

OPEN

# An epigenetic map of malaria parasite development from host to vector

Kathrin Witmer<sup>1\*</sup>, Sabine A. Fraschka<sup>2,3</sup>, Dina Vlachou<sup>1</sup>, Richárd Bártfai<sup>2</sup> & George K. Christophides<sup>1\*</sup>

The malaria parasite replicates asexually in the red blood cells of its vertebrate host employing epigenetic mechanisms to regulate gene expression in response to changes in its environment. We used chromatin immunoprecipitation followed by sequencing in conjunction with RNA sequencing to create an epigenomic and transcriptomic map of the developmental transition from asexual blood stages to male and female gametocytes and to ookinetes in the rodent malaria parasite *Plasmodium berghei*. Across the developmental stages examined, heterochromatin protein 1 associates with variably expressed gene families localised at subtelomeric regions and variant gene expression based on heterochromatic silencing is observed only in some genes. Conversely, the euchromatin mark histone 3 lysine 9 acetylation (H3K9ac) is abundant in non-heterochromatic regions across all developmental stages. H3K9ac presents a distinct pattern of enrichment around the start codon of ribosomal protein genes in all stages but male gametocytes. Additionally, H3K9ac occupancy positively correlates with transcript abundance in all stages but female gametocytes suggesting that transcription in this stage is independent of H3K9ac levels. This finding together with known mRNA repression in female gametocytes suggests a multilayered mechanism operating in female gametocytes in preparation for fertilization and zygote development, coinciding with parasite transition from host to vector.

Malaria is caused by apicomplexan parasites of the genus *Plasmodium* and is transmitted to humans through bites of anopheline mosquitoes. Clinical cases and deaths decreased significantly over the past decade but began to plateau since 2015 indicating that current measures have now reached their maximum capacity and that new measures are urgently needed<sup>1</sup>. Transmission through the mosquito vector is a natural bottleneck in parasite development and a favorable stage for interventions aiming at malaria control and elimination. Therefore, research towards understanding parasite development in the mosquito has been intensified in recent years.

Haploid parasites infect and asexually replicate in the red blood cells (RBCs) of the mammalian host causing disease. In each replication cycle, a fraction of parasites differentiates into sexual forms called gametocytes, the stage infective to mosquitoes. Upon a bite from a mosquito, gametocytes sense the change in environment (from mammalian host to mosquito vector) and are activated to form gametes: Female and male gametocytes both exit the RBCs, and female gametocytes develop into the macrogamete by releasing messenger RNAs (mRNAs) that were stored in a messenger ribonucleoprotein (mRNP) complex for translation<sup>2,3</sup>. The male gametocyte, on the other hand, undergoes three rapid rounds of endomitosis and forms eight flagellated microgametes, a process called exflagellation<sup>4</sup>. After fertilization of the macrogamete by the microgamete, the zygote embarks on a meiotic endoreplication cycle before traversing the mosquito midgut epithelium in the form of an ookinete that upon arrival at the midgut basal side transforms into an oocyst<sup>5</sup>. Over two weeks, endomitotic replication in the oocyst produces hundreds of sporozoites that, upon oocyst rupture, travel to the mosquito salivary glands, ready for inoculation into the vertebrate host with the next mosquito bite.

Epigenetic regulation is crucial for parasite survival within the human host<sup>6</sup>. Genes involved in host-parasite interactions or coding for virulence factors or ligands involved in RBC invasion are epigenetically regulated<sup>7,8</sup>, while some genes involved in drug resistance are epigenetically switched on or off in an environment-dependent manner<sup>9</sup>.

<sup>1</sup>Department of Life Sciences, Imperial College London, SW7 2AZ, London, UK. <sup>2</sup>Department of Molecular Biology, Radboud University, 6525, GA, Nijmegen, The Netherlands. <sup>3</sup>Present address: Institute of Medical Genetics and Applied Genomics, University of Tübingen, 72076, Tübingen, Germany. \*email: [kathrin.witmer@gmail.com](mailto:kathrin.witmer@gmail.com); [g.christophides@imperial.ac.uk](mailto:g.christophides@imperial.ac.uk)

Transcriptionally silent heterochromatin in *P. falciparum* is defined as the presence of tri-methylated histone 3 lysine 9 (H3K9me3) which is bound by HP1 (*Pf*HP1)<sup>7,10</sup>. Heterochromatin is largely confined to telomeric and subtelomeric regions as well as chromosome-central islands and is almost invariably associated with variably expressed multigene protein families in *P. falciparum*, *P. vivax*, *P. chabaudi*, *P. berghei*, *P. yoelii* and *P. knowlesi* asexual blood stage parasites<sup>7,10–14</sup>, *P. falciparum* oocysts<sup>15</sup> and *P. falciparum* and *P. vivax* sporozoites<sup>15–17</sup>. In *P. falciparum* gametocytes, heterochromatin domains expand into previously euchromatic regions harbouring genes encoding RBC remodeling proteins<sup>14,18</sup>, silencing genes that are used for asexual blood stage development. Euchromatic marks, on the other hand, dominate the *P. falciparum* genome: Acetylated histone 3 lysine 9 (H3K9ac) is the most investigated euchromatic mark to date and marks intergenic regions<sup>11</sup>. Its presence at promoter regions is a reliable predictor of gene expression in *P. falciparum* asexual blood stages<sup>13</sup> and oocysts<sup>15</sup>, and *P. falciparum* and *P. vivax* sporozoites<sup>15–17</sup>. H3K4me3 is another euchromatic mark in *P. falciparum*, and while associated with active genes, levels of occupancy do not follow changes in gene expression<sup>11–13,15</sup>.

The binding of transcription factors to promoter regions is heavily dependent on the state of the surrounding chromatin. Members of the apicomplexan-specific ApiAP2 family of transcription factors are found to control major cell fate decision events in the parasite lifecycle, in addition to housekeeping processes<sup>19–24</sup>. AP2-G is the master regulator of gametocytogenesis<sup>23,24</sup>, activating a number of gametocyte-specific genes<sup>25</sup>. Similarly, the ookinete-specific AP2-O, which is itself regulated by the mRNP complex, activates transcription of over 400 genes needed for ookinete development and mosquito midgut traversal<sup>20,26</sup>. Three additional ookinete-specific ApiAP2 transcription factors have been identified, which play a role just before or after ookinete formation<sup>19</sup>.

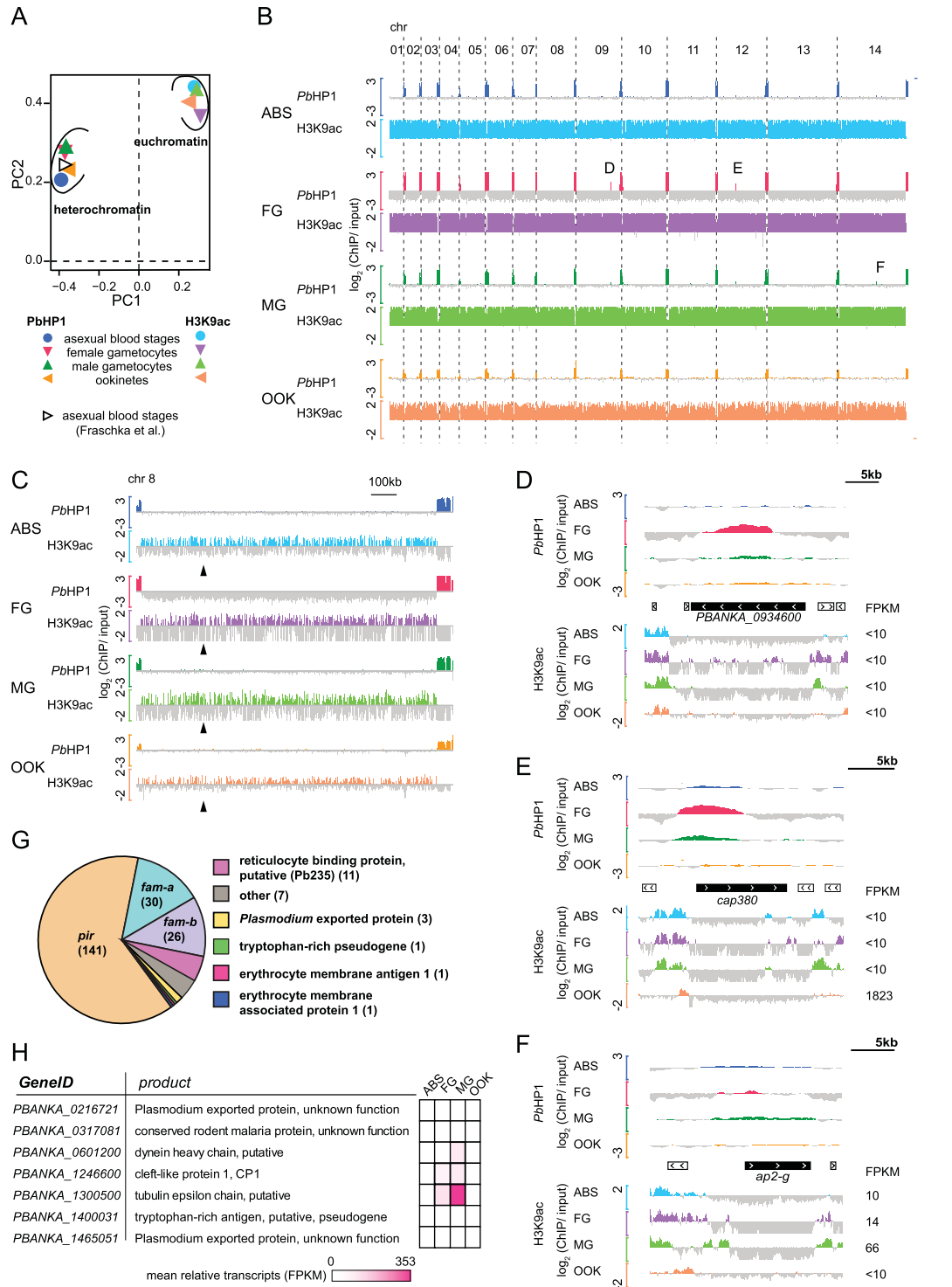
Here, we investigate how epigenetic traits change in malaria parasites during their transition from the murine host to the mosquito vector, using *P. berghei* asexual blood stages (ABS), female (FG) and male (MG) gametocytes, and ookinetes (OOK). We confirm that heterochromatin distribution is confined to subtelomeric regions in ABS in *P. berghei*<sup>14</sup>. We map for the first time the heterochromatin distribution in FG, MG and OOK in any *Plasmodium* spp. and find that heterochromatin distribution remains unaltered through parasite development and between *P. berghei* lines. We find heterochromatin occupancy at only two chromosome-central genes, namely the oocyst capsule protein Cap380 and a conserved protein of unknown function (*PBANKA\_0934600*). Implementing transcriptomics, we establish that variant transcription of multigene family genes occurs in *P. berghei* ABS, similar to *P. falciparum*<sup>27</sup>, albeit at lower levels. Additionally, we find evidence for variant multigene family transcription in MG, FG and OOK. In stark contrast to heterochromatin, the genomic distribution of H3K9ac is dynamic through parasite development: We find that while H3K9ac associates with 5' untranslated regions (5' UTRs) of genes, it specifically peaks around the start codon of ribosomal protein genes, a phenomenon previously not reported in any *Plasmodium* species. Consistent with previous findings in *P. falciparum* asexual blood stages<sup>13</sup>, H3K9ac enrichment in 5'UTRs correlates with transcript abundance in ABS in *P. berghei*. Additionally, we show for the first time that H3K9ac is a good indicator for transcript abundance in two more malaria parasite life cycle stages, namely MG and OOK. Surprisingly, we do not find a positive correlation between H3K9ac occupancy and transcript abundance in FG, suggesting a different epigenetic state of FG compared to other developmental stages. Finally, we identify four novel DNA motifs in the 5'UTR of OOK-specific genes, suggesting that transcription factors in addition to the known AP2-Os are involved in orchestrating transcription in OOK. Our study adds substantially to our understanding of epigenetic regulation of gene expression during parasite transition from the mammalian host to the mosquito vector.

## Results

**Generation of epigenetic and transcriptional profiles.** To investigate chromatin changes during *P. berghei* development, we performed chromatin immunoprecipitation (ChIP) using antibodies against *P. berghei* HP1 (*Pb*HP1) as a marker for heterochromatin<sup>14</sup> and against H3K9ac as a marker for euchromatin. Since both *P. falciparum* and *P. berghei* histones 3 (H3) show 100% sequence conservation, we decided to use the H3K9ac antibody that has previously been used in *P. falciparum* ChIP<sup>7,13</sup> in our study.

To keep samples as pure as possible and minimize contamination with another developmental stage we used three different parasite lines, one for each developmental stage, for the following reasons: Asexual blood stages (ABS) were sampled from the non-gametocyte producer PbANKA 2.33 parasite clone<sup>28</sup>, to rule out potential contamination with gametocytes. Male gametocytes (MG) and female gametocytes (FG) were isolated from the well-defined *820cl1m1cl1* (820) line<sup>29</sup>. The 820 line produces green fluorescent male gametocytes and red fluorescent female gametocytes, which can be easily sorted by flow cytometry<sup>25,30</sup>. As PbANKA 2.33 cannot produce ookinetes, and the 820 line is rarely used to produce ookinetes, we decided to use the well-established *507m6cl1* line<sup>31</sup> for our ookinete sample (OOK). OOK were prepared from *in vitro* cultures 24 hours post gametocyte activation and subsequently purified using an antibody against the ookinete surface protein P28<sup>32</sup>. Chromatin was crosslinked and sheared before performing chromatin immunoprecipitation (ChIP) using either an anti-*Pb*HP1 or anti-H3K9ac antibody. The immuno-precipitated DNA and unprecipitated input as a control for each stage were purified, amplified and subjected to next-generation sequencing (NGS). To compare the ChIPseq dataset to transcript abundance, we harvested RNA in biological triplicates from the respective stages from the lines used for ABS, MG, FG and OOK ChIP to perform RNAseq. To control for any amplification bias due to the AT-rich *P. berghei* genome we included sheared genomic DNA for the RNAseq samples. A GC-plot confirms equal amplification of all regions (Fig. S1A). A summary of mapped reads for both ChIPseq and RNAseq samples is shown in Table S1.

To get a first overview of our ChIP and RNAseq data, we performed principal component analysis (PCA) for each data set. Immunoprecipitated DNA of all the developmental stages clusters according to the antibody used in ChIP, suggesting that euchromatin and heterochromatin occupancies differ from each other but are similar between the different developmental stages and parasite lines (Fig. 1A). PCA of the RNAseq samples showed that transcription profiles of ABS and MG are somewhat related to each other and transcription profiles of FG are more related to OOK (Fig. S1B), as has been seen before<sup>33</sup>.



**Figure 1.** Heterochromatin distribution remains stable during malaria parasite development whereas euchromatin distribution is dynamic. **(A)** Principal component analysis of log<sub>2</sub> transformed ChIP over input data. Heterochromatin and euchromatin cluster away from each other. Data from ABS from a previous heterochromatin study is included<sup>14</sup>. **(B)** Screenshot of heterochromatin and euchromatin distribution across all 14 *P. berghei* chromosomes for each developmental stage. Peaks correspond to log<sub>2</sub> transformed data of either PbHP1-ChIP/input or H3K9ac-ChIP/input. Log<sub>2</sub> scale for PbHP1 is (-3 to 3) and (-2 to 2) for H3K9ac, respectively. Letters highlight chromosome-central genes showing enrichment in PbHP1 binding. Their close-up view is shown in **(D-F)**. **(C)** Bar chart close-up view of heterochromatin and euchromatin occupancy in chromosome 8. The approximate location of the centromere (syntenically inferred from *P. falciparum*<sup>34</sup>) is indicated with arrowheads. **(D,E)** Chromosome-central genes associated with heterochromatin in this study. **(F)** *ap2-G* does not qualify as heterochromatin in our study. Black boxes show gene of interest, white boxes indicate neighbouring genes and chevrons indicate the transcriptional orientation for each gene. Peaks correspond to

$\log_2$ -transformed data of *PbHP1*-ChIP/input (−3 to 3) and H3K9ac-ChIP/input (−2 to 2). (G) Pie chart of identified heterochromatic genes (including pseudogenes) grouped into gene families as described previously<sup>36</sup>. The number of genes in each family is shown in parentheses. (H) Description of seven heterochromatic genes not belonging to known multigene families. Colour bars indicate mean transcripts given as FPKM (kilobase of transcript per million mapped reads) of three independent biological replicates. ABS, asexual blood stages; FG, female gametocytes; MG, male gametocytes; OOK, ookinetes.

**Mutually exclusive profiles of heterochromatin and euchromatin.** To identify genomic regions that are associated with hetero- and euchromatin during *P. berghei* development, we visualized ChIPed chromatin over input using bar plots. This approach shows that heterochromatin is confined to telomeric and subtelomeric regions of all chromosomes, except for the right arm of chromosome 4, the only subtelomeric region devoid of multigene families (Fig. 1B). The euchromatic mark H3K9ac, on the other hand, is detected in all other chromosomal regions (Fig. 1B). In contrast to most other *Plasmodium* species<sup>14</sup>, *P. berghei* chromosomes are largely devoid of chromosome internal heterochromatin islands and remain so during development (Fig. 1B). As shown in the example of chromosome 8 (Fig. 1C), but observed for all chromosomes and all stages, centromeres do not show enrichment in heterochromatic marks<sup>34</sup>.

To identify heterochromatic genes we calculated the mean *PbHP1* occupancy ( $\log_2$  ratio of *PbHP1* ChIP to input) of the gene open reading frame (ORF) and performed hierarchical clustering of the ORF for all genes and all stages together. Using this approach, we found that only two non-telomeric and non-subtelomeric genomic genes showed clear heterochromatic profiles. The first gene identified using this approach is *PBANKA\_0934600*, a gene encoding a large, conserved protein of unknown function with orthologues in all *Plasmodium* species (Fig. 1D). Transcript levels of *PBANKA\_0934600* are very low (Fragments Per Kilobase of transcript per Million mapped reads (FPKM) < 10) suggesting that the gene is not expressed in any of the stages investigated here (Table S2), or indeed only expressed in a subset of cells. The latter seems to be the case, since single cell sequencing data shows that *PBANKA\_0934600* is clonally variably expressed throughout the life cycle<sup>33</sup>, explaining its heterochromatin status seen here. However, *PBANKA\_0934600* is redundant for *P. berghei* transmission<sup>35</sup> and therefore the role of this gene and why it is variably expressed remains unclear.

The second chromosome-central heterochromatic region corresponds to *cap380* and appears to be epigenetically silenced in ABS<sup>14</sup>, FG and MG but not in OOK (Fig. 1E). Indeed, OOK display increased levels of relative *cap380* transcripts, which coincides with a more condensed H3K9ac occupancy within the *cap380* 5'UTR, however, it is important to note that H3K9ac occupancy in the 5'UTR of *cap380* is present in all stages investigated here. *Cap380* transcription is controlled by AP2-O<sup>26,36</sup> and the protein is expressed in early-stage oocysts and localizes to the oocyst capsule<sup>37,38</sup>. Our data provide evidence that heterochromatic silencing is an additional regulatory level of *cap380* expression, presumably preventing its premature transcription; as H3K9ac occupancy alone is no clear indication of transcriptional abundance for this gene.

A third region that showed heterochromatic marks by visual inspection but did not classify as heterochromatic using our analysis algorithm encompasses the gene encoding AP2-G (*PBANKA\_1437500*; Fig. 1F). In asexual blood stages in *P. falciparum* AP2-G is epigenetically silenced by *PfHP1*, and *PfHP1* eviction leads to activation of this gene<sup>39</sup>. Here, weak levels of *PbHP1* marking AP2-G in ABS can be explained by our choice of the *PbANKA* 2.33 parasite line. This 2.33 line carries a mutation in the *ap2-g* gene resulting in expression of a truncated, non-functional protein unable to induce gametocytogenesis<sup>24</sup>, thereby making its epigenetic silencing redundant.

**Epigenetic silencing of subtelomeric multigene families.** Next, we asked which genes are epigenetically silenced by *PbHP1*. Our analysis revealed that as many as 223 genes (including 61 pseudogenes) located in subtelomeric regions are significantly enriched in *PbHP1* binding (Fig. 1G, Table S2). Of these, 214 belong to the rodent malaria parasite (RMP) multigene family<sup>36</sup>. It has been previously established that *Plasmodium* heterochromatin is largely associated with gene families involved in antigenic variation and host-parasite interactions<sup>7,14–16,18</sup> and our results show that the same is true for *P. berghei* gametocytes and ookinetes.

Only seven subtelomeric heterochromatic genes do not belong to RMP multigene families (Fig. 1H). They include: *PBANKA\_1246600* that encodes CP1, an atypical PEXEL protein exported to discrete structures in the cytosol of infected RBCs in asexual blood stages<sup>40</sup>; *PBANKA\_0216721* and *PBANKA\_1465051* both encoding exported proteins of unknown function; *PBANKA\_0317081* encoding a conserved rodent malaria parasite protein; *PBANKA\_1300500* encoding a putative tubulin epsilon chain; *PBANKA\_1400031*, a putative pseudogene encoding a tryptophan-rich antigen; and *PBANKA\_0601200* encoding a dynein heavy chain.

*PBANKA\_0601200* is one of 22 dynein-related proteins annotated in the *P. berghei* genome (Fig. S2). Dyneins are one of three cytoskeletal motor protein families in eukaryotes and made up of a protein complex of heavy, light and intermediate chain. Out of six dynein genes located close to telomeres only *PBANKA\_0601200* is heterochromatic in all developmental stages in this study. However, it is not orthologous to either of two heterochromatic dynein heavy chain-encoding genes found in *P. falciparum* asexual blood stages<sup>7</sup>. *PBANKA\_0601200* is only transcribed in MG<sup>33</sup> (Fig. S2), nonetheless, its 5'UTR is occupied by H3K9ac in both male and female gametocytes, suggesting an additional layer of control. Its knockout is associated with slow ABS growth<sup>41</sup>.

It has previously been shown that heterochromatin boundaries expand towards the centromere in *P. falciparum* late stage gametocytes<sup>14,18</sup>, silencing genes involved in invasion and RBC remodelling. In contrast, we did not find any expansion of heterochromatin boundaries in either gametocyte stage or ookinetes. This highlights differences between the two *Plasmodium* species, and can be explained by the absence of the genes found on



at the right arm of chromosome 7 display euchromatic traits correlating with transcription. Relative transcripts (in FPKM) for each gene are shown in numbers above ChIP-seq tracks. Black boxes show gene of interest, white boxes indicate neighbouring genes and arrows indicate the transcriptional orientation for each gene. Peaks correspond to  $\log_2$ -transformed data of *PbHP1*-ChIP/input (−3 to 3) and H3K9ac-ChIP/input (−2 to 2). ABS, asexual blood stages; FG, female gametocytes; MG, male gametocytes; OOK, ookinetes; pg, pseudogenes.

chromosome 2 in *P. berghei*. In summary, our data show that similar to other *Plasmodium* species<sup>14,18</sup>, subtelomeric multigene families are epigenetically silenced via *PbHP1* throughout *P. berghei* sexual development.

**Variant developmental expression of heterochromatic genes.** Stochastic changes in heterochromatin distribution result in clonal variant gene expression, which forms the basis of *P. falciparum* antigenic variation<sup>6</sup>. In *P. berghei*, members of multigene families show variable expression independent of life cycle stage in single-cell transcriptomic data<sup>33</sup>. Since different parasite lines were used in our study (except FG and MG samples that derived from the same line), we could not directly determine whether differences in heterochromatin distribution are due to either the epigenetic background of the parasite lines or the developmental stage. Therefore, our comparative heterochromatic profiling would capture the sum of both. We compared the heterochromatic profiles of genes across developmental stages, and found a surprisingly small number of 16 genes with differential heterochromatin occupancy between stages/strains (Fig. 2A). This small number stands in sharp contrast to the 252 genes shown to exhibit clonal variant expression in *P. falciparum* ABS alone<sup>27</sup>. Thirteen of the 16 genes belonged to multigene families, and 12 showed *PbHP1* enrichment in FG but not in MG. Since both gametocyte samples were derived from the same line<sup>29</sup>, these data indicate true developmental epigenetic differences that in turn imply that heterochromatin occupancy is reorganized during gametocyte development. However, transcript abundance of these 12 genes was very low (<10 FPKM) in all lifecycle stages examined (Table S2), suggesting that absence of *PbHP1* occupancy may not necessarily result in an active transcriptional state.

To identify if variegated clonal expression is present in the investigated *P. berghei* life cycle stages, we arbitrarily selected all heterochromatic genes with more than 10 FPKM (mean of three replicates) in at least one developmental stage. Using this approach, we identified an additional 35 genes in ABS, 17 genes in FG, 37 genes in MG and 8 genes in OOK, respectively, which are possibly variably expressed, as they show heterochromatic traits at the same time as transcripts (Fig. 2B). Most of these heterochromatic genes display additional euchromatic H3K9ac marks, mainly in their 5'UTR and some of them show very high transcript abundance (Fig. 2B). We compared all of these genes to the Malaria Cell Atlas data<sup>33</sup> and confirmed variant expression for almost all of them in the stages investigated (Fig. 2B).

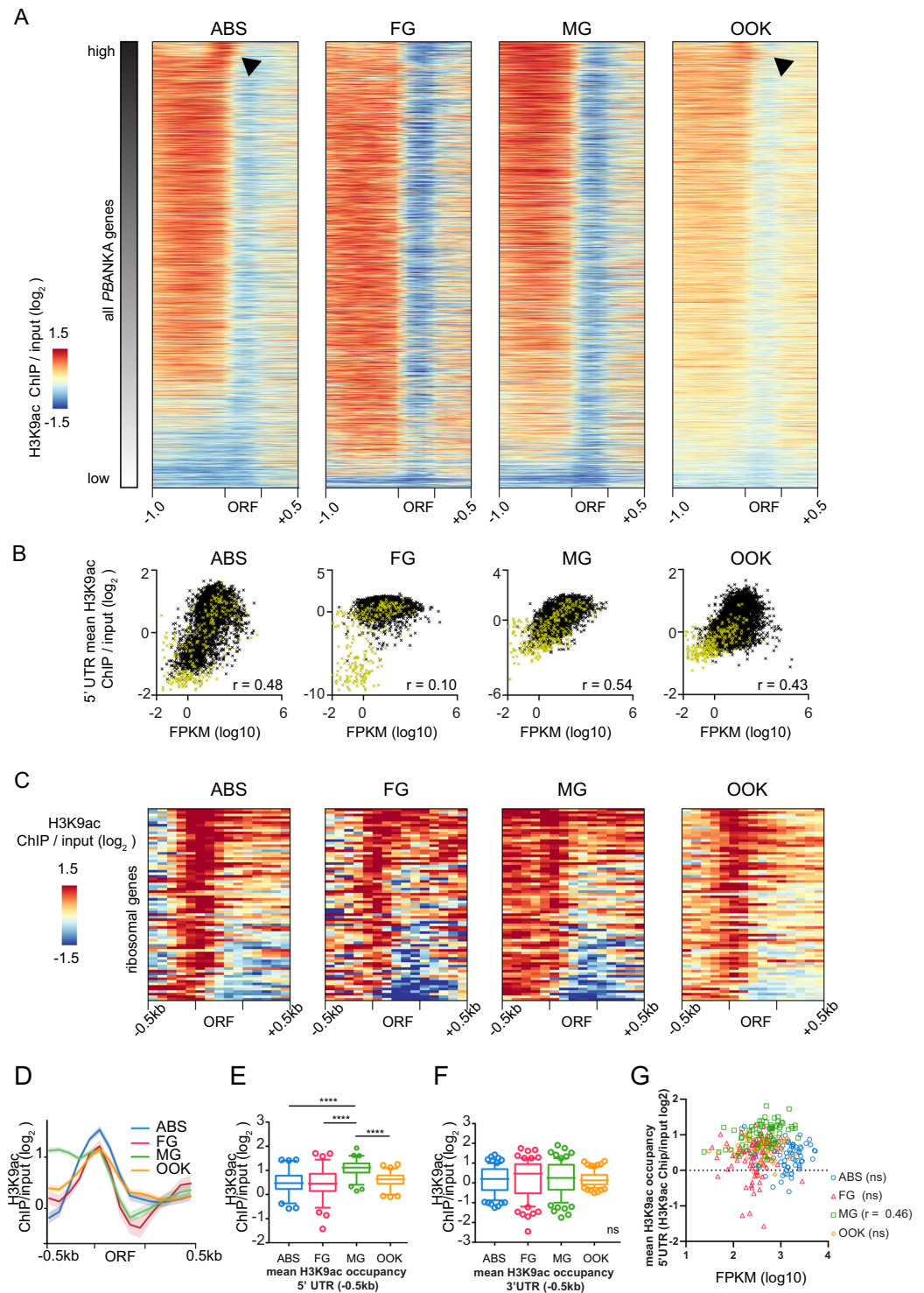
For example, despite the *PBANKA\_0300600* ORF being heterochromatic, its 5'UTR is enriched in H3K9ac and the gene is highly expressed in all stages (Fig. 2C). These findings suggest that *PBANKA\_0300600* is expressed in only a subset of cells. Indeed, single cell transcriptomic data<sup>33</sup> shows that *PBANKA\_0300600* is clonally variably expressed in asexual blood stage parasites, male and female gametocytes and ookinetes (Fig. S3A). Similarly, epsilon-tubulin (*PBANKA\_1300500*) is heterochromatic in all stages but shows euchromatic traits and high transcript abundance in both MG and FG (Fig. 2D). Again, compared to the single cell RNA transcriptomic data from Howick *et al.*<sup>33</sup>, epsilon tubulin is expressed in almost all male gametocytes, but shows variant expression in female gametocytes (Fig. S3B) as well as male gametocytes. Epsilon-tubulin protein marks the older of the two human centrioles upon centrosome duplication<sup>42</sup> and is an essential part of the basal bodies in *Tetrahymena*<sup>43</sup>, but its function is unknown in *Plasmodium* species.

Another example of clonally variegated expression includes multigene family members on the right arm of chromosome 7, displaying both heterochromatin and euchromatin occupancies and high transcript abundance (Fig. 2E). This is consistent with the finding that some heterochromatic genes are transcribed by a subset of cells, effectively displaying variegated expression as shown before for members of putative exported protein families in *P. berghei*<sup>44,45</sup>.

We show that genes displaying hallmarks of clonally variant expression are located at the heterochromatin-euchromatin boundaries, which may facilitate clonal variation. In conclusion, we show that bulk ChIP data is able to predict variably expressed genes defined as having both euchromatin and heterochromatin marks, as shown by our comparison to the Malaria Cell Atlas<sup>33</sup>.

**Distinct H3K9ac distribution in ribosomal protein genes.** H3K9ac is a universal histone mark associated with active promoters, including *P. falciparum* asexual blood stages, oocysts and sporozoites<sup>11,15,16</sup>. We investigated the relationship between H3K9ac distribution and gene transcription across the four *P. berghei* developmental stages of this study. H3K9ac enrichment in the gene ORF, 1 kb 5'UTR and 500 bp 3'UTR was examined against transcript abundance for each stage (Fig. 3A). Consistent with previous findings in *P. falciparum*, a positive correlation was detected between the gene 5'UTR H3K9ac enrichment and transcript levels in ABS<sup>11,12</sup>. In addition, we found that H3K9ac enrichment in the 5'UTR positively correlates with transcript abundance in MG and OOK but less so in FG (Fig. 3B), which will be investigated in more detail in the next chapter.

A shift of the peak of H3K9ac occupancy toward the gene ORF was observed for highly expressed genes, mainly in ABS and OOK (see arrowheads in Fig. 3A). Many of the genes with high expression encode ribosomal proteins. Closer investigation found that H3K9ac occupancy sharply peaks around the translation start codon of ribosomal protein genes in ABS, FG and OOK (Fig. 3C). Intriguingly, in MG, H3K9ac occupancy in these genes was extended further into the 5'UTR, resulting in no clear peak being detected around the start codon (Fig. 3D). Further analysis confirmed that the mean H3K9ac enrichment in the 500 bp upstream of the ATG (or translation



**Figure 3.** 5' UTRs of ribosomal protein genes display a distinctive H3K9ac pattern. (A) Heat map of H3K9ac distribution in malaria parasite development. Genes at each stage are sorted according to their relative transcription levels, with highly expressed genes on top. Arrowheads indicate a shift of H3K9ac occupancy towards the start codon. H3K9ac enrichment for each gene locus is shown as  $\log_2$ -transformed H3K9ac ChIP/input (1000 bp upstream of ATG the ORF, and 500 bp downstream of stop codon, respectively). (B) Scatterplot showing mean H3K9ac occupancy of 5' UTRs (1000 bp) against transcriptional strength (mean FPKM) for each gene for each developmental stage. Spearman's rank correlation coefficient for each scatterplot is shown ( $r$ ). Genes belonging to a multigene family are highlighted in yellow. (C) Heat map of H3K9ac enrichment for all 74 ribosomal protein genes (40 S and 60 S). H3K9ac is enriched in a sharp peak around the start codon in ABS, FG and OOK, but less in MG. (D) Summary plot of the data shown in C. The mean of H3K9ac enrichment of ribosomal protein genes is shown as a bold line for each life cycle stage, standard error is shown in lighter colours. (E) Boxplots show mean H3K9ac-enrichment values for 5' UTRs (500 bp) of ribosomal protein genes

for each stage. Whiskers indicate the 5<sup>th</sup> and 95<sup>th</sup> percentile, respectively. Individual symbols represent outliers. Asterisks mark significance (Wilcoxon signed rank test for matched-pairs,  $p \leq 0.0001$ ). (F) Boxplots show mean H3K9ac-enrichment values for 3'UTR of each ribosomal protein gene. Whiskers indicate the 5<sup>th</sup> and 95<sup>th</sup> percentile, respectively. Individual symbols represent outliers. Wilcoxon signed rank test for matched-pairs found no difference between the developmental stages (ns). (G) Scatter plot showing the correlation of the mean ribosomal protein gene transcript abundance (x-axis) against the mean H3K9ac enrichment value of the 5'UTR (500bp) (y-axis) of the same gene. The Spearman's rank correlation coefficient is shown (r). ABS, asexual blood stage; FG, female gametocyte; MG, male gametocyte; OOK, ookinete; ORF, open reading frame.

start) of ribosomal protein genes was significantly different between MG and all other stages analysed (Wilcoxon signed rank test,  $p \leq 0.0001$ ), but no difference was detected between any of the other stages (Fig. 3E). As a control, the same analysis using the first 500 bp of 3'UTR sequence of these genes did not show any significant differences between any of the four stages (Wilcoxon signed rank test,  $p > 0.3$ ) (Fig. 3F). This clearly shows that genes encoding ribosomal proteins exhibit different H3K9ac occupancy in comparison to all other genes in ABS, FG and OOK. Additionally, the H3K9ac intensity in the 5'UTR of ribosomal genes does not correlate with transcript abundance in ABS, FG and OOK (Fig. 3G). This also means that H3K9ac occupancy of ribosomal protein genes in MG does indeed correlate with transcript abundance.

H3K9ac occupancy is predominantly enriched around the transcriptional start site (TSS) in *P. falciparum* asexual blood stages<sup>46</sup>, thus we were interested to see if the shift in H3K9ac in ribosomal protein genes observed here can be explained by their TSSs lying closer to the ATG. As TSS are only mapped for *P. falciparum* but not *P. berghei*, we relied on the dataset of Adjalley and colleagues. We compared the distances from the middle of the main TSS to the ATG for ribosomal protein genes to the rest of the protein coding genes. While we found a trend for a TSS to be closer to the ATG in ribosomal protein genes compared to the rest of protein coding genes in *P. falciparum*, this was not statistically significant (unpaired t-test,  $P > 0.1$ ) (Fig. S4).

Taken together, these data show that a sharp peak around the start codon is no prediction of transcript abundance in ribosomal protein genes for ABS, FG and OOK, and that this peak does not correlate with the location of the TSS. The data further show that for ribosomal protein genes, male gametocytes exhibit a different H3K9ac occupancy pattern compared to the other developmental stages investigated here and that H3K9ac intensity correlates with transcript abundance. It has been previously suggested that a shift in peak shape can indicate a different function of a gene<sup>47</sup>, and highly dense and narrow distributions of H3K9ac near TSSs have been associated with constitutive expression of genes involved in translation in plants<sup>48</sup>.

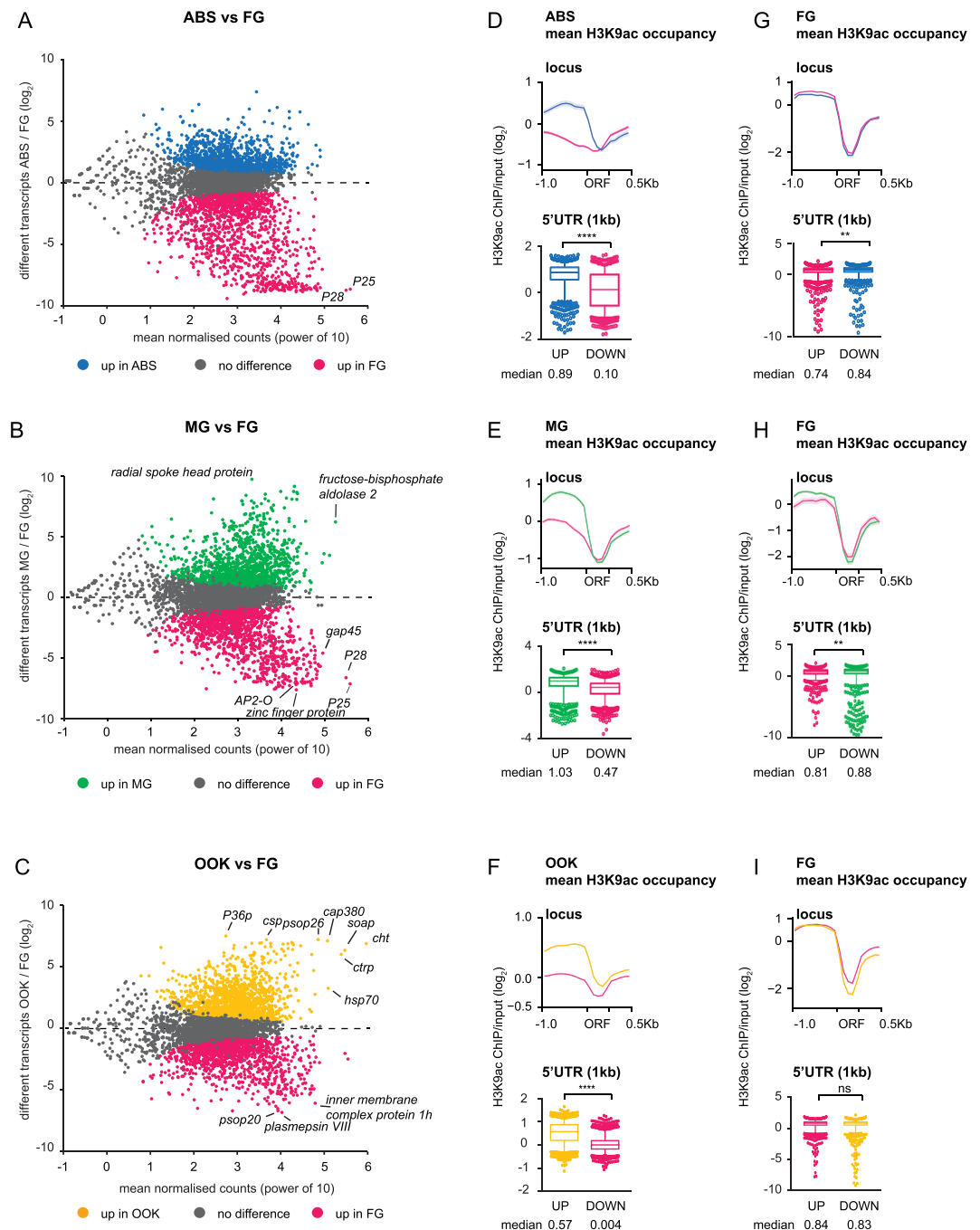
**H3K9ac enrichment does not positively correlate with transcript abundance in FG.** We further explored the relationship between H3K9ac occupancy and transcript abundance in *P. berghei* development. For this, we identified genes that are differentially expressed between two consecutive (ABS and FG, FG and OOK) or opposite (FG and MG) developmental stages and examined their H3K9ac occupancy for each of the two stages (Fig. 4A–C). Genes upregulated in ABS (1426), MG (1523) or OOK (1641) compared to FG exhibit greater H3K9ac enrichment in their 5'UTR (1000 bp upstream of the start codon) in each respective stage than genes downregulated in any of these three stages (1429, 1518 and 1534, respectively) compared to FG (Fig. 4D–F). These results indicate that H3K9ac occupancy in the gene 5'UTR is a good predictor for relative transcript abundance in ABS, MG and OOK. Surprisingly, when examining the same gene groups in FG, upregulated genes in FG showed less (FG vs. ABS and FG vs. MG) or the same (FG vs. OOK) H3K9ac enrichment than/as downregulated genes (Fig. 4G–I).

We further investigated the lack of positive correlation between H3K9ac occupancy and transcript abundance in FG by generating two high-confidence lists of genes that are either upregulated (591 genes) or downregulated (271 genes) in FG compared to all other stages, respectively (Fig. 5A and Table S3). Gene Ontology (GO) analysis using the GO slim version showed that the former list includes genes involved in cell adhesion, reproduction, motility, differentiation, cell cycle and locomotion, all of which are characteristic of cells preparing for fertilization and development into the motile ookinete stage (Fig. 5B and Table S4). In contrast, downregulated genes in FG are involved in ribosome biogenesis and translation, suggestive of a translationally less active FG stage compared to the other three stages. FG are known to produce and store mRNAs that are translationally repressed<sup>2</sup>. Similar to before, these downregulated FG genes show significantly higher H3K9ac enrichment in their 5'UTRs than genes that are upregulated in FG (Fig. 5C).

Together, these data indicate that H3K9ac, a histone mark previously associated with active promoters in *P. falciparum* asexual blood stages<sup>12,13</sup> and oocysts<sup>49</sup> as well as *P. falciparum*<sup>15,16</sup> and *P. vivax*<sup>17</sup> sporozoites, also associates with transcript abundance in *P. berghei* ABS, MG and OOK. Additionally, we show that H3K9ac occupancy does not directly correlate with transcript abundance in FG in *P. berghei*. Instead, it appears that H3K9ac marks the promoters of most genes in FG, irrespective of their transcript abundance.

Post-transcriptional regulation is known to occur in female gametocytes, where transcripts are stabilized by RNA binding proteins DOZI and CITH, and are poised for translation<sup>2,3,50</sup>. Therefore, large amounts of maternal transcripts accumulate in the female gametocyte not necessarily reflecting transcriptional states of these genes. To test if the lack of positive correlation between H3K9ac occupancy and transcript abundance in FG may be caused by poised transcripts, we compared the H3K9ac occupancy of translationally repressed genes to the sum of all genes (Fig. S5A). We find that mRNA of genes that are translationally repressed by the DOZI-complex show lower H3K9ac occupancy compared to all euchromatic genes (all genes minus multigene family genes, minus *PbHP1* occupied genes) (Fig. S5B), confirming that high transcript abundance due to storage does not positively correlate with H3K9ac occupancy in FG.





**Figure 4.** H3K9ac intensity correlates with relative transcripts in asexual blood stages, male gametocytes and ookinetes. (A–C) Comparative profiling of ABS, MG and OOK vs FG. Each dot represents a gene. Genes with a non-significant change in transcript abundance are shown in grey and genes that are significantly up- or downregulated are shown in their respective colour (p value adjusted for multiple testing with the Benjamini-Hochberg procedure which controls false discovery rate (FDR), here  $p < 0.001$ ). (D) Mean H3K9ac distribution correlates with transcript levels in asexual blood stages. The upper panel shows the mean H3K9ac distribution for all genes that are either up- or downregulated in ABS vs FG. The mean H3K9ac enrichment of all loci is indicated with a solid line, and the standard error is shown in a lighter colour. 1000 bp upstream and 500 bp downstream of each gene is included. Boxplots show mean H3K9ac enrichment for the 5' UTR for both gene groups. In ABS, (compared to FG), mean H3K9ac enrichment in the 5' UTR of a gene positively correlates with its transcript levels (Mann Whitney test,  $p < 0.0001$ ). Boxes extend from the 25th to 75th percentiles, and the median is shown as a line in the middle of the box and as a number below the boxes. Whiskers indicate the 5<sup>th</sup> and 95<sup>th</sup> percentile, respectively. Individual symbols represent outliers. The median is shown. (E) In male gametocytes (compared to FG), mean H3K9ac enrichment in the 5' UTR of a gene positively correlates with its transcript levels (Mann Whitney test,  $p < 0.0001$ ). Labelling is the same as in D. (F) In ookinetes (compared to FG), mean H3K9ac enrichment in the 5' UTR of a gene positively correlates with its transcript levels (Mann Whitney test,  $p < 0.0001$ ). Labelling is the same as in D. (G) Mean H3K9ac distribution of in female

gametocytes using the same gene groups as in A and D. Genes that are upregulated in FG (compared to ABS) are less enriched for H3K9ac in their 5'UTR of than downregulated genes in FG (Mann Whitney test,  $p = 0.0016$ ). (H) Mean H3K9ac distribution of in female gametocytes using the same gene groups as in B and E. Genes that are upregulated in FG (compared to MG) are less enriched for H3K9ac in their 5'UTR of than downregulated genes in FG (Mann Whitney test,  $p = 0.0024$ ). (I) Mean H3K9ac distribution of in female gametocytes using the same gene groups as in C and F. Genes that are upregulated and downregulated in FG (compared to OOK) show the same H3K9ac enrichment in FG (Mann Whitney test, non-significant (ns),  $p = 0.57$ ).

After fertilization, ookinete-specific transcripts are transcribed *de novo*, activated by the ookinete AP2 transcription factor AP2-O<sup>20,26</sup>. To test if H3K9ac enrichment in FG predicts future transcriptional activity, we compared H3K9ac occupancy of AP2-O regulated genes to the sum of all genes (Fig. S5C). We found that H3K9ac enrichment in FG is unrelated to future transcriptional activity, as AP2-O-regulated genes display the same H3K9ac enrichment compared to all euchromatic genes in the *P. berghei* genome (Fig. S5D). Taken together, these findings indicate a H3K9 hyperacetylated genome in FG where H3K9ac does not predict transcript abundance. Further experiments will be needed to determine if and which histone modifications are predictive of transcript abundance in FG.

**Novel and known DNA motifs control ookinete gene expression.** In most metazoans as well as in flowering plants, maternally deposited proteins and mRNAs are responsible for directing early stages of development post fertilization, while *de novo* transcription is resumed via specific transcription factors, a phase called maternal-to-zygotic transition (MZT). In *Plasmodium*, maturation of the fertilized female gamete to the ookinete is dependent upon both de-repression of maternal mRNAs<sup>2,50</sup> and *de novo* transcription mediated by ookinete-specific transcription factors such as AP2-O<sup>19,20,26</sup>.

We further investigated the mechanisms of MZT in *P. berghei* starting with genes that are differentially upregulated in OOK compared to FG (Fig. 4C), and compared them to 464 genes previously annotated as controlled by AP2-O<sup>26</sup> (Fig. 6A). Of the 1641 ookinete-specific genes, only 302 qualified as AP2-O controlled (Fig. 6A), suggesting that not all AP2-O controlled genes have been annotated to date and/or that additional transcription factor(s) might be involved in ookinete-specific transcription. Therefore, we searched for DNA-motifs<sup>51</sup> that are significantly enriched in the 5'UTRs of all 1641 OOK-enriched genes and found two highly significantly enriched motifs: the AP2-O CTAGCT/CA motif<sup>20</sup> present in 265 genes, and the AAAAAAAAA motif found in as many as 1571 genes (Fig. 6A).

We repeated the same search using a more stringent list of 357 genes that are at least 8-fold upregulated in OOK compared to FG; and only 122 qualified as controlled by AP2-O (Fig. 6B). Four DNA-motifs were identified, the most significant of which again resembles the PfAP2-O motif (TAGCT/CAA/G) and was found in almost half of all genes (164). The other three novel motifs were CT/AGAACA (in 62 genes), A/GGAGAAAA (in 124 genes) and GTTGTAACA/C (in 61 genes).

These data confirm previous findings that AP2-O is a master regulator of *de novo* transcription and MZT in *Plasmodium* and indicate the possibility of the existence of additional transcription factors involved in this process. It is important to note that while additional OOK transcription factors have been described<sup>19,33</sup>, none of the published DNA-motifs matches those found here.

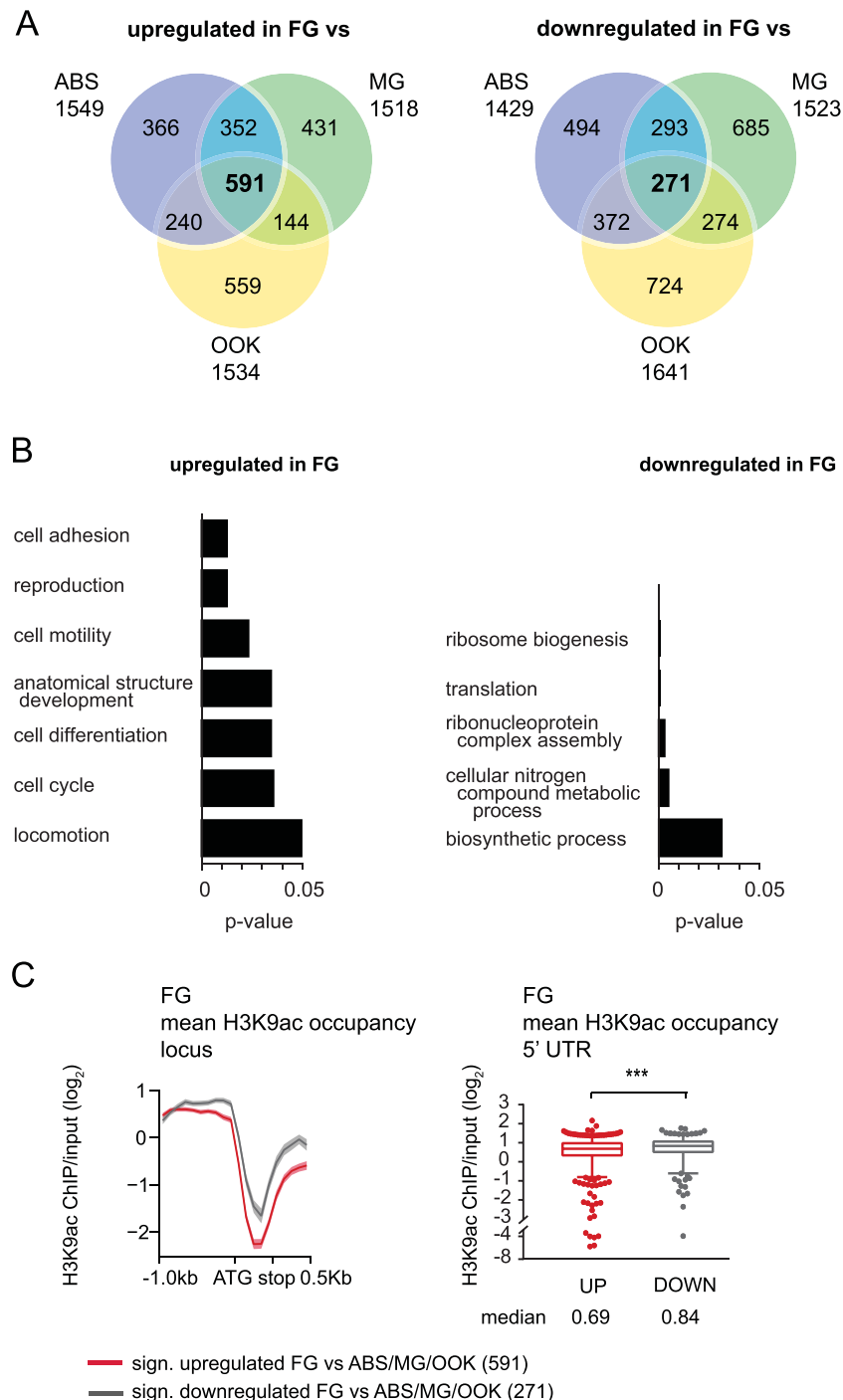
## Discussion

Our study provides new insights into the epigenetic regulation of *Plasmodium* gene expression during its developmental transition from asexual to sexual and zygotic stages, which occurs as the parasite moves from its vertebrate host to the insect vector and is essential for parasite transmission.

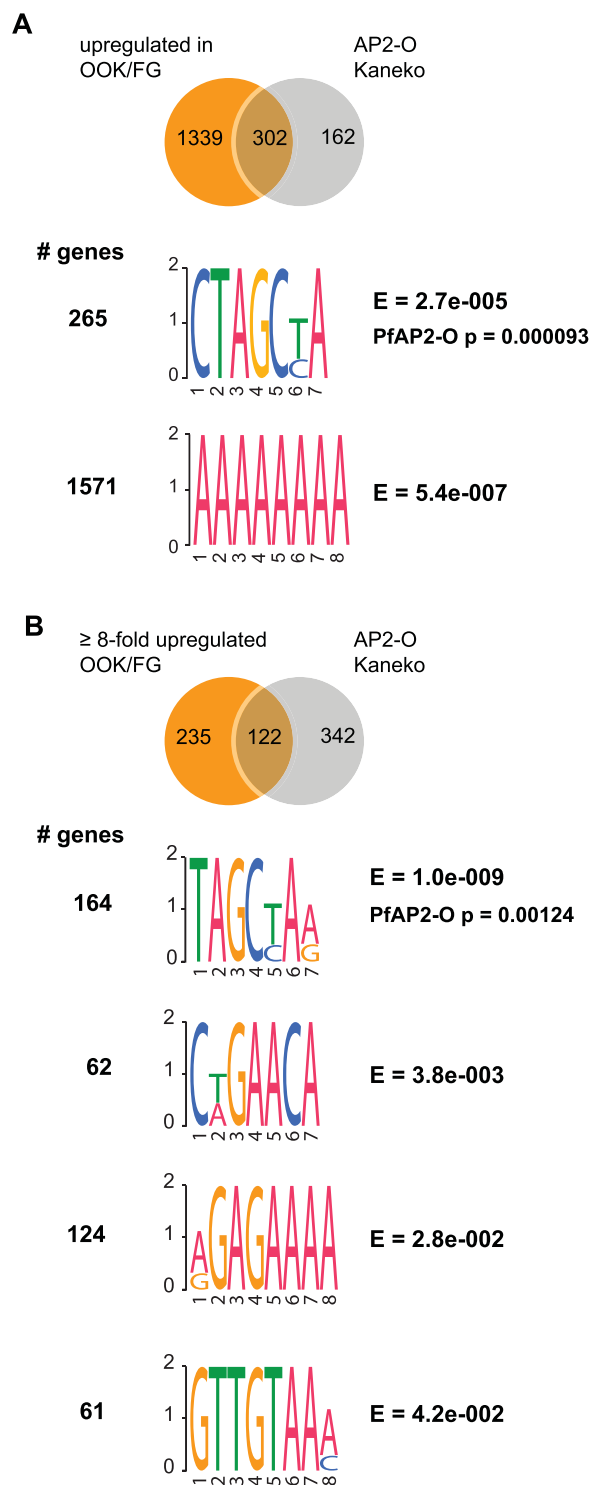
Epigenetic silencing has been associated with key strategies of parasite development and environmental adaptations. Our data reveal that HP1-mediated epigenetic silencing remains largely unchanged during this transition and confined to chromosomal subtelomeric regions. Mapping *Plasmodium* subtelomeric genes is very challenging as most genes belong to multigene families with high levels of sequence identity between them. Indeed, as many as 51 heterochromatic genes could not be mapped to *P. berghei* chromosomes in our analysis: 14 *fam-a*, 8 *fam-b*, 1 *fam-c*, 27 *pir*, and one encoding a tryptophan-rich antigen.

It is important to note that the study by Fraschka *et al.*<sup>14</sup> analyzed *PbHP1* occupancy in *P. berghei* ANKA asexual blood stages. Their strain is the parental strain of the PBANKA 2.33 line used in this study. Comparison between the two studies showed that 177 heterochromatic genes are shared between the two studies, 46 are specific to our study and 14 are specific to the study of Fraschka *et al.* (Fig. S6A). Whereas the vast majority of differences pertain to subtelomeric regions, where our analysis appears to be more sensitive, the combination of the two studies identified four non-subtelomeric heterochromatic genes, of which only *cap380* encoding an oocyst capsule protein is detected in both studies. On the one hand, *PBANKA\_0934600* that encodes a protein of unknown function was only detected in our analysis but a closer visual inspection of the profiles obtained by Fraschka *et al.*<sup>14</sup> shows extensive occupancy of the gene by *PbHP1* (Fig. S6B). On the other hand, *ap2-g* and *ap2-sp3/ap2-tel* are not detected as heterochromatic in our analysis but visual inspection indicates increased *PbHP1* occupancy (Fig. S6C). These data suggest that automated detection of moderate heterochromatin enrichment remains a challenge and that conclusions about absence of epigenetic silencing in such regions should be treated with caution. Nonetheless, the two studies together reveal that *P. berghei* heterochromatin formation, maintenance and inheritance are largely hardwired throughout development.

It has been previously proposed that epigenetic reprogramming or resetting of the expression of genes involved in parasite virulence occurs during parasite passage through the vector<sup>52,53</sup>. However, no such evidence



**Figure 5.** H3K9ac enrichment in the 5'UTR of a gene does not positively correlate with its relative transcripts in female gametocytes. **(A)** Venn diagram of all genes that are significantly up- or downregulated in female gametocytes compared to asexual blood stages, male gametocytes and ookinetes, respectively. **(B)** Enriched GO-terms (slim) of both up- and downregulated genes in female gametocytes compared to all other life cycle stages. Each respective p-value is indicated. The full list of genes and GO-terms can be found in Table S4. **(C)** H3K9ac enrichment in female gametocytes for female gametocyte-specific genes. The left panel shows H3K9ac enrichment in female gametocytes for genes that are significantly up- or downregulated in female gametocytes only. The mean H3K9ac enrichment of all loci is indicated with a solid line, and the standard error is shown in a lighter colour. 1000 bp upstream and 500 bp downstream of each gene is included. Boxplots show mean H3K9ac enrichment for the 5' UTR for both gene groups. Mean H3K9ac enrichment in the 5'UTR of a gene negatively correlates with its transcript levels (Mann-Whitney test, p-value=0.0009). Boxes extend from the 25th to 75th percentiles, and the median is shown as a line in the middle of the box. Whiskers indicate 5 and 95 percentiles, respectively, and outliers are individual dots. The median is shown.



**Figure 6.** Novel and known DNA-binding motifs enriched in the 5' UTR of ookinete-specific genes. **(A)** Venn diagram of genes that are significantly up regulated in ookinetes vs female gametocytes and their overlap with AP2-O genes from Kaneko *et al.*<sup>26</sup>. DNA motifs significantly enriched in the 5' UTR of all significantly upregulated ookinete genes compared to the 5' UTR of all genes. The number of genes exhibiting each motif is shown as well as the E-value for each motif. The AP2-O motif is highly enriched. **(B)** Venn diagram of genes that are at least 8-fold significantly up regulated in ookinetes vs female gametocytes and their overlap with AP2-O genes from Kaneko *et al.*<sup>26</sup>. DNA motifs significantly enriched in the 5' UTR of 8-fold upregulated ookinete genes compared to the 5' UTR of all genes. The number of genes exhibiting each motif is shown as well as the E-value for each motif. The AP2-O motif is highly enriched.

for a reset of virulence is detected pre- and post-meiotically in our study, which could be explained by two scenarios. Firstly, such reset is likely to start immediately after nuclear fusion as in higher eukaryotes<sup>54</sup> and thus may have been missed by our study design. Secondly, any such presumed reset may involve different epigenetic marks than those examined here. Finally, the virulence-reset hypothesis is based on parasites serially maintained in rodents, which are known to become more virulent. Therefore, it is probable that the observed phenotype by Spence *et al.* is based on ill-managed heterochromatin maintenance undetectable by our study.

Gametocytes are poised cells that are rapidly activated and transform into respective gametes upon ingestion by a mosquito. In *P. falciparum*, heterochromatin boundaries are shown to expand outside telomeres as the parasite lifecycle progresses from ABS to gametocytes<sup>14</sup>. However, our data do not show any major differences in heterochromatin distribution between the various *P. berghei* lifecycle stages examined. This result can be explained by the fact that genes that become heterochromatic in *P. falciparum* gametocytes and are involved in iBC remodeling (such as knob formation) have no clear orthologues in *P. berghei* and highlights differences of heterochromatin maintenance between the two parasite species.

Male and female gametocytes are characterized by specific transcriptomes and proteomes<sup>25,30,55,56</sup>. However, in our analysis, differential heterochromatin occupancy between these two stages is limited to 12 genes, suggesting that epigenetic gene silencing is not a key regulatory mechanism of differential gene expression between male and female gametocytes. In addition, all these genes are located in subtelomeric regions and 11 of them belong to a multigene family, and thus some of the detected differences may be due to the method's sensitivity. It remains to be seen whether any of the 12 genes show true expression differences between male and female gametocytes and relate to sex-specific functions.

Clonally variant expression of subtelomeric gene families has been extensively examined in *P. falciparum*<sup>27,57,58</sup>. In contrast, variant gene expression in other *Plasmodium spp.* is less well investigated<sup>33</sup>. Due to the nature of clonal variant expression, its identification is always hampered by the fact that bulk RNA sequencing captures the sum of all the transcripts in the population, hence the identification of variant gene expression traditionally involved the analysis of clones derived from the same starting population<sup>27</sup> and more recently, single cell sequencing<sup>33</sup>. Here, our approach of combining bulk RNA sequencing data with euchromatic and heterochromatic marks identified heterochromatic genes that are transcribed and sometimes even display a euchromatic environment, suggesting that they are variably expressed within the population. Indeed, single cell sequencing data confirmed that most of these genes are variably expressed in a different *P. berghei* line<sup>33</sup>. Therefore, combining euchromatic, transcriptional and heterochromatic traits can be used to predict variably expressed genes as an alternative to single cell sequencing.

In line with previous findings<sup>33</sup>, we confirm that the number of variably expressed genes is significantly lower in *P. berghei* than in *P. falciparum*. We currently do not have any explanation for this finding but can provide two hypotheses. Firstly, it is possible that by keeping *P. berghei* in its “unnatural” host of outbred mice under controlled conditions decreases its active multigene repertoire, and as a result, *P. berghei* fix on a low number of genes to be expressed, similar to *P. falciparum* under *in vitro* culture conditions<sup>58</sup>. Secondly, it is known that in the rodent malaria parasite *P. chabaudi*, the number of expressed *cir* genes is highest directly after mosquito transmission and decreases with multiple passages between mice<sup>53</sup>. It is possible that our observations of the low number of variably expressed genes are based on this phenomenon.

In several *Plasmodium* species and life stages, it is now proven that the occupancy of the euchromatic mark H3K9ac in the 5'UTR of a gene correlates with its transcription<sup>11,13,15–17</sup>. Here we show for the first time that the same is true for *P. berghei* asexual blood stages. Additionally, for the first time for any *Plasmodium spp.*, we reveal that the correlation between H3K9ac and gene expression also applies to male gametocytes and ookinetes.

Here, the peak shape of H3K9ac occupancy was further investigated. We show that ribosomal protein genes display a very specific H3K9ac pattern peaking around the translational start codon in asexual blood stages, female gametocytes and ookinetes. This pattern is irrespective of transcript intensity but is connected to the gene group instead. Additionally, using mapped TSS from *P. falciparum*<sup>46</sup>, we do not find that TSSs of ribosomal protein genes to be closer to the ATG than TSSs from other genes. A similar sharp H3K9ac peak around the TSS has been found in *A. thaliana* for genes involved in translation<sup>48</sup> which could explain our findings. We have currently no explanation why this pattern is not found in male gametocytes.

A striking difference between male and female gametocytes is the level of H3K9ac occupancy in the 5'UTR of genes, which correlates with transcript abundance in male but not in female gametocytes. Thus it appears that gene transcript abundance in female gametocytes is independent of H3K9ac levels. We propose that the female gametocyte is a stage of epigenetic remodeling that involves universal marking of gene promoters with H3K9ac. At the same time, genes that are involved in ribosome biogenesis and translation are downregulated, a phenomenon that together with translational mRNA repression by the DOZI-complex suggests that a multilayered mechanism regulating zygotic development operates in *Plasmodium*. It will be interesting to see how other euchromatic marks for example H3K4me3 are distributed in female gametocytes. In mice, H3K4me3 occupancy usually remains narrowly restricted to promoter regions but its occupancy expands during oocyte development<sup>59–61</sup>. It is possible that a similar occupancy shift occurs in female gametocytes.

Mature oocytes in mammals, flies and worms are transcriptionally silent when passing through meiosis I preceding fertilization<sup>62</sup>. Additionally, early embryonic development depends solely on maternally deposited proteins and RNAs and coincides with low or undetectable transcription<sup>63</sup>. Zygotic transcription often resumes after several hours or days post fertilization, a process called maternal-to-zygotic transition accompanied by zygotic genome activation, and in flies, for example, is controlled by a master transcription factor that is not necessarily associated with H3K9ac<sup>64</sup>. Additionally, epigenetic reprogramming occurs before and after fertilization, ensuring the transition from a highly specific cell type back to a totipotency state able to form a new organism. However, direct comparison of malaria parasite female gametocytes to oocytes of multicellular organisms is complicated by the fact that malaria parasites undergo meiosis directly after fertilization<sup>65</sup>, in contrast to metazoan cells where

meiosis precedes fertilization. Nonetheless, it is tempting to hypothesise that, similar to higher eukaryotes, (i) the female gametocyte is transcriptionally poised and (ii) transcription in the ookinete is resumed only from a selected set of genes. While there is no evidence so far that the female gametocyte is transcriptionally poised, there is evidence for the second part of the hypothesis. Firstly, when expression of a fluorescent reporter protein is controlled by an ookinete-specific gene promoter the protein is detected from both the maternal and paternal genomes in the first 24 hours after fertilization<sup>66</sup>. Secondly, if expression of a fluorescent reporter protein is controlled by an housekeeping gene promoter the protein is detected only from the maternal genome<sup>66</sup>. This could either mean that the paternal genome is silenced, or more likely, that both parental genomes are transcriptionally poised, and the detected expression stems from inherited maternal mRNA.

Two histone variants associated with active promoters in *P. falciparum* ABS parasites (H2A.Z and H2B.Z)<sup>13,67</sup> are significantly downregulated in female gametocytes in this study (Table S3). As mRNA is stored in the RNP complex in female gametocytes, we do not know which mRNA is translated into protein and which mRNA is translationally poised at this stage. Thus, it will be interesting to see if either female gametocytes have a low occurrence of H2A.Z and H2B.Z, or if these histone variants are not deposited into the chromatin during the early development of the zygote. Proteomic studies identified both H2A.Z and H2B.Z in female gametocytes in both *P. berghei* and *P. falciparum*<sup>55,56</sup>, pointing towards the latter possibility. Supportively, in mice, H2A.Z remains undetectable in embryonic chromatin before the late 2-cell stage<sup>68</sup>. Thus, we propose a scenario where the female gametocyte is transcriptionally poised, and where transcription of stage-specific genes is resumed in the ookinete stage via AP2-O, and potentially by additional transcription factors, as we and others<sup>19,33</sup> identified additional DNA motifs abundant in ookinete-specific genes. Still, it is important to note that these newly identified motifs need experimental validation in future studies.

## Material and Methods

**Ethics statement.** All animal experimental procedures were reviewed and approved by the Imperial College London Animal Welfare and Ethical Review Body (AWERB) and the United Kingdom Home Office. These procedures were in accordance with the Animal Scientific Procedures Act 1986, under the UK Home Office Licenses PPL70/7185 and PPL70/8788.

**Plasmodium berghei maintenance.** *P. berghei* clone ANKA 2.33 (for asexual blood stages)<sup>28</sup> and the 507m6cl1 (c507) line (for ookinetes)<sup>31</sup> were maintained in 6–8 week old female Tuck's Ordinary (TO) (Harlan, UK). The 820cl1m1cl1 (wt-fluo) line (for gametocytes)<sup>29</sup> was maintained in 6–8 week old female CD1 mice (Harlan, UK).

**Chromatin extraction and fragmentation.** Asexual blood stage parasites were harvested via heart puncture and passed through a Plasmodipur filter (Europroxima) to remove leucocytes, resuspended in RPMI-1640 medium (Sigma-Aldrich) and crosslinked with 1% formaldehyde in PBS for 10 min at 37 °C. Crosslinking was quenched adding glycine to an end concentration of 0.125 M. RBC were then lysed with 0.15% saponin (in PBS) on ice for 5–10 minutes. To obtain nuclei the resulting parasite pellet was lysed with cell lysis buffer (20 mM HEPES, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% NP-40, 1 mM DTT, 1x protease inhibitor (Roche)). The pellet was resuspended in sonication buffer (1% SDS, 50 mM Tris pH8, 10 mM EDTA, 1x protease inhibitor (cComplete, Mini, EDTA-free, Roche)) and sheared for 25 minutes (30 sec ON, 30 sec OFF; settings HIGH) using a Bioruptor Plus sonication device (Diagenode) to obtain DNA fragments of around 100–300 bp.

The method for the purification of gametocytes was modified from<sup>69</sup>. Briefly, mice were pretreated by intra peritoneal injection of 0.2 ml phenylhydrazine (6 mg/ml in PBS) to stimulate reticulocyte formation two days prior to infection with parasites. Gametocyte-enriched blood was harvested via heart puncture and blood was immediately resuspended in 4 °C coelenterazine loading buffer (CLB) (1x PBS, 20 mM HEPES, 20 mM glucose, 4 mM sodium bicarbonate, 1 mM EDTA, 0.1% BSA in PBS, pH 7.24–7.31) and magnet-purified using D Columns on a SuperMACS II Separator (Miltenyi Biotec). Magnet-purified parasites were crosslinked in 1% formaldehyde for 10 minutes at 37 °C, quenched with glycine to an end concentration of 0.125 M and resuspended in FACS buffer (PBS with 2 mM HEPES, 2 mM glucose, 0.4 mM NaHCO<sub>3</sub>, 0.01% BSA, 2.5 mM EDTA). Male and female gametocytes were sorted according to color (GFP for males and RFP for females) on a FACSAria III with a 70 μm nozzle at 4 °C. Purified gametocyte pellets were lysed in 150 μl sonication buffer and chromatin prepared as above. FACS was performed with four biological replicates, and a total number of 3.1E7 female gametocytes and 2.2E7 male gametocytes were used for ChIP, respectively.

Ookinetes were cultured *in vitro* as described<sup>70</sup>. Briefly, mice were pretreated by intra peritoneal injection of 0.2 ml phenylhydrazine (6 mg/ml in PBS) to stimulate reticulocyte formation two days prior to infection with parasites. Gametocyte-enriched blood was harvested via heart puncture and blood was immediately resuspended in 30 ml of ookinete medium (RPMI-HEPES complemented with 100 μM xanthurenic acid, 200 μM hypoxanthine and 10% BSA, pH 7.4) and incubated at 21 °C for 24 hours. After 24 hours, ookinetes were magnet-purified using 1 μl monoclonal anti-P28 antibody 13.1<sup>32</sup> coupled to 10 μl magnetic beads (Dynabeads M-280 Sheep Anti-Mouse IgG). Purified ookinetes were crosslinked in 1% formaldehyde for 10 minutes at 37 °C, quenched with glycine to an end concentration of 0.125 M. Purified ookinete pellets were lysed in 150 μl sonication buffer and chromatin was prepared as above. Two ookinete cultures were pooled for the ChIP to obtain enough material, with female to ookinete conversion rates of 66% and 71.5%, respectively.

**Chromatin immunoprecipitation.** Antibodies used for ChIP were rabbit anti H3K9ac (Diagenode Cat# C15410004, RRID:AB\_2713905) and rabbit anti-PbHP1<sup>14</sup>. 1 μg antibody was incubated with up to 500 ng chromatin in ChIP buffer (5% TritonX-100, 750 mM NaCl, 5 mM EDTA, 2.5 mM EGTA, 100 mM HEPES) at 4 °C over

night. The next day, 50  $\mu$ l Protein A Dynabeads (Fisher 10001D) were added and further incubated for 1–2 h. After washing with buffers containing 100, 150 and 250 mM NaCl, immunoprecipitated DNA was eluted and purified using PCR minelute purification columns (Qiagen).

For each antibody several ChIP reactions were performed in parallel to obtain sufficient amount of DNA for ChIP-seq: (asexual blood stages (3xH3K9ac and 8xPbHP1); female/male gametocytes (3x H3K9ac and 6xPbHP1); ookinetes (2x H3K9ac and 5xPbHP1)).

The following number of biological replicates were pooled to obtain enough material for ChIP: ABS (1); FG (4), MG(4), OOK (2).

**ChIP sequencing.** For each sequencing library up to 10 ng of ChIP or input DNA were end-repaired, extended with 30 A-overhangs and ligated to barcoded NextFlex adapters (Bio Scientific) as described previously<sup>71</sup>. Libraries were amplified (98 °C for 2 min; four cycles 98 °C for 20 sec, 62 °C for 3 min; 62 °C for 5 min) using KAPA HiFi HotStart ready mix (KAPA Biosystems) and NextFlex primer mix (Bio Scientific) as described<sup>72</sup>. 225–325 bp fragments (including the 125 bp NextFlex adapter) were size- selected using a 2% E-Gel Size Select agarose gel (Thermo Fisher Scientific) and amplified by PCR for ten (asexual blood stages and ookinetes) or eleven (female and male gametocytes) cycles under the same condition as described above. Library purification and removal of adapter dimers was performed with Agencourt AMPure XP beads in a 1:1 library:beads ratio (Beckman Coulter). ChIP-seq libraries were sequenced for 75 bp single-end reads using the NextSeq. 500/550 High Output v2 kit (Illumina) on the Illumina NextSeq. 500 system.

Sequencing reads were mapped against the *P. berghei* ANKA reference genome v3 using BWA samse (v0.7.12-r1039)<sup>73</sup> (and filtered to mapping quality  $\geq 15$  (SAMtools v1.2)<sup>74</sup>). Only uniquely mapped reads were used for further analysis. The PCA plot was calculated using default settings in the DeepTools2 suite<sup>75</sup> using non-overlapping 1000 bp bins.

**PbHP1 analysis.** The bam files containing mapped reads from each ChIP were normalised against their input using *bamcompare* in DeepTools2<sup>75</sup>. For PbHP1, the default settings were used with the following changes: 100 bp bin size, 0.01 pseudocount and 1000 bp smoothing.

The mean of the log2 PbHP1/input ratio was then extracted for each gene, making each gene fit 500 bp using one 500 bp bin using *computematrix* in DeepTools2<sup>75</sup> with default settings with the following changes: missing values = 0 and skip 0. The resulting *bigwig* files were hierarchically clustered with average linkage and Euclidean distance as similarity metric in Cluster3.0<sup>76</sup>. The resulting heatmap and tree was inspected in Treeview<sup>77</sup>.

The multigene family list is based on Table S4 from Otto *et al.*,<sup>36</sup> and we updated gene IDs to *P. berghei* genome version 3 and found 379 genes.

For visualising our data next to the data from Frasnka *et al.*<sup>14</sup> we used the author's PbHP1 over input bed-graph file from the GEO database (GSE102695) and viewed it in IGV<sup>78</sup>. Chromosome plots were drawn using Phenogram<sup>79</sup>.

**H3K9ac analysis.** The bam files containing aligned reads from each ChIP were normalised against their input using *bamcompare* from the DeepTools2 suite<sup>75</sup>. For H3K9ac, default settings were used with the following changes: bin size 50 bp, 0.01 pseudocount, 100 bp smoothing. Analyses using heatmaps and summary plots were performed using the *deeptools2*<sup>75</sup> suite on *usegalaxy.org* as well as *usegalaxy.eu*.

For the ribosomal subunit genes, we used the *P. berghei* orthologues from *P. falciparum* reported in<sup>80</sup>. For statistical analysis and drawing of boxplots, we used Graphpad Prism 7.03. Wilcoxon matched-pairs signed rank test was performed on each data set and P-values  $< 0.01$  are reported, with (\*\*\*)  $P \leq 0.0001$ .

**Motif identification.** 5' UTRs were extracted from PBANKA genome version 3. Sequences between annotated genes were attributed to the nearest gene as follows: For head-to-head genes, the 5' UTRs were split 1:1. For tail-to-head genes, the sequence was split 1:2 (1/3 was assigned as 3' downstream region to the gene ending and 2/3 was assigned as 5' UTR to the gene starting). To find motifs enriched in ookinete-expressed genes, we used DREME (v 5.0.0)<sup>51</sup> with default settings locally on the MEME suite<sup>81</sup>.

We used the build-in Gene Ontology tool of PlasmoDB with default settings to identify enriched GO terms<sup>82</sup>.

**RNA preparation.** Asexual blood stage parasites were harvested via heart puncture and resuspended in 15 ml RPMI-HEPES and passed through a Plasmodipur filter (Europroxima) to remove white blood cells. Parasite/RBC pellet was lysed with 10 ml RBC-lysis buffer (150 mM NH<sub>4</sub>Cl, 10 mM KHCO<sub>3</sub>, 1 mM EDTA) for 20 min on ice. Parasite-pellets were washed once in ice-cold PBS and lysed in 500  $\mu$ l TRIzol (Invitrogen) and stored at  $-80$  °C.

The method for the purification of gametocytes was modified from<sup>69</sup>. Briefly, mice were pretreated by intra peritoneal injection of 0.2 ml phenylhydrazine (6 mg/ml in PBS) to stimulate reticulocyte formation two days prior to infection with parasites. Gametocytes were harvested via heart puncture and blood was immediately resuspended in 4 °C coelenterazine loading buffer (CLB) (1x PBS, 20 mM HEPES, 20 mM glucose, 4 mM sodium bicarbonate, 1 mM EDTA, 0.1% BSA in PBS, pH 7.24–7.31) and magnet-purified using D Columns on a SuperMACS II Separator (Miltenyi Biotec). Magnet-purified parasites were resuspended in FACS buffer (PBS with 2 mM HEPES, 2 mM glucose, 0.4 mM NaHCO<sub>3</sub>, 0.01% BSA, 2.5 mM EDTA). Male and female gametocytes were sorted according to color (GFP for males and RFP for females) on a FACSaria III with a 100  $\mu$ m nozzle at 4 °C. Purified gametocyte pellets were lysed in 500  $\mu$ l TRIzol (Invitrogen) and stored at  $-80$  °C.

Ookinetes were cultured *in vitro* as described<sup>70</sup>. Briefly, mice were pre-treated by intra peritoneal injection of 0.2 ml phenylhydrazine (6 mg/ml in PBS) to stimulate reticulocyte formation two days prior to infection with parasites. Gametocytes-enriched blood was harvested via heart puncture and blood was immediately resuspended in

30 ml of ookinete medium (RPMI-HEPES complemented with 100 μM xanthurenic acid, 200 μM hypoxanthine and 10% BSA, pH7.4) and incubated at 21 °C for 24 hours. After 24 hours, ookinetes were magnet-purified using 1 μl monoclonal anti-P28 antibody 13.1<sup>32</sup> coupled to 10 μl magnetic beads (Dynabeads M-280 Sheep Anti-Mouse IgG) and resuspended in 500 μl TRIzol (Invitrogen) and stored at −80 °C. Female to ookinete conversion rates for the three biological replicates were as follows: 65% 67% and 75%, respectively.

Genomic DNA was used as a control for the library preparation protocol. For this, asexual blood stage parasites were harvested via heart puncture, resuspended in 15 ml RPMI-HEPES, and passed through a Plasmodipur filter (Europroxima) to remove white blood cells. Parasite/RBC pellet was lysed with 10 ml RBC-lysis buffer (150 mM NH<sub>4</sub>Cl, 10 mM KHCO<sub>3</sub>, 1 mM EDTA) for 20 min on ice. Parasite-pellets were washed once in ice-cold PBS and lysed in buffer A (500 mM NaAc, 100 mM NaCl, 1 mM EDTA, pH5.2) and 3% SDS. DNA was extracted using phenol:chloroform and sheared to 150 bp fragments on a Bioruptor Plus sonication device (Diagenode).

**RNA sequencing.** RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo). Residual gDNA was digested with the TurboDNA-free kit (Ambion). Stranded RNA sequencing libraries were prepared using the RNA HyperPrep Kit (KAPA) following manufacturer's protocol with the exception of the amplification step set to 60 °C. The RNA library was sequenced in 43 bp paired-end reads on an Illumina NextSeq 500 sequencer and each sample was split over four non-independent lanes. RNAseq was performed using biological triplicates for each condition.

We used sheared genomic DNA (gDNA) as a technical control and did not observe a bias towards GC-rich(er) sequences (Fig. S1A), a known problem in *Plasmodium falciparum* NGS<sup>13</sup>.

**RNA sequencing analysis.** Quality of the reads were checked by eye using FASTQC. Reads were aligned separately for each lane to *PBANKA* genome v3 using HiSat2 with default settings (v 2.0.5.2; −max intron size 5000; −fr)<sup>83</sup>. Mapped reads were pooled into one bam file per condition and replicate. Importantly, to adjust for potential GC-bias gDNA was processed alongside RNA samples as a control for the RNA library preparation. GC-bias was calculated according to<sup>84</sup> using the default settings for *computeGCBias* in DeepTools2 suite<sup>75</sup>.

The PCA plot was calculated using default settings in the DeepTools2 suite<sup>75</sup>.

We calculated the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) value for each gene using featurecounts with default settings<sup>85</sup>.

For comparative transcriptomics we used Deseq 2 (v2.11.39) with default settings<sup>86</sup>. We considered significantly differently regulated genes as having a p value of lower than 0.001 (adjusted for multiple testing with the Benjamini-Hochberg procedure which controls false discovery rate (FDR)).

All of the analysis has been performed using usegalaxy.org (an open source, web-based platform for data intensive biomedical research)<sup>87</sup> unless stated otherwise. Additionally, usegalaxy.eu was used for some calculations.

## Data Availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus<sup>88</sup> and are accessible through GEO Series accession number GSE130278 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130278>). This includes raw read data files as well as output from Deseq 2<sup>86</sup> and featurecounts<sup>85</sup>. Bam files are deposited on zenodo.com.

Received: 19 November 2019; Accepted: 24 March 2020;

Published online: 14 April 2020

## References

1. WHO. World Malaria Report 2018. 4. WHO ISBN 978-92-4-156565-3 (2018). doi:ISBN 978 92 4 1564403.
2. Mair, G. R. *et al.* Regulation of sexual development of *Plasmodium* by translational repression. *Science* **313**, 667–9 (2006).
3. Mair, G. R. *et al.* Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. *PLoS Pathog.* **6**, e1000767 (2010).
4. Janse, C. J., van der Klooster, P. F., van der Kaay, H. J., van der Ploeg, M. & Overdulve, J. P. DNA synthesis in *Plasmodium berghei* during asexual and sexual development. *Mol. Biochem. Parasitol.* **20**, 173–82 (1986).
5. Sinden, R. E. *Plasmodium* differentiation in the mosquito. *Parassitologia* **41**, 139–148 (1999).
6. Voss, T. S., Bozdech, Z. & Bártfai, R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr. Opin. Microbiol.* **20**, 88–95 (2014).
7. Flueck, C. *et al.* *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* **5**, e1000569 (2009).
8. Cortés, A. *et al.* Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog.* **3**, 1023–1035 (2007).
9. Sharma, P. *et al.* An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. *J. Biol. Chem.* **288**, 19429–40 (2013).
10. Lopez-Rubio, J.-J., Mancio-Silva, L. & Scherf, A. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5**, 179–90 (2009).
11. Salcedo-Amaya, A. M. *et al.* Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9655–60 (2009).
12. Gupta, A. P. *et al.* Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*. *PLoS Pathog.* **9**, e1003170 (2013).
13. Bártfai, R. *et al.* H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog.* **6**, e1001223 (2010).
14. Fraschka, S. A. *et al.* Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites. *Cell Host Microbe* **23**, 407–420.e8 (2018).
15. Gómez-Díaz, E. *et al.* Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. *Sci. Rep.* **7** (2017).



16. Zanghi, G. *et al.* A Specific PfEMP1 Is Expressed in *P. falciparum* Sporozoites and Plays a Role in Hepatocyte Infection. *Cell Rep* **22**, 2951–2963 (2018).
17. Muller, I. *et al.* Transcriptome and histone epigenome of *Plasmodium vivax* salivary-gland sporozoites point to tight regulatory control and mechanisms for liver-stage differentiation in relapsing malaria. *Int. J. Parasitol.* **49**, 501–513 (2019).
18. Bunnik, E. M. *et al.* Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages. *Nat. Commun.* **9**, 1910 (2018).
19. Modrzynska, K. *et al.* A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the *Plasmodium* Life Cycle. *Cell Host Microbe* **21**, 11–22 (2017).
20. Yuda, M. *et al.* Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol. Microbiol* **71**, 1402–14 (2009).
21. Iwanaga, S., Kaneko, I., Kato, T. & Yuda, M. Identification of an AP2-family protein that is critical for malaria liver stage development. *PLoS One* **7**, e47557 (2012).
22. Santos, J. M. *et al.* Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I Transcription Factor. *Cell Host Microbe* **21**, 731–741 (2017).
23. Kafsack, B. F. C. *et al.* A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 253–7 (2014).
24. Sinha, A. *et al.* A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* **507**, 253–7 (2014).
25. Kent, R. S. *et al.* Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite *Plasmodium berghei*. *Nat. Microbiol* **3**, 1206–1213 (2018).
26. Kaneko, I., Iwanaga, S., Kato, T., Kobayashi, I. & Yuda, M. Genome-Wide Identification of the Target Genes of AP2-O, a *Plasmodium* AP2-Family Transcription Factor. *PLoS Pathog.* **11**, e1004905 (2015).
27. Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.* **22**, 925–38 (2012).
28. Dearsly, A. L., Sinden, R. E. & Self, I. A. Sexual development in malarial parasites: Gametocyte production, fertility and infectivity to the mosquito vector. *Parasitology* **100**, 359–368 (1990).
29. Ponzi, M. *et al.* Egress of *Plasmodium berghei* gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. *Cell. Microbiol* **11**, 1272–1288 (2009).
30. Yeoh, L. M., Goodman, C. D., Mollard, V., Mcfadden, G. I. & Ralph, S. A. Comparative transcriptomics of female and male gametocytes in *Plasmodium berghei* and the evolution of sex in alveolates. *BMC Genomics* **18** (2017).
31. Janse, C. J. *et al.* High efficiency transfection of *Plasmodium berghei* facilitates novel selection procedures. *Mol. Biochem. Parasitol.* **145**, 60–70 (2006).
32. Winger, L. A. *et al.* Ookinete antigens of *Plasmodium berghei*. Appearance on the zygote surface of an Mr 21 kD determinant identified by transmission-blocking monoclonal antibodies. *Parasite Immunol.* **10**, 193–207 (1988).
33. Howick, V. M. *et al.* The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science (80-)*. **365**, eaaw2619 (2019).
34. Hoeijmakers, Wa. M. *et al.* *Plasmodium falciparum* centromeres display a unique epigenetic makeup and cluster prior to and during schizogony. *Cell. Microbiol.* **14**, 1391–401 (2012).
35. Lasonder, E. *et al.* Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog.* **4**, (2008).
36. Otto, T. D. *et al.* A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol.* **12**, (2014).
37. Srinivasan, P., Fujioka, H. & Jacobs-Lorena, M. PbCap380, a novel oocyst capsule protein, is essential for malaria parasite survival in the mosquito. *Cell. Microbiol.* **10**, 1304–1312 (2008).
38. Itsara, L. S. *et al.* PfCap380 as a marker for *Plasmodium falciparum* oocyst development *in vivo* and *in vitro*. *Malar. J.* **17**, 135 (2018).
39. Brancucci, N. M. B. *et al.* Heterochromatin Protein 1 Secures Survival and Transmission of Malaria Parasites. *Cell Host Microbe* **16**, 165–176 (2014).
40. Haase, S., Hanssen, E., Matthews, K., Kalanon, M. & de Koning-Ward, T. F. The Exported Protein PbCP1 Localises to Cleft-Like Structures in the Rodent Malaria Parasite *Plasmodium berghei*. *PLoS One* **8**, e61482 (2013).
41. Bushell, E. *et al.* Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260–272.e8 (2017).
42. Chang, P., Giddings, T. H., Winey, M. & Stearns, T. E-tubulin is required for centriole duplication and microtubule organization. *Nat. Cell Biol.* **5**, 71–76 (2003).
43. Ross, I., Clarissa, C., Giddings, T. H. & Winey, M. epsilon-tubulin is essential in *Tetrahymena thermophila* for the assembly and stability of basal bodies. *J. Cell Sci.* **126**, 3441–3451 (2013).
44. Fougère, A. *et al.* Variant Exported Blood-Stage Proteins Encoded by *Plasmodium* Multigene Families Are Expressed in Liver Stages Where They Are Exported into the Parasitophorous Vacuole. *PLoS Pathog.* **12**, 1–37 (2016).
45. Pardini, E. M. *et al.* Proteomic and genetic analyses demonstrate that *Plasmodium berghei* blood stages export a large and diverse repertoire of proteins. *Mol. Cell. Proteomics* **12**, 426–48 (2013).
46. Adjalley, S. H., Chabbert, C. D., Klaus, B., Pelechano, V. & Steinmetz, L. M. Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*. *Cell Rep* **14**, 2463–2475 (2016).
47. Cremona, M. A. *et al.* Peak shape clustering reveals biological insights. *BMC Bioinformatics* **16** (2015).
48. Ha, M., Ng, D. W.-K., Li, W.-H. & Chen, Z. J. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res.* **21**, 590–598 (2011).
49. Gómez-díaz, E., Rivero, A., Chandre, F. & Corces, V. G. Insights into the epigenomic landscape of the human malaria vector *Anopheles gambiae*. *Front. Genet* **5**, 1–12 (2014).
50. Guerreiro, A. *et al.* Genome-wide RIP-Chip analysis of translational repressor-bound mRNAs in the *Plasmodium* gametocyte. *Genome Biol.* **15**, 493 (2014).
51. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–9 (2011).
52. Bachmann, A. *et al.* Mosquito Passage Dramatically Changes var Gene Expression in Controlled Human *Plasmodium falciparum* Infections. *PLoS Pathog.* **12**, e1005538 (2016).
53. Spence, P. J. *et al.* Vector transmission regulates immune control of *Plasmodium* virulence. *Nature* **498**, 228–31 (2013).
54. Zhou, Lquan & Dean, J. Reprogramming the genome to totipotency in mouse embryos. *Trends Cell Biol* **25**, 82–91 (2015).
55. Lasonder, E. *et al.* Integrated transcriptomic and proteomic analyses of *P. falciparum* gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res* **44**, 6087–6101 (2016).
56. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* **121**, 675–687 (2005).
57. Janes, J. H. *et al.* Investigating the Host Binding Signature on the *Plasmodium falciparum* PfEMP1 Protein Family. *PLoS Pathog.* **7**, e1002032 (2011).
58. Peters, J. M., Fowler, E. V., Krause, D. R., Cheng, Q. & Gatton, M. L. Differential Changes in *Plasmodium falciparum* var Transcription during Adaptation to Culture. *J. Infect. Dis.* **195**, 748–755 (2007).
59. Dahl, J. A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).

60. Liu, X. *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
61. Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
62. Seki, Y. *et al.* Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134**, 2627–2638 (2007).
63. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–42 (2009).
64. Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *Elife* **3**, 1–20 (2014).
65. Sinden, R. E. & Hartley, R. Identification of the meiotic division of malarial parasites. *J. Protozool* **32**, 742–744 (1985).
66. Ukegbu, C. V., Cho, J.-S., Christophides, G. K. & Vlachou, D. Transcriptional silencing and activation of paternal DNA during Plasmodium berghei zygotic development and transformation to oocyst. *Cell. Microbiol.* **17**, 1230–1240 (2015).
67. Hoeijmakers, W. A. M. *et al.* H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the Plasmodium falciparum genome. *Mol. Microbiol.* **87**, 1061–73 (2013).
68. Bošković, A. *et al.* Analysis of active chromatin modifications in early mammalian embryos reveals uncoupling of H2A.Z acetylation and H3K36 trimethylation from embryonic genome activation. *Epigenetics* **7**, 747–757 (2012).
69. Beetsma, A. L., van de Wiel, T. J., Sauerwein, R. W. & Eling, W. M. Plasmodium berghei ANKA: purification of large numbers of infectious gametocytes. *Exp. Parasitol.* **88**, 69–72 (1998).
70. Rodríguez, M. C. *et al.* Plasmodium berghei: routine production of pure gametocytes, extracellular gametes, zygotes, and ookinetes. *Exp. Parasitol.* **101**, 73–6 (2002).
71. Hoeijmakers, W. A. M., Bártfai, R., François, K.-J. & Stunnenberg, H. G. Linear amplification for deep sequencing. *Nat. Protoc.* **6**, 1026–36 (2011).
72. Kensch, P. R. *et al.* The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res* **44**, 2110–2124 (2016).
73. Li, H., Li, H., Durbin, R. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
74. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
75. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, 160–165 (2016).
76. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
77. Saldanha, A. J. Java Treeview - Extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
78. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–92 (2013).
79. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **6**, (2013).
80. Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **3**, e03080 (2014).
81. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8 (2009).
82. Aurrecochea, C. *et al.* PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009).
83. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
84. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72–e72 (2012).
85. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biol.* **15**, 550 (2014).
87. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
88. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–10 (2002).

## Acknowledgements

The authors wish to acknowledge the support from the FACS-facility from Imperial College London, namely Jess Rowley and Jane Srivastava. Andy Brockman for extracting the UTRs from the *P. berghei* ANKA genome, Tony Brooks and Paola Niola from UCL for help with RNA sequencing. Bob MacCallum, Amie Jaye and Dan Lawson for assistance in all things computational. Till S. Voss for sharing the *PbHP1* antibody before its publication. Virginia M. Howick for providing high quality images of selected genes from the malaria cell atlas. Sophie Adjalley for providing the TSS sites for *P. falciparum* and TSS discussions. Christian Flueck for critical reading of the manuscript. We are grateful to Jen Hillman for guidance/assistance on usegalaxy.org and the Freiburg Galaxy team for their assistance using usegalaxy.eu. This work was supported by the Wellcome Trust [093587/Z/10/Z to G.K.C. and D.V., 107983/Z/15/Z to G.K.C.], the The Netherlands Organization for Scientific Research [NWO-Vidi 864.11.007 to R.B.]; by an advanced Postdoc.Mobility fellowship from the Swiss National Science Foundation [P300P3\_158527 to K.W.]; a PhD fellowship from the European Community's Seventh Framework Program [242095, 290080 to S.K.]. Funding for open access charge: [Wellcome Trust 107983/Z/15/Z].

## Author contributions

conceptualization, K.W., R.B., D.V., G.K.C.; investigation, K.W.; formal analysis, K.W., S.A.F.; resources, R.B., D.V., G.K.C.; writing – original draft, K.W.; writing – review & editing, K.W., S.A.F., D.V., R.B., G.K.C.; visualization, K.W.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63121-5>.

**Correspondence** and requests for materials should be addressed to K.W. or G.K.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020