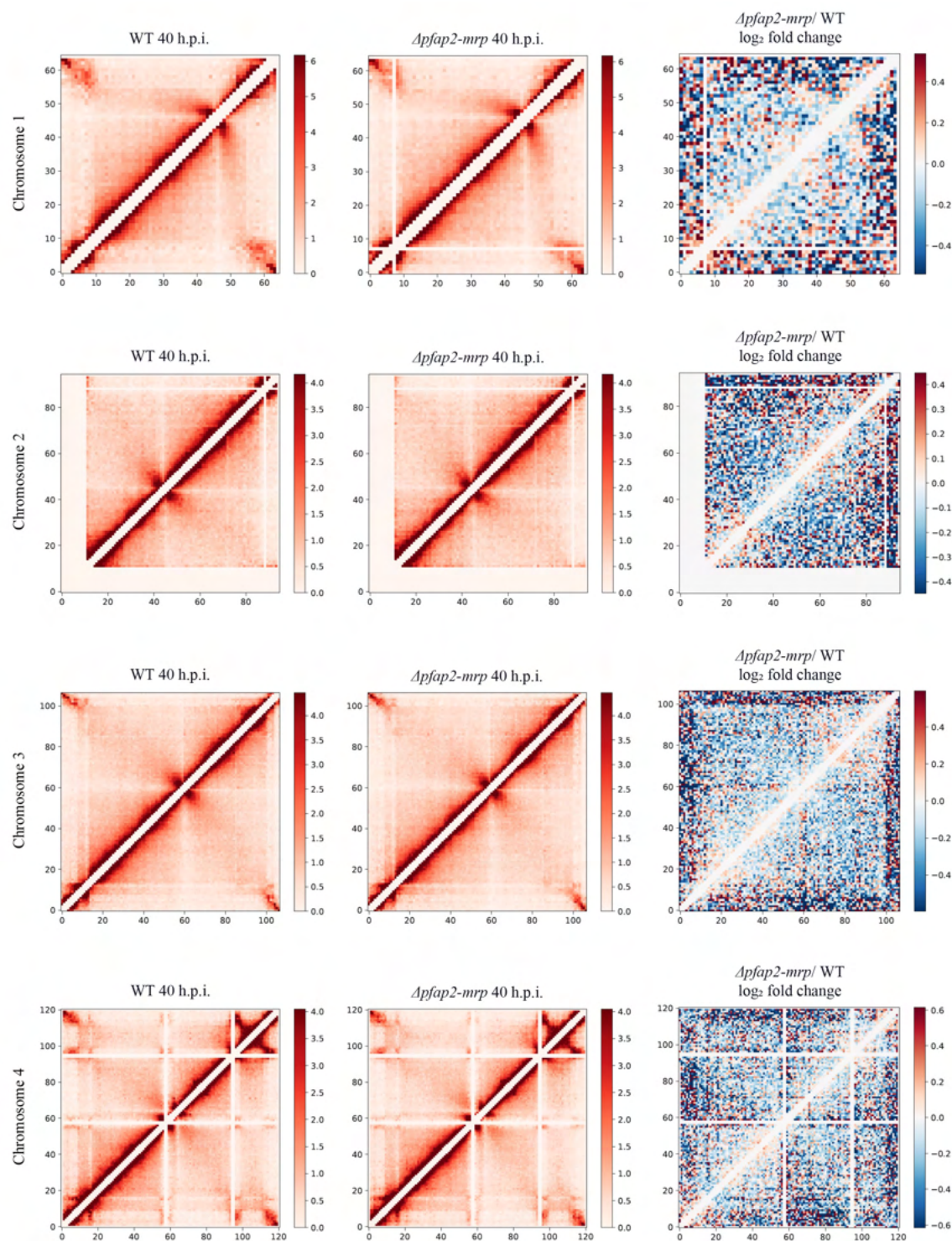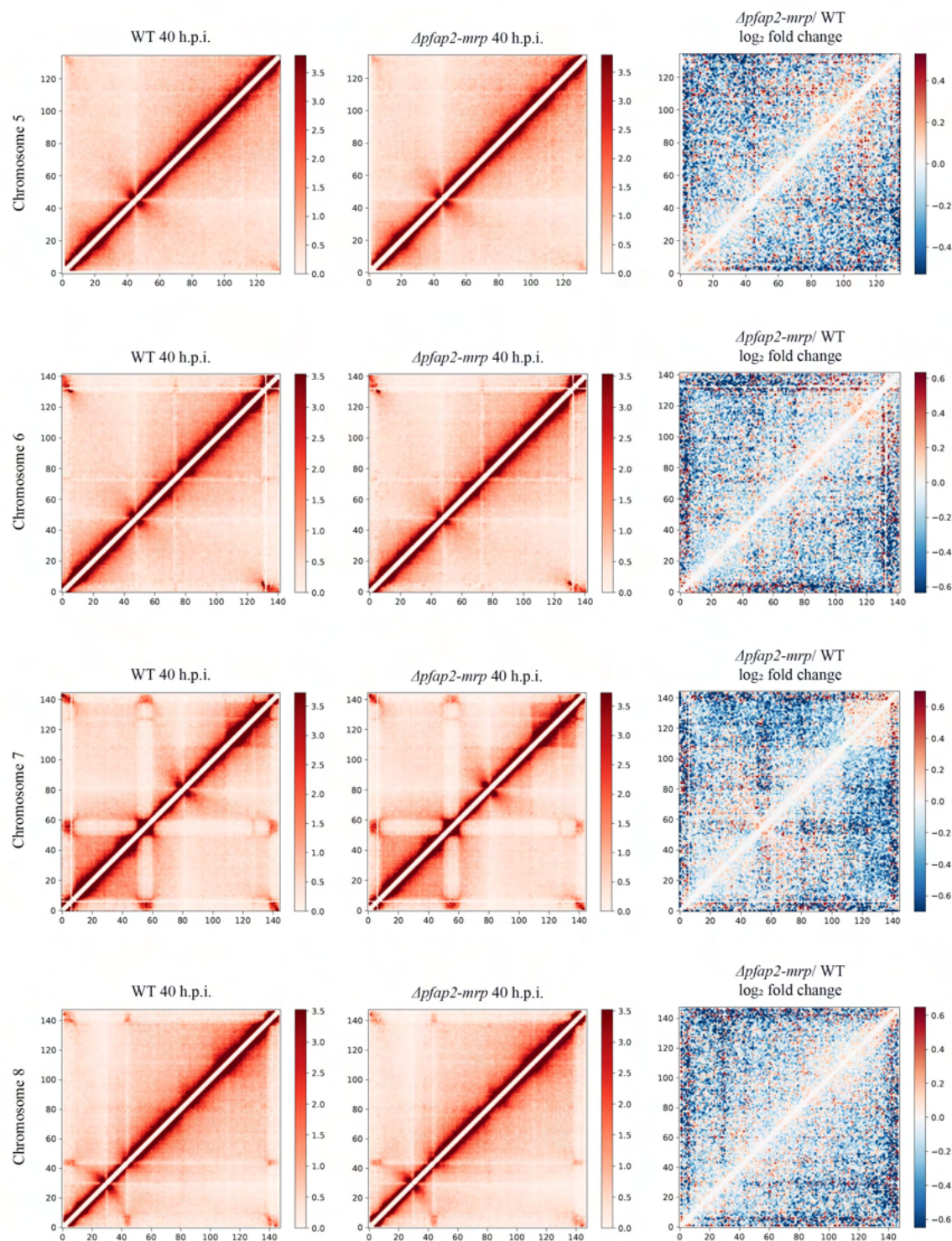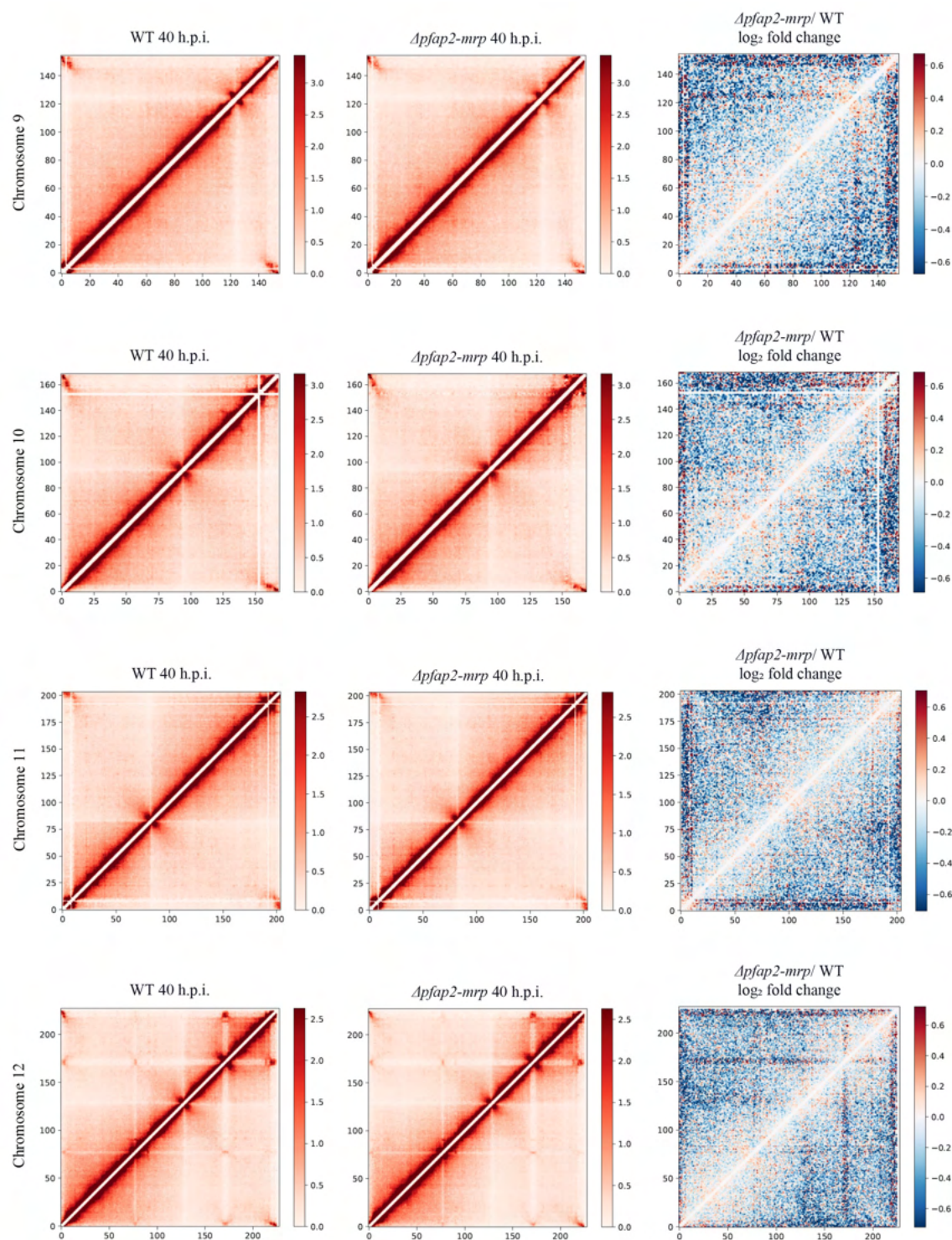
1417
1418
1419
1420
1421
1422
1423

1424
1425
1426
1427
1428
1429
1430

1431
1432
1433
1434
1435
1436
1437

50
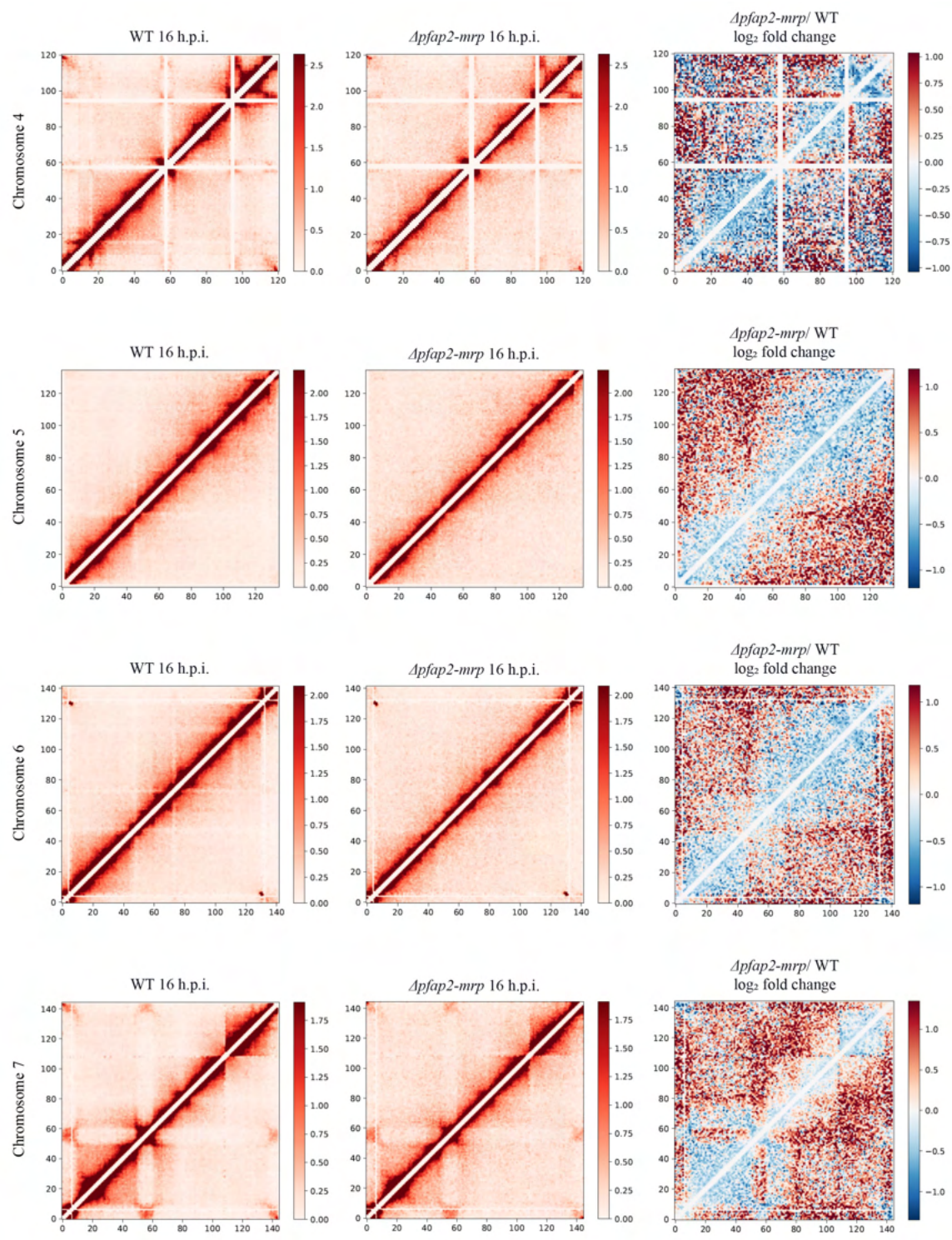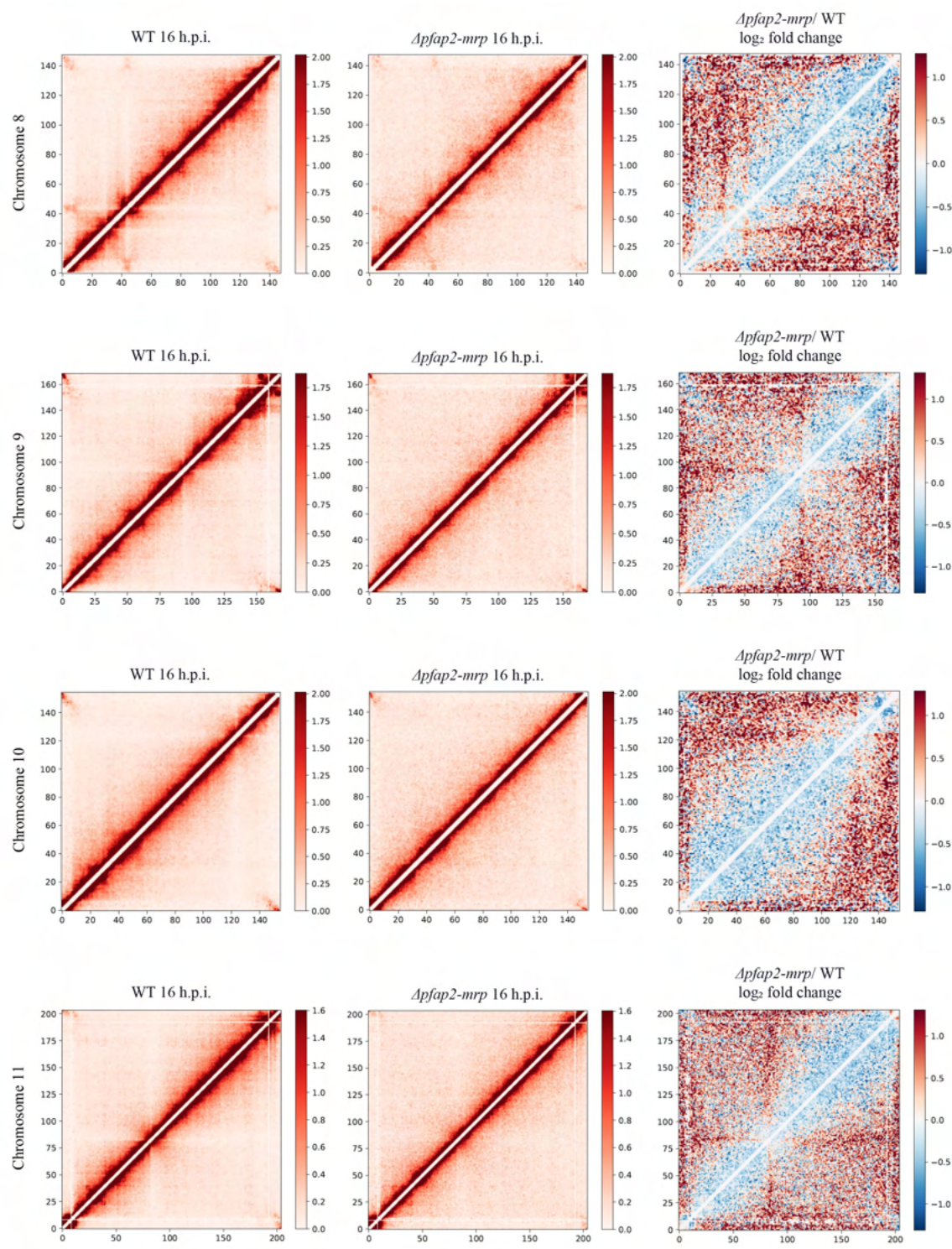
1438
1439
1440
1441
1442
1443
1444
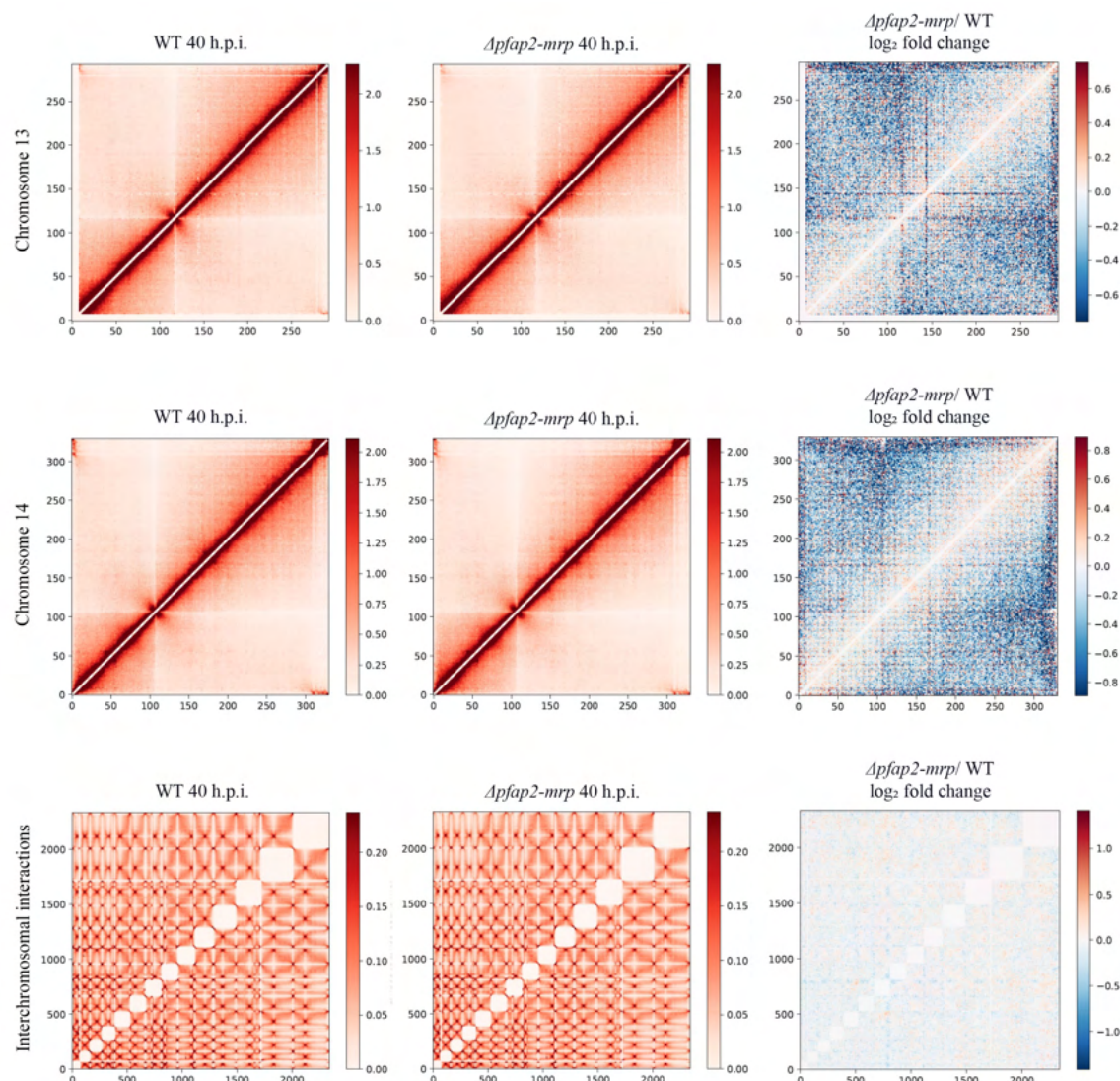
1459
1460
1461
1462
1463
1464
1465

54

1466
1467 **Extended Data Fig. 8| Visualizations of in-situ Hi-C generated chromatin**
1468 **interactions**
1469 ICE-normalized contact count heatmaps at 10kb resolution of intrachromosomal interactions for the 14 chromosomes, as well as a
1470 genome-wide contact count heatmap showing interchromosomal interactions, are given for both the 16 h.p.i. and 40 h.p.i. time points.
1471 Each row represents a single chromosome of the wild type (left), *Δpfap2-mrp* (center), and log$_2$ fold change differential interactions
1472 (right). The data for the two biological replicates were merged using a weighted average based on the total read count and then counts-
1473 per-million normalized prior to generating heatmaps. The scale of the legend was also normalized to improve comparisons between
1474 WT and *Δpfap2-mrp*. For each bin *i,* all interactions within *i* +/- 2 are set to 0 (see white line at diagonal) to enhance visualization of
1475 remaining bins due to intra-bin and very short-range contacts being significantly higher
1476
1477
1478
1479
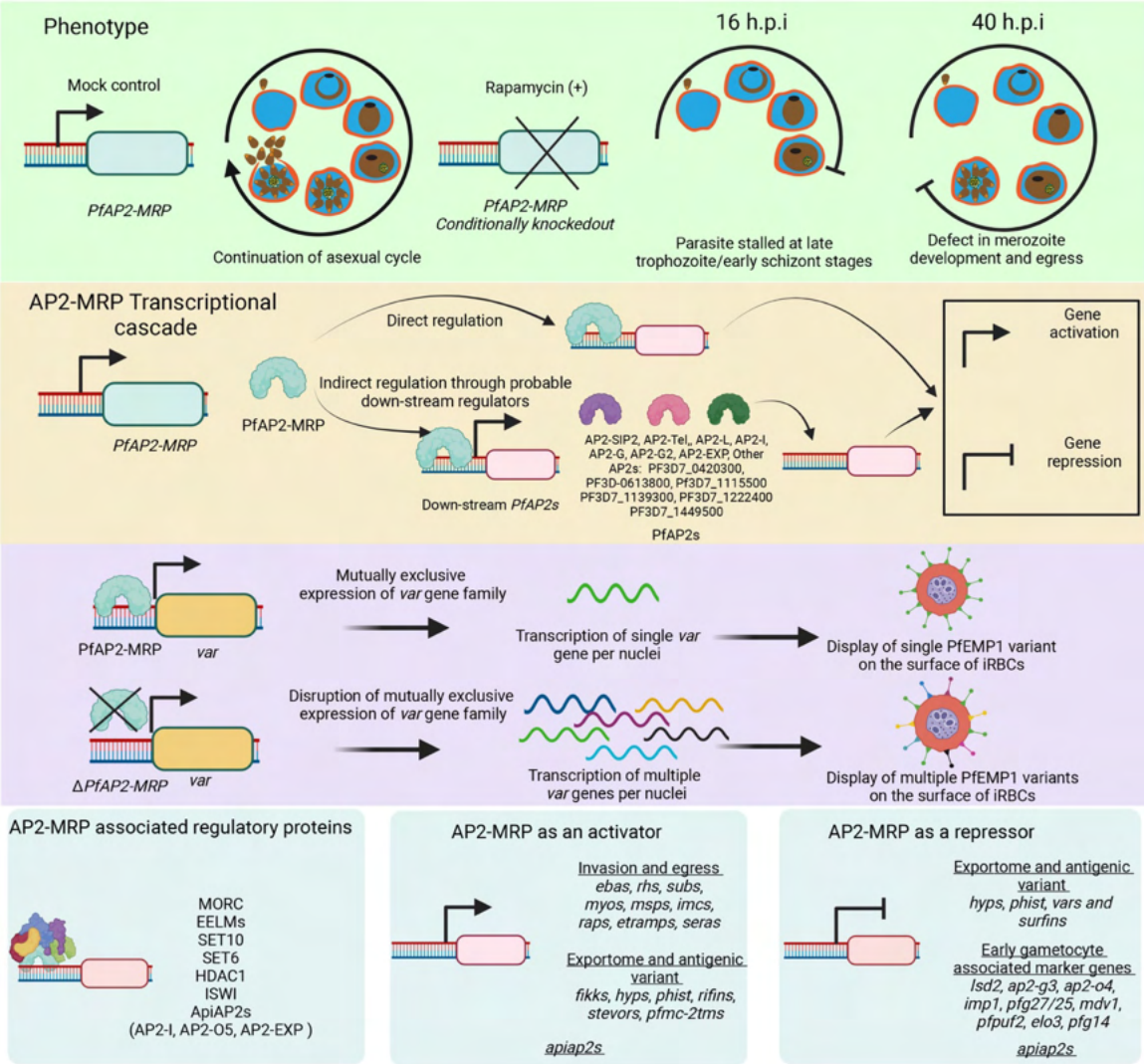
Extended Data Fig. 9| PfAP2-MRP is a master regulator of malaria pathogenesis

Deletion of *pfap2-mrp* before the first peak of expression at 16 h.p.i. affects parasite development beyond late trophozoite/early schizont stages. Deletion of *pfap2-mrp* well before 40 h.p.i. but after the first peak of expression at 16 h.p.i. affects merozoite development and blocks parasite egress from infected RBCs. At this late stage of intraerythrocytic development, the second peak of *pfap2-mrp* expression activates many genes associated with invasion, egress, antigenic variation, host cell remodeling and protein phosphorylation, either directly by binding to their promoter or indirectly through other downstream ApiAP2 transcription factors and regulators. PfAP2-MRP is a direct repressor of *var* genes and an indirect repressor of many gametocytogenesis-associated marker genes. Deletion of *pfap2-mrp* derepresses expression of most of *var* genes leading to the displayed of the corresponding PfEMP1 on the iRBC surface. PfAP2-MRP acts as a direct activator of many other genes such as *rifins*, *stevors*, and *pfmc-2tms* coding for antigenically variant proteins, and binds to the promoter of 14 other ApiAP2s (50% of all *P. falciparum* ApiAP2 genes), suggesting that it is an upstream regulator of gene expression cascades during the IDC. PfAP2-MRP associates with many known and putative histone modifiers and chromatin remodelers that probably participate in PfAP2-MRP associated gene regulation. Altogether, PfAP2-MRP regulates the expression of most known pathogenic factors (associated with antigenic variation and parasite growth) in *P. falciparum* suggesting it is a master regulator of malaria pathogenesis.