Published in final edited form as:
Cell Host Microbe. 2013 December 11; 14(6): 696–706. doi:10.1016/j.chom.2013.11.007.

Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum*

Nadia Ponts[1,6], **Lijuan Fu**[4], **Elena Y. Harris**[2], **Jing Zhang**[3,4], **Duk-Won D. Chung**[1], **Michael C. Cervantes**[1], **Jacques Prudhomme**[1], **Vessela Atanasova-Penichon**[6], **Enric Zehraoui**[6], **Evelien Bunnik**[1], **Elisandra M. Rodrigues**[1], **Stefano Lonardi**[2], **Glenn R. Hicks**[5], **Yinsheng Wang**[4], and **Karine G. Le Roch**[1]

[1]Department of Cell Biology and Neuroscience, University of California, 900 University Avenue, Riverside, CA 92521, USA.

[2]Department of Computer Science and Engineering, University of California, 900 University Avenue, Riverside, CA 92521, USA.

[3]School of Chemistry & Materials Science, Shaanxi Normal University, 199 South Chang'an Road, Xi'an 710062, China.

[4]Department of Chemistry, University of California, 900 University Avenue, Riverside, CA 92521, USA.

[5]Center for Plant Cell Biology and Dept of Botany & Plant Sciences, University of California, 900 University Avenue, Riverside, CA 92521, USA.

[6]INRA, UR1264-MycSA, 71 avenue E. Bourlaux, CS20032, 33882 Villenave d'Ornon Cedex, France.

SUMMARY

Cytosine DNA methylation is an epigenetic mark in most eukaryotic cells that regulates numerous processes, including gene expression and stress responses. We performed a genome-wide analysis of DNA methylation in the human malaria parasite *Plasmodium falciparum*. We mapped the positions of methylated cytosines and identified a single functional DNA methyltransferase, PfDNMT, that may mediate these genomic modifications. These analyses revealed that the malaria genome is asymmetrically methylated, in which only one DNA strand is methylated, and shares common features with undifferentiated plant and mammalian cells. Notably, core promoters are hypomethylated and transcript levels correlate with intra-exonic methylation. Additionally, there are sharp methylation transitions at nucleosome and exon-intron boundaries. These data suggest that DNA methylation could regulate virulence gene expression and transcription elongation. Furthermore, the broad range of action of DNA methylation and uniqueness of PfDNMT suggest that the methylation pathway is a potential target for anti-malarial strategies.

Contact: Correspondence and requests for materials should be addressed to Karine G. Le Roch (karine.leroch@ucr.edu).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, 7 figures, 2 tables, and Supplemental References.

INTRODUCTION

In eukaryotes, DNA marking with methylated cytosines (5-methylcytosine or me^5C) is involved in a wide array of processes such as genomic imprinting, DNA repair, response to stress and regulation of gene expression (Boyko and Kovalchuk, 2008; Li et al., 1993; Tost, 2009). The role of DNA methylation in host-virus interactions and virulence is well documented. By contrast, there is little information with regard to other human pathogens such as the entire phylum of apicomplexan parasites including the human malaria parasite, *Plasmodium falciparum*, responsible for more than one million deaths per year. The 23Mb genome of *P. falciparum* consists of 14 chromosomes, encodes about 5,500 genes and is the most AT-rich genome sequenced to date (more than 90% in intergenic regions; (more than 90% in intergenic regions; Gardner et al., 2002). For years, the very low GC-content of the parasite's genome challenged the classical methods of me^5C detection (problems of detection thresholds in mass spectrometry-based methods, Gissot et al., 2008, and bias towards me^5C in CpG context using restriction enzyme and immunoprecipitation-based methods). As a consequence, the methylation status of *P. falciparum*'s DNA remains unclear. Previous mass spectrometry-based analyses failed to identify methylated nucleosides in *P. falciparum* (Choi et al., 2006; Pollack et al., 1982). Nonetheless, experiments involving methylase-sensitive restriction analyses suggested the presence of partial cytosine methylation at the locus of the gene coding for the dihydrofolate reductase-thymidylate synthase (Pollack et al., 1991), providing evidence that *P. falciparum*'s genome can carry methylations.

In the present study, we clarify the methylation status of *P. falciparum*'s genome and provide a genome-wide map of me^5C distribution. We show that *P. falciparum*'s genome is methylated using mass spectrometry and hypomethylating drugs assays. We also identify a unique candidate DNA methyltransferase in the parasite's genome and demonstrate its cytosine methyltransferase activity, both *ex vivo* and *in vitro*. Finally, we mapped the cytosines that are methylated in the *P. falciparum* genome during the intraerythrocytic cycle. To do so, we used the state-of-the-art technique of bisulfite conversion of unmethylated cytosines coupled to high throughput sequencing (or BS-seq), which allows the study of DNA methylation in an AT-rich context (Cokus et al., 2008; Lister et al., 2008, 2009). Our results revealed that non-CG methylations, generally overlooked by other methods, could be of major importance for the regulation of transcription elongation, splicing, and the silencing of virulence genes. Applications of such works to different organisms could remodel the current perception of their methylomes.

RESULTS

Detection of methylcytosines in *P. falciparum*'s genome

We analyzed the nucleoside mixture arising from the enzymatic digestion of *P. falciparum* strain 3D7 genomic DNA by LC-MS/MS. We used the highly sensitive Thermo TSQ Vantage triple-quadrupole Mass Spectrometer to prevent insufficient detection capacity. In addition, increased sensitivity was achieved by using formic acid as a proton donor for positive electrospray ionization Finally, more efficient ionization was obtained in our measurements when the 5-mdC was separated from the nucleosides mixture plus 5-mC by liquid chromatography. Indeed, impurities present in the sample decrease sensitivity and proper separation prior ionization is essential. Using this set up, we successfully detected the presence of 5-methyl-2'-deoxycytidine in three independent genomic DNA preparations from asynchronous populations of *P. falciparum* (Figure 1A). The proportion of methylcytosines in the samples was estimated to be about 0.67% of the total cytosines, depending on the proportion of each parasite stage in the asynchronous sample, by matrix effect-free external calibration (see Experimental Procedures and Supplemental Figure S1A

and S1B). The identity of me^{5}C was confirmed by mass spectrometric measurement, which revealed the characteristic *m/z* 242→126 transition, corresponding to the elimination of a 2-deoxyribose moiety from the $[M+H]^+$ ion of 5-methyl-2'-deoxycytidine (Figure 1B). We verified that methylcytosines were not significantly detected in non-infected red blood cells spiked-in with commercial unmethylated DNA. The data showed that only 0.0037% 5-mdC came from the background (Supplemental Figure S1C). Similar assays were performed in synchronized populations of *P. falciparum*. We found that 1.16 % (−0.11), 1.31 % (−0.04), and 0.36 % (−0.08) of the total cytosines are methylated at ring, trophozoite, and schizont stages, respectively (Supplemental Figure S1D). Whilst methylation levels at ring and trophozoite stages are comparable, there is a remarkable diminution of methylcytosine content at schizont stage. These observations may indicate that DNA methylation in *P. falciparum* only occurs *de novo* and is lost during replication, thus diluting the total methylcytosine content. Finally, drugresponse curves to hypomethylating drugs – we used the cytosine analogues 5-azacytidine and 5-aza-2'-deoxycytidine or decitabine – show that the parasite's viability is affected (Figure S2A). Similarly to the results obtained in acute myeloid leukemia cell lines (Hollenbach et al., 2010), decitabine half maximum inhibitory concentration (IC_{50}) is lower than 5-azacytidine IC_{50} whereas maximum viability reduction is higher with 5-azacytidine than with decitabine (about 80 % *vs.* 60 % reduction, respectively). These observations certainly reflect the time-restricted incorporation of decitabine at the time of DNA replication by comparison with a more continuous effect of 5-azacytidine, which can be incorporated in both DNA and RNA. These cytosine analogs cause the parasite's development to stop *in vitro* (Figure S2B). As a whole, these results suggest the presence of important methylation events in *P. falciparum*'s genome during the intra-erythrocytic stages.

Detection of C5-DNA methyltransferase activity in *P. falciparum*'s nucleus

We then investigated the presence of DNA cytosine methyltransferase (DNMT) activity in *P. falciparum* nuclear protein extract using an ELISA-like *in vitro* cytosine methylation assay (see Experimental Procedures). Methylation of cytosine-rich DNA coated on a 96-well plate was detected by fluorescence. DNMT activity was expressed in relative units of fluorescence per hour and per milligram of proteins (RFU/h/mg) and measured after 10 min (Figure 1C). The RFU obtained for the *Plasmodium* nuclear protein extract was 297 ± 31 RFU (± standard deviation), which is significantly different from background (blank = 97 ± 17 RFU, $n = 2$). Therefore, *P. falciparum* nuclear extracts showed significant DNMT activity. The weaker intensity of the signal when compared to the control (1 μg of purified bacterial DNA) is consistent with a complex mixture of proteins in the parasite's protein extract and therefore diluted levels of DNMT. The experiment was repeated in the presence of 100, 200, or 500 nM of the methyltransferase inhibitor hydralazine (Zambrano et al., 2005) and 100 nM of the rationally designed DNMT inhibitor RG108 (Brueckner et al., 2005). The signal was monitored for 15 min (Figure 1D). Our results indicate that the methyltransferase activity detected in *P. falciparum*'s nuclear extracts is sensitive to both hydralazine and RG108. These results support the presence of a functional DNMT in *P. falciparum*.

In silico identification of candidate C5-DNA methyltransferases

In eukaryotes, multiple families of DNMTs fulfill different functions and are regulated by different pathways (Goll and Bestor, 2005). To identify putative DNMTs in *P. falciparum*, we searched its genome for the presence of proteins that contain the DNA methylase PFAM domain (Bateman et al., 2004; Finn et al., 2006). For comparison, validation, and identification purposes, we added to our analysis four other *Plasmodium* species (*vivax*, *yoelii*, *chabaudi*, *berghei*), the Apicomplexa *Cryptosporidium parvum*, *Cryptosporidium hominis*, and *Toxoplasma gondii*, and the model eukaryotic organisms *Arabidopsis thaliana*,

Homo sapiens, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, and *Neurospora crassa*. We used a hidden Markov model-driven domain recognition approach (Eddy, 1998) and identified 31 putative DNMTs (Table S1). Among them, we identified PF3D7_0727300 in *P. falciparum* that is expressed during the erythrocytic cycle (Otto et al., 2010; Le Roch et al., 2003; expression was further confirmed by real-time RT-PCR, data not shown). For comparison purpose, quantitative mass spectrometry analyses find 2 spectral counts for our putative PfDNMT against 115 spectral counts for the constitutive elongation factor PF3D7_1451100 in schizont stage parasites (Bowyer et al., 2011).

We found various isoforms of the *de novo* DNMT1/3 in humans and their equivalent in plants, the *de novo* DNMT2 and the three plant-specific chromomethylases in *A. thaliana*, the two *N. crassa* DNMT Dim-2 and RIP defective, and one DNMT2 homolog in *S. pombe* that is believed to be inactive. No DNMT was found in *S. cerevisiae*, consistent with literature data. For each *Plasmodium* and *Cryptosporidium*, we found one candidate DNMT related to DNMT2 (Figure 2A). In *Toxoplasma gondii*, we found two proteins, one annotated as putative DNMT2 and one annotated as putative methyltransferase. Previous analyses, however, showed that only TGME49_027660 is expressed in *T. gondii*. None of the identified proteins fell in the DNMT2 group that contained *Plasmodium* candidate DNMTs.

We aligned the protein sequences for the group of putative DNMT2 (Figure 2B). In eukaryotes, the DNMT-specific motif IV contains a prolyl-cysteinyl dipeptide (or PC, Figure 2B, red arrow) that is necessary to enzyme activity. This motif IV is well conserved among *Plasmodium spp.* but is missing in *Cryptosporidium* and altered in *S. pombe* (Figure 2B, black frame). These observations are consistent with previous data showing that *Cryptosporidium*'s genome is not methylated (Gissot et al., 2008) and that the fission yeast's DNMT is not functional (Pinarbasi et al., 1996; Wilkinson et al., 1995). In *Plasmodium* species, however, the crucial PC dipeptide is complete and functional. Altogether our observations could indicate that PF3D7_0727300 may be an active DNMT responsible for the presence of me^5C in *P. falciparum*'s genome.

Cloning, expression and *in vitro* activity of the putative PfDNMT PF3D7_0727300

The DNMT domain of PF3D7_0727300 was cloned into pGS21a downstream of a GST-HIS tag, between SpeI and SacII restriction sites. This construction included the catalytic PC motif (or *complete domain*, Figure 2C). The resulting expression vector was then transfected and expressed into *E. coli*. After purification on glutathione-bound resin, the purified DNMT domain of PF3D7_0727300 was eluted and the presence of the GST tagged PfDNMT was verified by western blot anti-GST (Figure 2D). Two sharp bands are visible. The first one corresponds to the ~42 kDa protein domain combined to the GST tag. The second one corresponds to the 29.9 kDa GST tag only. The purified PfDNMT was immediately tested for *in vitro* DNMT activity. Activity was measured every minute for 10 min and expressed in RFU (Figure 2E, black line with black squares). Purified protein extracts showed significant DNMT activity. This activity was very strongly reduced in the presence of both hydralazine and RG108 (Figure 2E, dashed lines and dotted line, respectively). Similarly, we constructed a truncated version of our protein, in which the catalytic PC motif was missing (or *truncated domain*, Figure 2C), that showed reduced activity compared to the fully expressed DNMT domain (Figure 2E, black line with empty squares). These results strongly suggest that PF3D7_0727300 is a functional methyltransferase in *P. falciparum*.

As a whole, our data show that the malaria parasite's genome is methylated. We detected significant amounts of me^5C and we identified a potentially functional putative DNMT2-like enzyme encoded by PF3D7_0727300. To explore further the methylation status of *P. falciparum*, we analyzed the m^5C patterning in its genome by BS-seq.

Genome-wide mapping of cytosine *loci* methylated during the intra-erythrocytic cycle

Genomic DNA was extracted from an asynchronous population of *P. falciparum*-infected erythrocytes, bisulfite-treated and sequenced on a Genome Analyzer II platform. A total of 26,148,165 very high quality reads were aligned to the *P. falciparum* reference genome version 3 (Table S2) using the software BRAT in *bisulfite* mode (see Harris et al., 2010, and Supplemental Experimental Procedures). A total of 20,145,321 reads were mapped at a unique location with up to one sequencing error-related mismatch (*i.e.*, mismatches other than C/T, Table S2). A total of nearly 96% of the genome was covered by at least one read, with ~10.6× genome coverage (Figure S3 and Table S2), consistent with preliminary *in silico* simulations (Supplemental Experimental Procedures). More than two million cytosines were sequenced on each strand, which represents 92.6% of the total cytosines in the genome (Table S2). We used the apicoplast as a non-methylated internal reference and estimated an error rate of 1.41 % (*i.e.*, proportion of mismatches resulting from both incomplete bisulfite conversion and sequencing error) that we used for statistical filtering of false positives (FDR=0.05, Lister et al., 2009). Prior *in silico* tests indicated that the apicoplast genome is a suitable internal reference for measuring the non-conversion rate in the case of *P. falciparum* (see Table S2D, S2E S2F and Supplemental Experimental Procedures). In addition, the use of a commercial spike-in non-methylated DNA is currently not suitable in *P. falciparum* due to the high (A+T) content of its genome. Indeed, as pinpointed by Kruger and colleagues (Krueger et al., 2012), not all sequences have the same conversion properties and a spike-in control should have compatible nucleotide distributions. There is, to our knowledge, no commercial unmethylated AT-rich DNA that can be used for that purpose in *Plasmodium*. Finally, we used the multiple read counts at a given me^5C locus as a measure of the fraction of sequences that are methylated at that locus (*i.e.* the number of sequenced cytosine divided by the total read depth at the considered locus; Lister et al., 2008). Since our sample consisted in an asynchronous population of the parasite intra-erythrocytic stages, this measure at each methylcytosine locus reflects the fraction of the cycle when the considered locus is methylated.

We found that a total of 26,152 highly confident cytosine *loci* were methylated, representing 0.58% of the total genomic cytosine *loci* or 0.63% of the sequenced ones, consistent with our LC-MS/MS measurements. These values were within the ranges [0.47%; 0.85%] of the total genomic cytosines, and [0.56%; 1.21%] of the sequenced cytosines, for three independent biological replicates (independent populations of mixed intra-erythrocytic stages). For comparison, more than two million methylated cytosine *loci* were found in the genome of *A. thaliana* (about 5% of the total genomic cytosine *loci*, Lister et al., 2008), half of them being undetected by previous mass spectrometry, restriction enzymes and/or antibody-based analysis. We examined our results looking for the presence of the partial methylation previously discovered by Pollack and Colleagues (Pollack et al., 1991) in the bifunctional dihydrofolate reductase-thymidine synthase gene body (DHFR-TS). We did not find the expected methylation at position 749,117 on chromosome 4, but we identified three other methylated *loci* within the gene body, and one in the upstream region (Figure S4A). We re-examined the methylation status of position +1,030 of the DHFR-TS gene using methylation-sensitive/insensitive restriction enzymes and methylated DNA immunoprecipitation (MeDIP), in a bisulfite-independent manner (Figure S4Ba). Our results are consistent with Pollack and Colleagues' observations and indicate that this particular cytosine is methylated. Its absence from our BS-seq dataset could be related to a lower coverage on this particular *locus*, and may therefore be a false negative. We repeated the same experiments for other sites selected from our dataset and confirmed the presence of methylcytosines (Figure S4Ba). Finally, successful bisulfite conversion of unmethylated regions was confirmed by PCR (no restriction enzymes used, Figure S4C).

We monitored the distribution of me^{5}C along the *P. falciparum* chromosomes (Figure 3A and Figure S3B). Regions with higher me^{5}C content are distributed on the whole length of chromosomes, mostly asymmetrically. Levels of methylation are stable along the chromosomes, including in telomeric and sub-telomeric regions despite a higher GC content (Figure 3B). We further examined the context of genome-wide methylations and confirmed that 78% of them are asymmetrical (CHH context where H can be any nucleotide but G), the remaining 22% being almost equally distributed between CG (10%) and CHG methylations (12%; Figure 3C), consistent with the fact that most cytosines of *P. falciparum*'s genome are in CHH contexts regardless their methylation status (distribution conserved among biological replicates, data not shown). While DNA methylation occurs almost exclusively in a CG context in differentiated human cells, non-CG methylations were recently found in higher proportions (up to 45%) in plants and undifferentiated human cells (Cokus et al., 2008; Lister et al., 2008, 2009). For each context, we found that ~75–79% of the highly confident methylations occur during 1/3 of the cycle or lower (*i.e.*, frequencies ≤ 0.33, Figure 3D), suggesting that DNA methylation occurs exclusively *de novo* in *P. falciparum* and could be related to erythrocytic cycle progression. The low number of methylated cytosines in the parasite's genome and the very high proportion of CHH methylations explain why less sensitive classical sequence-biased methods failed to detect methylations in *P. falciparum*.

Sequence context and preferences of methylcytosines

We expanded our analyses of the context in which me^{5}C occurs to the neighboring bases (Figure S4D). In the case of unmethylated cytosines, surrounding positions are most often occupied by an adenine or a thymine, which reflects the extreme AT-richness of the *P. falciparum* genome, with a slight preference for thymines except at position +1 where adenines are more frequent (Figure S4D). Positions surrounding conserved me^{5}C, however, show a clear and significant preference for thymines, including at position +1, that is particularly marked at position −1 ($p < 0.01$, Figure S4D). In addition, sequences surrounding me^{5}C are generally depleted in guanines, and in cytosines when conserved methylations are considered (Figure S4D). Positions immediately surrounding cytosines seem, however, to behave differently: cytosines immediately followed by another cytosine are likely unmethylated, which could be explained by steric effects (Figure S4D). Such sequence preferences were previously observed in *A. thaliana* (Lister et al., 2008). In eukaryotes, sequence context specificity is driven by multiple parameters such as the different affinities of various DNA MTases, and histone tail methylation or interactions with RNA molecules. Since we identified only one putative DNMT in the entire genome (Figure 2), sequence context specificity is unlikely to be driven by enzymatic recognition. Nonetheless, specificity could be mediated by direct RNA-DNA interaction (Hawkins and Morris, 2008; Matzke et al., 2009), the recently suggested methylation-directing activity of introns (Dalakouras et al., 2009), or the complex relationship between histone modification and DNA methylation (Cedar and Bergman, 2009).

Our observations raise the question of a potential effect of sequence preference on the distribution of me^{5}C in various regions of the parasite's genome. *P. falciparum*'s coding regions are known to have a higher CG content than non-coding regions that are extremely AT-rich. We measured the proportion of me^{5}C that are found in the various compartments of the genome, *i.e.*, the exons, the introns, the putative promoters and terminators of genes (up to 1kb-long non-coding regions located upstream or downstream of the start or stop codons, respectively). We observed an increased distribution of me^{5}C in exons compared to the exonic GC-content (66.7 % of the genomic cytosines are found in exons; $p < 0.0001$) with 72% of the total me^{5}C being found within exons (Figure 4A). This over-representation of methylated cytosines in exons is consistent with recent data showing that intragenic DNA

methylation occurs at higher density in plants (Cokus et al., 2008). Similarly, cytosines tend to be more often methylated in exons (0.67% of the cytosines in average) than in non-coding regions (differences not statistically significant; (Figure 4B). Since *P. falciparum*'s exons are enriched in nucleosomes (Ponts et al., 2010), we paralleled me^5C marks to the positions of nucleosomes genome wide (Ponts et al., 2010). We examined the distribution and the intensity of DNA methylation in the neighborhood of all nucleosomes (Figure 4C and Figure S4E). We observed an oscillation of the average level of methylation, local minima being phased with the boundaries of the nucleosome-bound regions consistent with the marking of DNA with methylations and nucleosome positioning being tightly linked (Pennings et al., 2005). When both DNA strands are considered separately, we observe that methylations clearly localize on the template strand within the gene body whereas no general strand-specificity is visible within flanking non-coding regions (Figure 4D and Figure S4F). Such a strand specificity of DNA methylation patterns could have major consequences on the affinity of the RNA polymerase II for the template DNA, and directly impact the speed of transcription elongation to facilitate the inclusion of constitutive exons during splicing, as seen in various eukaryotes (Zilberman et al., 2007).

Exons/introns distribution of methylcytosines and gene expression

We deepened our analysis and examined methylation levels at the extremities and the exon/intron boundaries of *P. falciparum*'s genes. We found that the extremities of the gene body are marked by DNA methylation (Figure S5A). In particular, the start codon and the end of genes appear hypermethylated. Such results are consistent with previous methylation patterns established in murine and human genes, for which hypermethylations of genes' 3' ends reduce transcription elongation efficiency (Choi et al., 2009). Recently, hypermethylation of start/stop codons was suspected to secure the first and last exon from exon skipping during splicing to ensure accurate translation (Choi et al., 2009). At exon/intron boundaries, we found splice junctions to be more methylated on both 5' and 3' ends of introns (Figure 5 and Figure S5B), which is consistent with a role for DNA methylation in splicing. A similar pattern was recently observed in human embryonic cells at 5' splicing junction sites, the 3' splicing junction site being nonetheless strongly hypomethylated (Laurent et al., 2010). In *P. falciparum*, we further observed that these strong methylations occur almost exclusively on the template strand, the sense strand presenting less variation across the considered region. One hypothesis is that strand-specific hypermethylations of splicing sites can regulate alternative splicing in *P. falciparum* while constitutive exons are secured from exon skipping by hypermethylation of coding regions that reduces the speed of transcription elongation.

We further investigated the relationship between DNA methylation and transcription regulation. Previous observations showed that intragenic DNA methylation could inhibit gene expression in plants (Hohn et al., 1996). We therefore examined the methylation level of every first exon and compared it to the mRNA levels measured by Le Roch and colleagues (2003) for genes weakly or highly expressed during the intra-erythrocytic cycle (Le Roch et al., 2003). We found a negative relationship between methylation of the first exon and mRNA abundance: highly expressed genes appear hypomethylated whereas weakly expressed genes are hypermethylated (Figure 6A and Figure S6). In a general manner, methylation is more intense within the first half of the first exon, on the template strand (Figure 6B). When genes with high and low expression levels are considered separately, methylation is more intense on the sense strand and on the second half of the first exon only when weakly expressed genes are considered (Figure 6C). These results suggest that intragenic methylation could regulate gene expression in the malaria parasite, the 5' flanking region being always hypomethylated.

DNA methylation and other epigenetic marks

In eukaryotes, acetylation of histone H3 on lysine 9 (H3K9Ac) and methylation of lysine 4 (H3K4me3) are epigenetic marks associated to euchromatin whereas methylation of lysine 9 (H3K9me3) and methylation of lysine 20 on histone H4 (H4K20me3) are involved in gene silencing. We examined the methylation status of previously published regions containing the active marks H3K9Ac and H3K4me3, spread genome-wide, the silencing mark H4K20me3 (also broadly spread across the genome), and the silencing mark H3K9me3, localized in the sub-telomeric regions and associated with the silencing of clonally variant genes involved in virulence (Lopez-Rubio et al., 2009). We find that regions containing the permissive marks H3K9Ac and H3K4me3 have very similar methylation profiles; on the template strand, methylation levels tend to increase downstream of the mark (Figure 7A and C, and Figure S7A and C). The strict restriction of the silencing mark H3K9me3 to virulence genes in sub-telomeric regions is consistent with a general transcriptionally active state of the parasite's genes. We find that regions surrounding the repressive marks H3K9me3 and H4K20me3 are associated to a fairly constant methylated state regardless the considered strand (Figure 7B and D, and Figure 14B and D). These observations may suggest that DNA methylation and H3K9me3 could be linked and together participate in the silencing of the parasite's virulence genes. Since knockout experiments demonstrated that the activity of the characterized histone deacetylase PfSIR2 alone is not sufficient to account for all the H3K9me3, DNA methylation could be another mechanism acting instead of (or together) with another deacetylase.

DISCUSSION

The present study identifies methylcytosines in an AT-rich genome. Its success was enabled by the advent of unbiased bisulfite conversion coupled with deep sequencing. We demonstrate that the genome of *P. falciparum* is methylated. In addition, we identify a functional DNMT, PF3D7_0727300, with homologues in other *Plasmodium* species. These findings provide insights for the previously unexplained higher mutation rate observed in regions of high GC content (Carlton et al., 2008), possibly due to deamination of methylcytosines into thymines (Lutsenko and Bhagwat, 1999). In addition, we demonstrate that non-CG methylations, generally overlooked by other methods, can be of major importance for the regulation of transcription elongation, splicing, or silencing of virulence genes.

Our results are consistent with previous nucleosome positioning data showing that the parasite's genome is mostly hard-wired in a transcriptionally permissive state, with the exception of the sub-telomeric regions that contain silenced virulence genes (Ponts et al., 2010). *Plasmodium*'s genome active state seems to be mostly maintained by epigenetic marks such as DNA methylation, nucleosome positioning, and post-translational histone modifications. Such a pattern of hyperactive transcription has been previously observed in pluripotent embryonic stem cells and contributes to their plasticity (Efroni et al., 2008), which could indicate that *P. falciparum* is a transcriptionally undifferentiated cell throughout its intraerythrocytic cycle. Previous work has already suggested that post-transcriptional regulations, such as mRNA stability, could be involved in the differentiation of the parasite into its different intra-erythrocytic morphological stages (Shock et al., 2007). In this model, a transition in methylation levels and/or nucleosome landscape could occur during sexual differentiation. Gametocyte differentiation is indeed known to be a general response to a wide array of *stimuli* (*e.g.* drug-induced stress) and is mediated by an arrest of the erythrocytic cycle and the transcriptional activation gametocytogenesis genes (Le Roch et al., 2008). Such behavior mimics the refinement of transcription events during pluripotent cell differentiation (Efroni et al., 2008).

Finally, we found striking genome-wide strand specificity in *P. falciparum* that, to our knowledge, has not been described anywhere else. Partial strand-specificity has nonetheless been previously observed in *A. thaliana*'s centromeric regions (Luo and Preuss, 2003), which are hypomethylated by comparison with the rest of the genome. Strand specificity could nevertheless exist in other organisms where asymmetrical methylation-related features may be embedded within CG-related patterns and could be harder to detect. Future applications of such works could reshape the current knowledge of the methylation status of many different organisms. Our work opens perspectives in the field of epigenetics in general, and infectious disease in particular.

EXPERIMENTAL PROCEDURES

P. falciparum strain and culture conditions

P. falciparum parasite strain 3D7 was maintained in human erythrocytes according to previously described protocols (Le Roch et al., 2003; Trager and Jensen, 1976). Cultures were harvested at 8% parasitemia.

DNA Digestion and LC-MS/MS analysis

Genomic DNA from *Plasmodium* (50 μg) was denatured by heating to 95°C and chilled immediately on ice. Denatured DNA was subsequently digested with two units of nuclease P1 in a buffer containing 30 mM sodium acetate (pH 5.5) and one mM zinc acetate at 37°C for 4 h. To the digestion mixture were then added 25 units of alkaline phosphatase in a 50 mM Tris-HCl pH 8.6. The digestion was continued at 37°C for 2.5 h, and the enzymes were removed by chloroform extraction. The aqueous DNA layer was dried using a Speed-Vac and re-dissolved in water. The amount of nucleosides in the mixture was quantified by UV spectrometry. Analysis of me^5C in the DNA hydrolysates was performed by online capillary HPLC-ESI-MS/MS using an Agilent 1200 capillary HPLC pump interfaced with an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, USA). A 0.5×150 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies) was used for the separation of the DNA hydrolysis mixture with a flow rate of 12.0 μL/min. A mixture consisting of 50 pmol of total nucleosides from the enzymatic digestion of *Plasmodium* gDNA or 2 nmol of total nucleosides from control unmethylated DNA was injected in each analysis. A gradient of 0–90% methanol (in 10 min) followed by 90% methanol (in 5 min) in 0.1% aqueous solution of formic acid was employed. The effluent from the LC column was directed to the LTQ mass spectrometer, which was set up for monitoring the fragmentation of the protonated ions of me^5C (m/z 242). The area for peak found in the selected-ion chromatograms for monitoring the m/z 242→126 transition, which correspond to the elimination of a 2-deoxyribose from me^5C, was then determined.

Extraction of nuclear proteins

5.10^9 parasites were harvested in PBS and released from their host red blood cells by saponin lysis. After 15 min of incubation on ice, parasites were resuspended in 1 mL of cytoplasm lysis buffer (20 mM HEPES pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.5 mM AEBSF, 0.65% Igepal®, Roche Complete cocktail protease inhibitor) and lysed on ice for 5 min. Nuclei were pelleted by 10 min of centrifugation at 1500g and 4°C, washed three times with ice-cold PBS, and resuspended in 100 μL of nuclei lysis buffer (20 mM HEPES pH 7.9, 0.1 M NaCl, 0.1 mM EDTA, 0.1 mM EGTA, 1.5 mM MgCl2, 1 mM DTT, 25% glycerol, 1 mM AEBSF, Roche Complete cocktail protease inhibitor). Nuclear extracts were obtained after 20 min of lysis at 4°C with vigorous shaking, and clearing by 10 min of centrifugation at 6000g and 4°C. Protein content was quantified by Bradford assay and DNMT activity was measured immediately after extraction.

DNA methyltransferase assays

DNMT activity of protein extracts was measured using the EpiQuik™ DNA methyltransferase activity kit (Epigentek cat # P-3004, fluorometric) following the manufacturer's instructions. Activity was measured every minute for 10 min. Assays were realized in triplicate on two independent sample preparations (nuclear protein extracts or purified domains of PF3D7_0727300). The positive control (1 μg of purified bacterial DNMT) and blank (buffer only, used for background subtraction) were run in duplicate. DNMT activity was expressed in relative units of fluorescence per hour and per milligram of proteins (RFU/h/mg).

In silico search for putative DNMTs

The program hmmsearch (HMMER v3.0; Eddy, 1998) was used to extract sequences that carry the Pfam DNMT domain (PFAM v22.0 accession number PF00145; Bateman et al., 2004; Finn et al., 2006, 2010) from the genomes of the studied organisms (E-value ≤ 0.1). Protein sequences were aligned with MUSCLE (Edgar, 2004a, 2004b) and a tree was built using the Neighbor-Joining method (1000 bootstraps).

Extraction of gDNA, library preparation and bisulfite conversion

Parasite cultures were harvested at 50% haematocrit in PBS and three volumes of Cell Lysis Solution (cat. # A7933, Promega Corporation, Madison, USA) were added. After 10 min of incubation at room temperature, parasites were precipitated by 15 min of centrifugation at 2000 g. Cell lysis was performed in 400 μL of parasite lysis buffer (guanidine HCl 3.75 M, SDS 0.625% v/v, proteinase K 250 μg/mL) for 30 min at 55°C and then overnight at 4°C. DNA was extracted with phenol-chloroform followed by ethanol precipitation and RNase A treatment. Twenty micrograms of DNA were solubilized in 400 μL of TE and sheared by sonication into fragments ranging from 50 to 500 bp. Five micrograms of sheared DNA was processed using the Illumina® Paired-End Sample Prep Kit in which we substituted the regular adaptors with the Illumina® Early Access Methylation Adapter Oligo. Library preparation was performed according to the manufacturer's instructions with some modifications. First, on-gel fragment size selection was performed at room temperature. DNA was then bisulfite-converted twice using the EpiTect QIAGEN kit. Libraries were amplified with 18 cycles of PCR using a blend of the polymerases TaKaRa Ex Taq™ (Clontech) and Platinum® High Fidelity Polymerase (Invitrogen) with an elongation temperature of 62°C. Sequencing was performed on an Illumina® Genome Analyzer II (paired-end reads 2×26bp) at the Institute for Integrative Genome Biology (UC Riverside). Raw sequence data files for this study are available at the Short Read Archive under the accession number SRA026090.

Mapping to the reference genome and identification of me^5C

Reference genome was downloaded from the malaria resource PlasmoDBv9.2 (http://www.plasmoDB.org). Paired-end reads were mapped to the reference genome with BRAT (Harris et al., 2010). Only reads that matched a single locus in the genome, up to one non bisulfite-related mismatch, were used in our analysis. Methylcytosine identification was performed according to a previously published method (Lister et al., 2008, 2009) for a maximum false discovery rate of 0.05 for nucleotides with a read coverage within the average read coverage per base ± two standard deviations, and equal to at least five reads. Positive and negative strands were analyzed separately. Biological variation was estimated from three independent samples for which, due to lower read coverage per base in average, the minimum coverage threshold was lowered to two (replicates 1 and 2) and one (replicate 3) reads so that about two third of the total genomic cytosines are analyzed and the dataset remains representative of the entire genome. At each significant me^5C site, the methylation

level was estimated from the number of reads that carry an unconverted cytosine divided by the number of reads that carry a thymine, that is to say a converted cytosine (me^5C/C).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Thomas Girke, Tyler Backman, Rebecca Sun, and Barbara Walter (IIGB UC Riverside) for assistance with Illumina® sequencing and Pipeline analysis. They also thank Dr. Felix Krueger (Babraham Institute, UK) for his expertise and Vance C. Huskins for his attentive proofreading. The study was supported by the National Institute of Allergy and Infectious Diseases and the National Institutes of Health #R01AI085077 (KLR), #R01CA101864 (YW) and T34GM062756, by NSF-CAREER IIS-0447773 (SL) and NSF IIS-1302134 (SL) and by the Human Frontier Science Program # LT00507/2011-L (EMB).

REFERENCES

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, et al. The Pfam protein families database. Nucleic Acids Res. 2004; 32:D138–D141. [PubMed: 14681378]

Bowyer PW, Simon GM, Cravatt BF, Bogyo M. Global profiling of proteolysis during rupture of Plasmodium falciparum from the host erythrocyte. Mol. Cell Proteomics. 2011; 10:M110.001636. [PubMed: 20943600]

Boyko A, Kovalchuk I. Epigenetic control of plant stress response. Environ Mol Mutagen. 2008; 49:61–72. [PubMed: 17948278]

Brueckner B, Garcia Boy R, Siedlecki P, Musch T, Kliem HC, Zielenkiewicz P, Suhai S, Wiessler M, Lyko F. Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. Cancer Res. 2005; 65:6305–6311. [PubMed: 16024632]

Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 2009; 10:295–304. [PubMed: 19308066]

Choi JK, Bae J-B, Lyu J, Kim T-Y, Kim Y-J. Nucleosome deposition and DNA methylation at coding region boundaries. Genome Biol. 2009; 10:R89. [PubMed: 19723310]

Choi S-W, Keyes MK, Horrocks P. LC/ESI-MS demonstrates the absence of 5-methyl-2'-deoxycytosine in Plasmodium falciparum genomic DNA. Mol. Biochem. Parasitol. 2006; 150:350–352. [PubMed: 16934885]

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008; 452:215–219. [PubMed: 18278030]

Dalakouras A, Moser M, Zwiebel M, Krczal G, Hell R, Wassenegger M. A hairpin RNA construct residing in an intron efficiently triggered RNA-directed DNA methylation in tobacco. Plant J. 2009; 60:840–851. [PubMed: 19702668]

Eddy SR. Profile hidden Markov models. Bioinformatics. 1998; 14:755–763. [PubMed: 9918945]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004a; 32:1792–1797. [PubMed: 15034147]

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004b; 5:113. [PubMed: 15318951]

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RDG, Buetow KH, et al. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell. 2008; 2:437–447. [PubMed: 18462694]

Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al. Pfam: clans, web tools and services. Nucleic Acids Res. 2006; 34:D247–D251. [PubMed: 16381856]

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. The Pfam protein families database. Nucleic Acids Res. 2010; 38:D211–D222. [PubMed: 19920124]

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. 2002; 419:498–511. [PubMed: 12368864]

Gissot M, Choi S-W, Thompson RF, Greally JM, Kim K. Toxoplasma gondii and Cryptosporidium parvum lack detectable DNA cytosine methylation. Eukaryotic Cell. 2008; 7:537–540. [PubMed: 18178772]

Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 2005; 74:481–514. [PubMed: 15952895]

Harris EY, Ponts N, Levchuk A, Roch KL, Lonardi S. BRAT: bisulfite-treated reads analysis tool. Bioinformatics. 2010; 26:572–573. [PubMed: 20031974]

Hawkins PG, Morris KV. RNA and transcriptional modulation of gene expression. Cell Cycle. 2008; 7:602–607. [PubMed: 18256543]

Hohn T, Corsten S, Rieke S, Müller M, Rothnie H. Methylation of coding region alone inhibits gene expression in plant protoplasts. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:8334–8339. [PubMed: 8710871]

Hollenbach PW, Nguyen AN, Brady H, Williams M, Ning Y, Richard N, Krushel L, Aukerman SL, Heise C, MacBeth KJ. A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. PLoS ONE. 2010; 5:e9001. [PubMed: 20126405]

Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. Nat. Methods. 2012; 9:145–151. [PubMed: 22290186]

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, et al. Dynamic changes in the human methylome during differentiation. Genome Research. 2010; 20:320–331. [PubMed: 20133333]

Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993; 366:362–365. [PubMed: 8247133]

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008; 133:523–536. [PubMed: 18423832]

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–322. [PubMed: 19829295]

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. 2009; 5:179–190. [PubMed: 19218088]

Luo S, Preuss D. Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:11133–11138. [PubMed: 12960391]

Lutsenko E, Bhagwat AS. Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications. Mutat. Res. 1999; 437:11–20. [PubMed: 10425387]

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJM. RNA-mediated chromatin-based silencing in plants. Curr. Opin. Cell Biol. 2009; 21:367–376. [PubMed: 19243928]

Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Böhme U, Lemieux J, Barrell B, Pain A, Berriman M, et al. New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Mol. Microbiol. 2010; 76:12–24. [PubMed: 20141604]

Pennings S, Allan J, Davey CS. DNA methylation, nucleosome formation and positioning. Brief Funct Genomic Proteomic. 2005; 3:351–361. [PubMed: 15814025]

Pinarbasi E, Elliott J, Hornby DP. Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. J. Mol. Biol. 1996; 257:804–813. [PubMed: 8636983]

Pollack Y, Katzen AL, Spira DT, Golenser J. The genome of Plasmodium falciparum. I: DNA base composition. Nucleic Acids Res. 1982; 10:539–546. [PubMed: 6278419]

Pollack Y, Kogan N, Golenser J. Plasmodium falciparum: evidence for a DNA methylation pattern. Exp Parasitol. 1991; 72:339–344. [PubMed: 1851101]

Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG. Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res. 2010; 20:228–238. [PubMed: 20054063]

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 2003; 301:1503–1508. [PubMed: 12893887]

Le Roch KG, Johnson JR, Ahiboh H, Chung DW, Prudhomme J, Plouffe D, Henson K, Zhou Y, Witola W, Yates JR, et al. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum. BMC Genomics. 2008; 9:513. [PubMed: 18973684]

Shock JL, Fischer KF, DeRisi JL. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. Genome Biol. 2007; 8:R134. [PubMed: 17612404]

Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. Methods Mol Biol. 2009; 507:3–20. [PubMed: 18987802]

Trager W, Jensen JB. Human malaria parasites in continuous culture. Science. 1976; 193:673–675. [PubMed: 781840]

Wilkinson CR, Bartlett R, Nurse P, Bird AP. The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res. 1995; 23:203–210. [PubMed: 7862522]

Zambrano P, Segura-Pacheco B, Perez-Cardenas E, Cetina L, Revilla-Vazquez A, Taja-Chayeb L, Chavez-Blanco A, Angeles E, Cabrera G, Sandoval K, et al. A phase I study of hydralazine to demethylate and reactivate the expression of tumor suppressor genes. BMC Cancer. 2005; 5:44. [PubMed: 15862127]

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. 2007; 39:61–69. [PubMed: 17128275]

HIGHLIGHTS

- The malaria genome is methylated.
- Transcript levels correlate with exonic rather than promoter methylation levels.
- Methylcytosines are particularly found at exon-intron boundaries.
- The positions of methylcytosines correlate with those of nucleosomes.

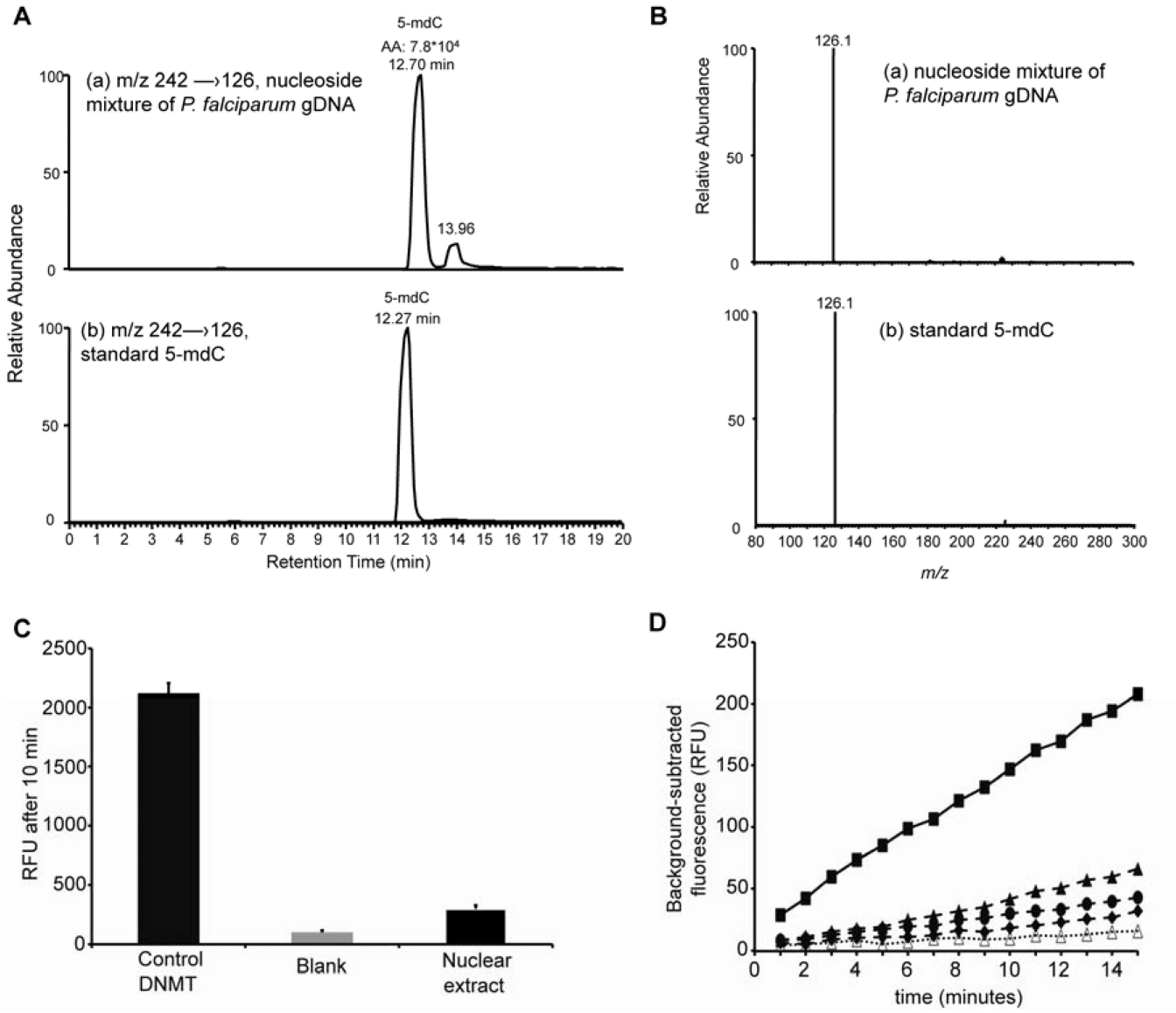

Figure 1. Biochemical evidences of the presence of DNA methylations

(A and B) LC-MS/MS detection of me^{5}C. (A) Selected-ion chromatograms for monitoring the *m/z* 242→126 transition (corresponding to the loss of a 2-deoxyribose) obtained from LC-MS/MS with the injection of: *(a)* 2.16 nmol of total nucleosides from the enzymatic digestion mixture of *Plasmodium* gDNA; *(b)* standard me^{5}C (5-mdC). The integrated peak areas mentioned above the 5-mdC peak show the presence of 5-mdC in the sample. (B) Tandem mass spectra (MS/MS) for monitoring the fragmentation of the $[M+H]^+$ ion (*m/z* 242) of 5-mdC averaged from the 5-mdC peak in Figure 1A, *i.e.*, the 12.70 and 12.27 min peaks in panels *(a)* and *(b)*. See also Figure S1. (C and D) DNMT activity in nuclear protein extracts. (C) Measurement of relative fluorescence units (RFU; mean ± SD) after 10 min of incubation for reactions performed with 1 μg of purified bacterial DNMT (Control DNMT), 10 μg of full *Plasmodium* nuclear protein extract, or buffer only (blank). (D) DNMT activity of 10 μg of full *Plasmodium* nuclear protein extract in the presence of hydralazine (dashed lines) 100 nM (full triangles), 200 nM (full circles), 500 nM (full diamonds), RG108 100 nM (dotted line, empty triangles), or without inhibitor (plain line, full squares). DNMT activity was expressed in RFU/h/mg after background subtraction. Our results demonstrate

the presence of DNMT activity in the nucleus of *P. falciparum*, consistent with the presence of me^5C. See also Figure S1

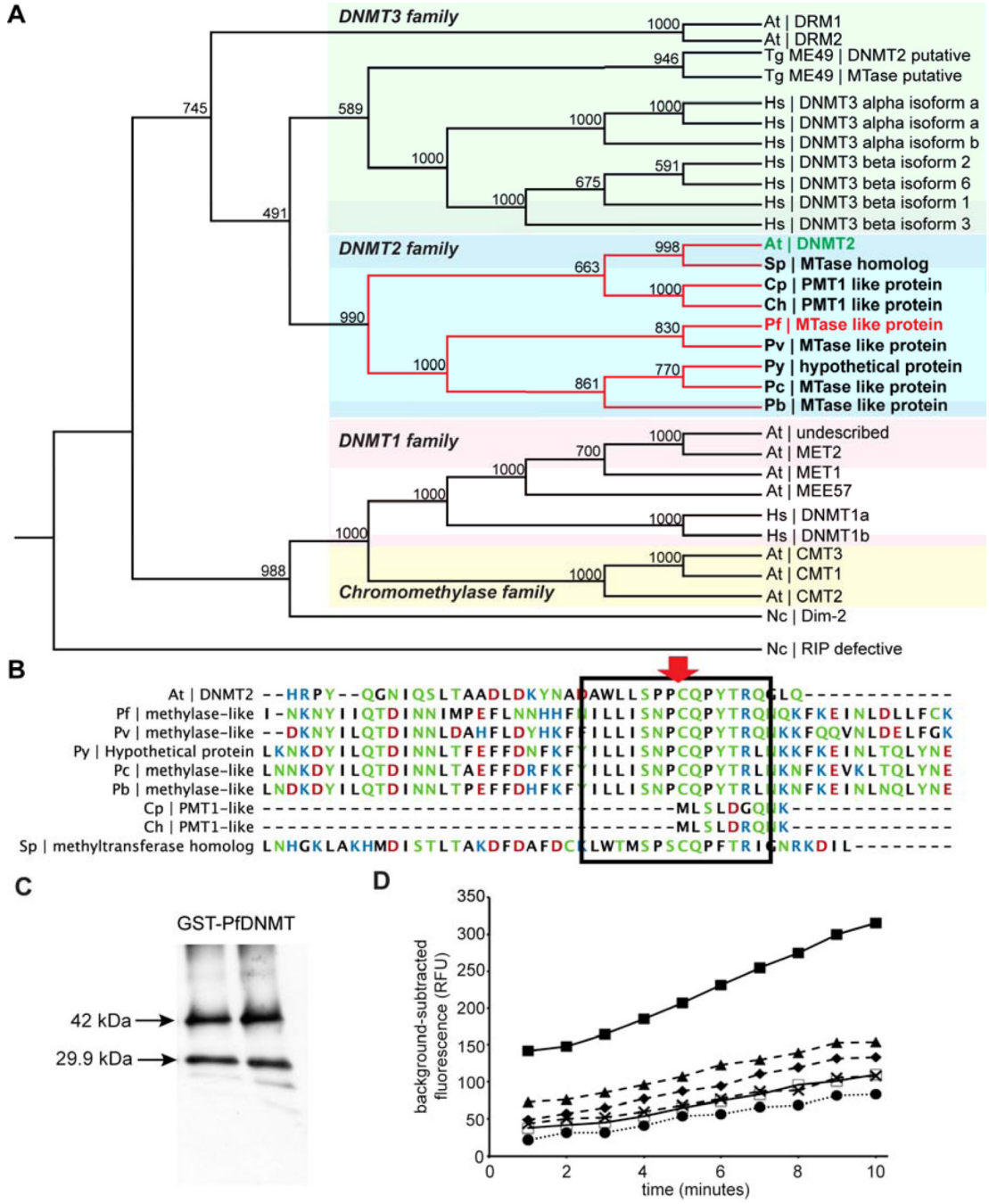

Figure 2. Identification of a functional C5-DNA methyltransferase

(A and B) *In silico* identification of candidate DNMTs. The genome of *A. thaliana* (At), *H. sapiens* (Hs), *S. pombe* (Sp), *N. crassa* (Nc), *T. gondii* (Tg), *Cryptosporidium spp. parvum* (Cp) and *hominis* (Ch), and *Plasmodium spp. falciparum* (Pf), *vivax* (Pv), *yoelii* (Py), *berghei* (Pb), and *chabaudi* (Pc) were investigated (See Table S1 for accession numbers). (A) Phylogenetic tree of the identified DNMT. Bootstrap values are indicated on the branches. PF3D7_0727300 was identified as a putative PfDNMT2. (B) Multiple alignments of the DNMT2 family of proteins. The conserved DNMT motif IV is highlighted with a black frame. The red arrow shows the presence of the catalytic prolyl-cysteinyl (PC)

dipeptide in all *Plasmodium* and its absence in *Cryptosporidium* and *S. pombe*. (C and D) Validation of PF3D7_0727300 as a functional DNMT. (C) Two constructions were prepared: the cloned complete domain included the PC motif whereas the cloned truncated domain did not contain the catalytic PC motif. (D) The DNMT domain of PF3D7_0727300 was GST-tagged, expressed, and purified. The presence of a 42 kDa, representing the protein domain combined to the GST tag, was resolved by SDS-PAGE and revealed by anti-GST western blot. The tag only is also visible at 29.9 kDa. (E) The purified domain was tested for a DNMT activity by fluorimetric ELISA-like assay in the presence of the inhibitors hydralazine (dashed lines) 100 nM (full triangles), 200 nM (full diamonds), 500 nM (black ×), RG108 100 nM (dotted line, full circles), or without inhibitor (plain line, full squares). The truncated domain (missing the PC motif) was also tested for DNMT activity (plain line, empty squares). Activities were measured every minute for 10 min and expressed in RFU/h/mg of protein. The DNMT domain of PF3D7_0727300 containing a PC motif can effectively methylate cytosines. See also Figure S2 and Table S1.

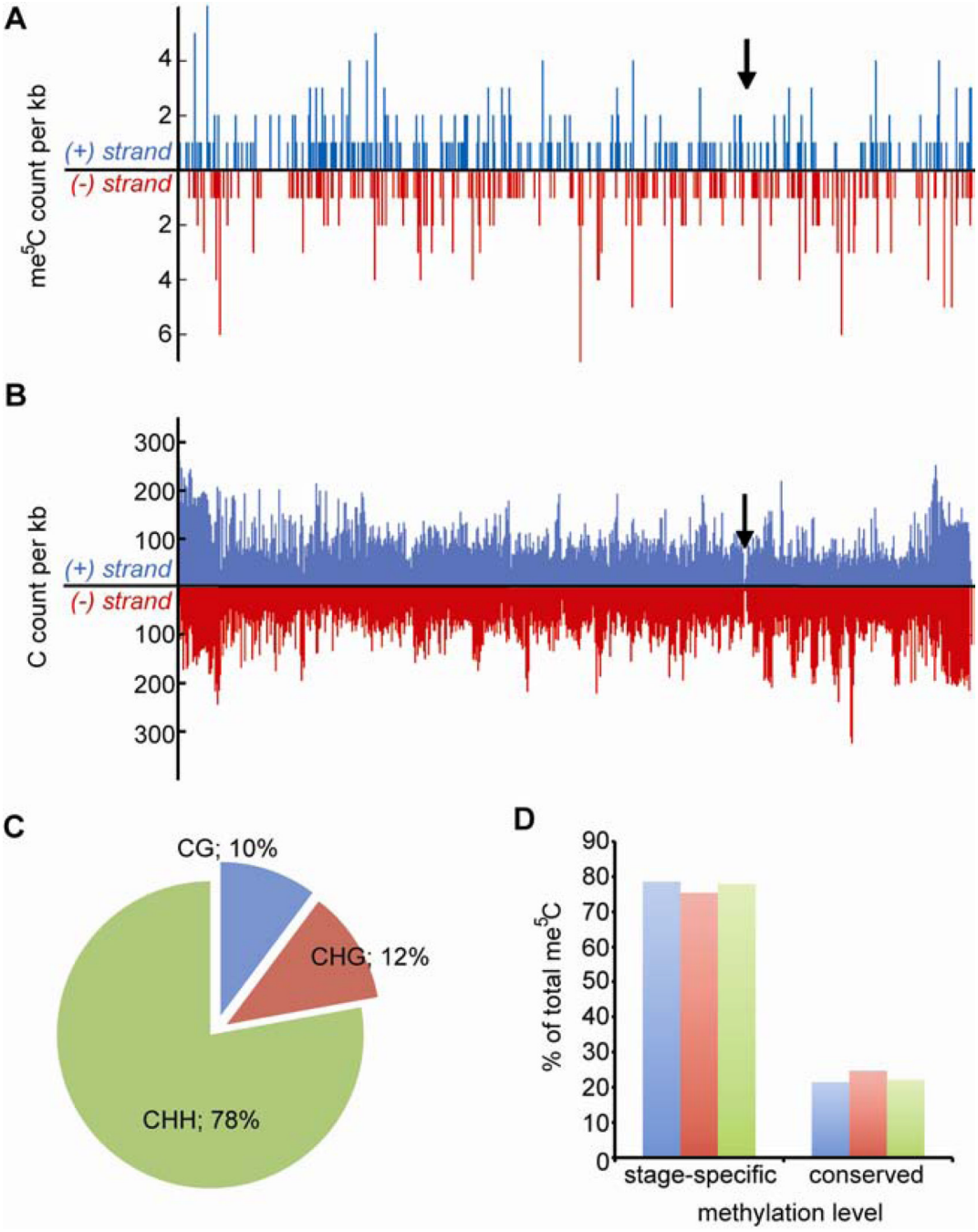

Figure 3. Methylation status of *P. falciparum*'s genome during the intraerythrocytic cycle

(A) Density profile of me^{5}C content in chromosome 1. The total number of me^{5}C found in 1kb-long non-overlapping windows was counted for each strand. Blue = positive strand; red = negative strand. The arrow shows the position of the centromere. See also Figure S3B. (B) CG-content of chromosome 1. The total number of cytosines was counted on each strand using 1kb-long non-overlapping windows. (C) Methylation context distribution of me^{5}C. The number of me^{5}C present in all possible contexts, *i.e.*, CG, CHG and CHH, was counted. (D) Distribution of me^{5}C according to their level of methylation, for each context. Stage specific = frequency at *locus* not exceeding 0.33; Conserved = frequency at *locus* above

0.33. Blue = CG context; red = CHG context; green = CHH context. See also Figure S3 and Table S2.

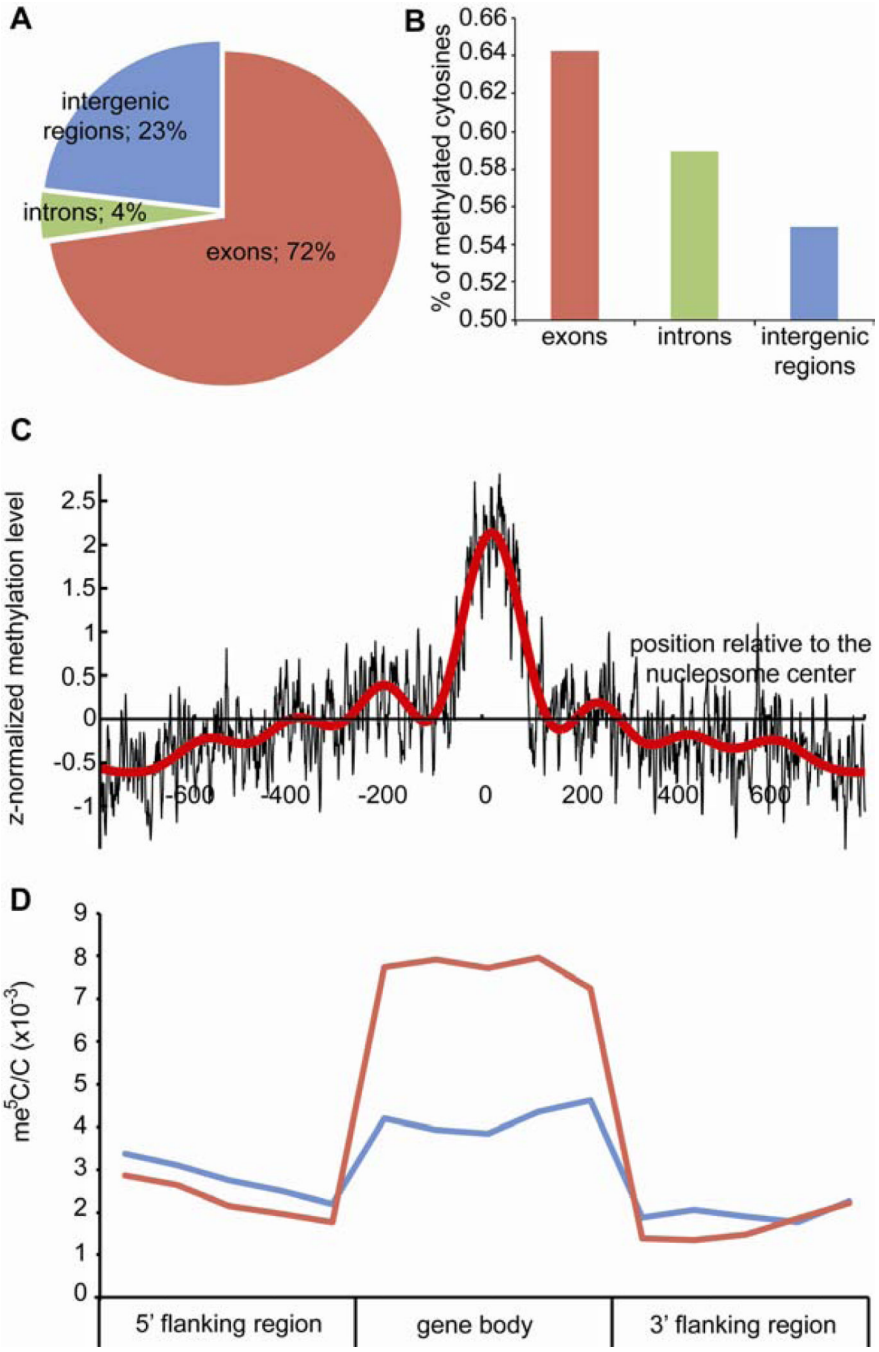

Figure 4. Genomic distribution of methylcytosines

(A) Repartition of me^5C within different compartments of the genome. (B) Proportion of methylated cytosines within each compartment of the genome. (C) Methylation status of nucleosomal DNA. For each position in the region spanning 1600bp around the center of nucleosomes (Ponts et al., 2010), me^5C/C are averaged and z-normalized (black curve; red curve = Fourier transform of the profile). All replicates are considered. The hypomethylated region spans ~40–80bp around the central position, which is the length of the DNA fragment tightly bound to the histone surface (Brower-Toland et al., 2002). See also Figure S4E. (D) Strand specificity of intragenic me^5C (all biological replicates). All genes are considered. Flanking regions and gene bodies are divided into five bins. For each bin, me^5C/C are

normalized by the size of the bin and averaged among all genes. Red = template strand; blue = non-template strand. See also Figure S4F.

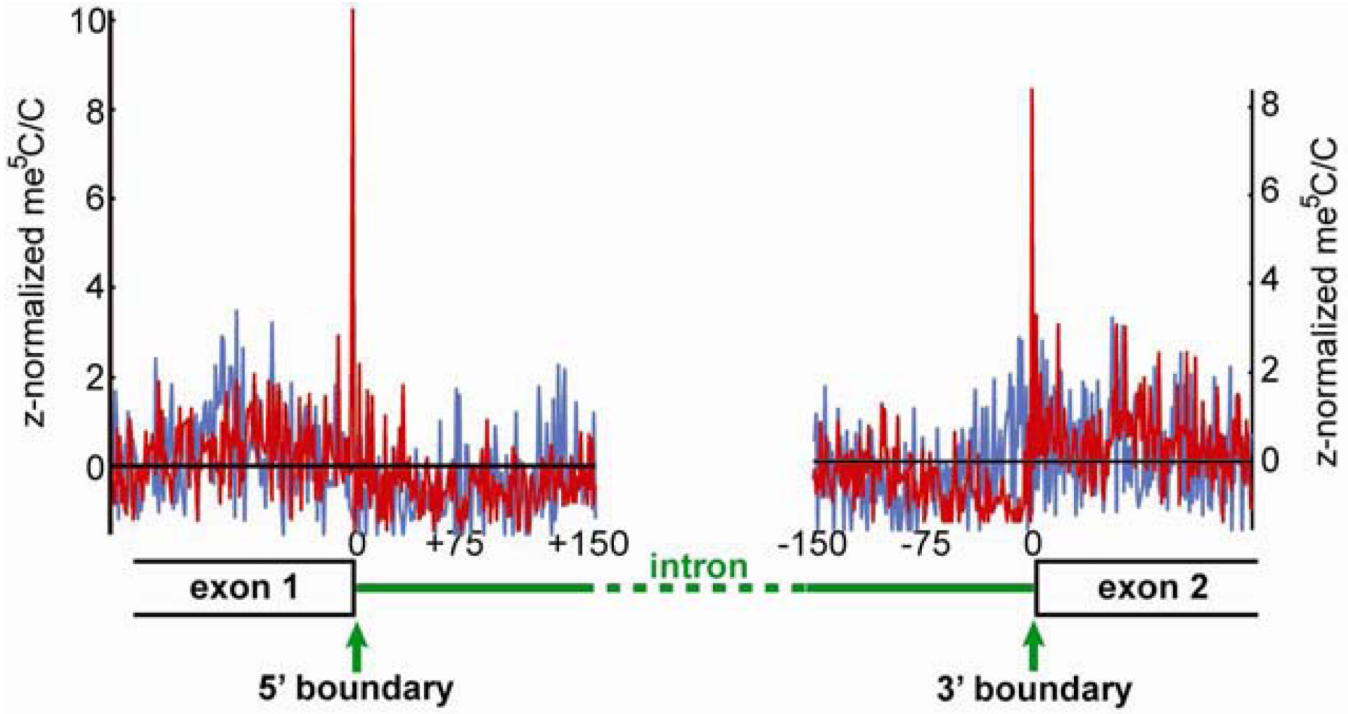

Figure 5. Methylation status of exon/intron boundaries

Methylation status in regions spanning 150bp around 5' and 3' splicing junctions. All exon/intron 5' and 3' junctions are considered. All me^5C/C ratios measured for each position around each exon/intron junction are averaged and z-normalized position-wise (red = template strand; blue = non-template strand). See also Figure S5.

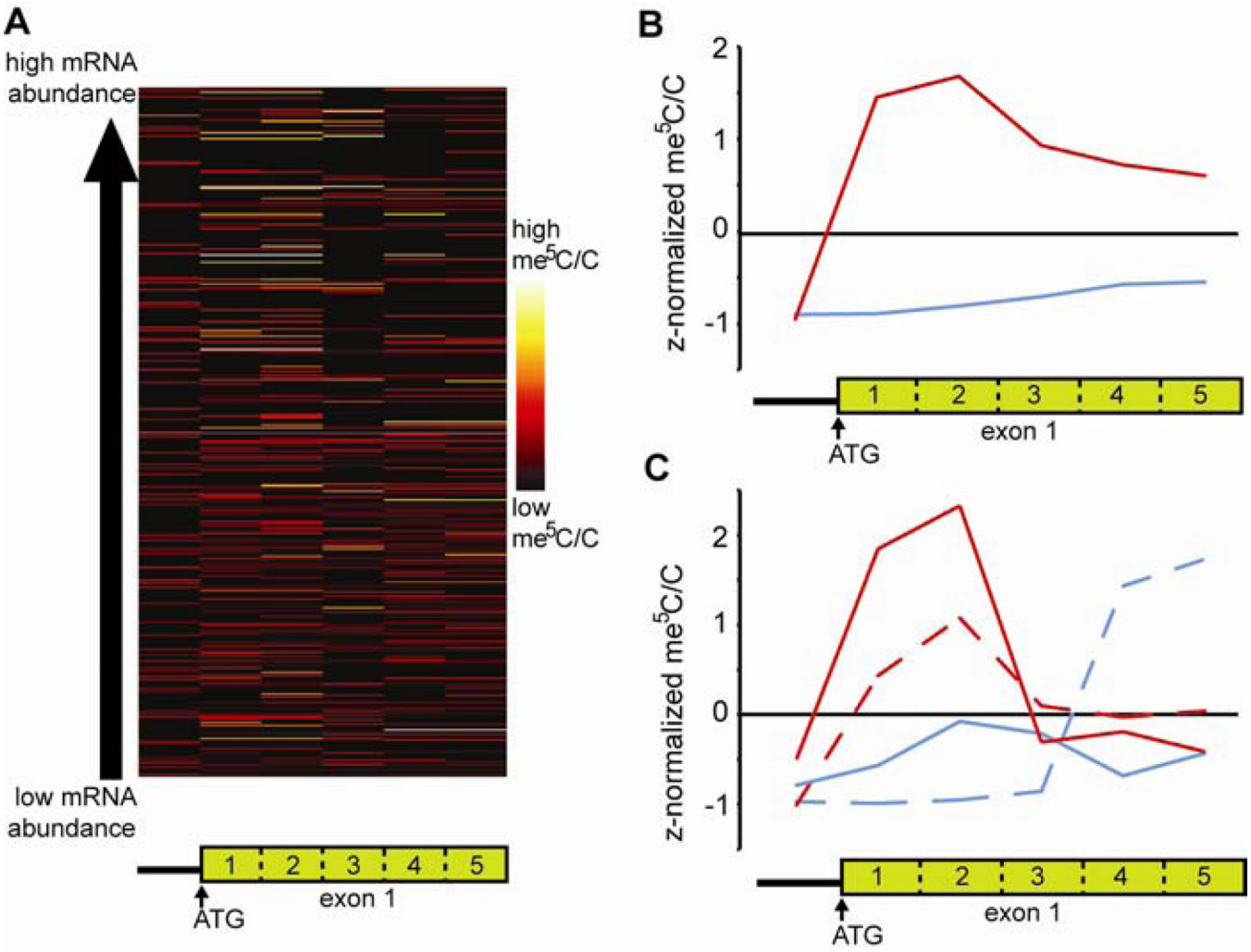

Figure 6. DNA methylation and gene expression

(A) Methylation levels of first exon and mRNA abundance. Highly expressed (95th percentile) and weakly expressed (5th percentile) genes were retrieved and ranked according to their average mRNA abundances across the erythrocytic cycle (Le Roch et al., 2003). First exons were binned into five bins. For each bin, me^{5}C/C are normalized by the size of the bin and z-scored. The representation uses a color scale from black (low methylation) to white (high methylation). See also Figure S6. (B) Average methylation levels of each bin among all genes (red = template strand; blue = non-template strand). (C) Average methylation levels of each bin among selected genes (red = template strand; blue = non-template strand; plain lines = highly expressed; dashed lines = weakly expressed). For each position, values are z-normalized. See also Figure S6.

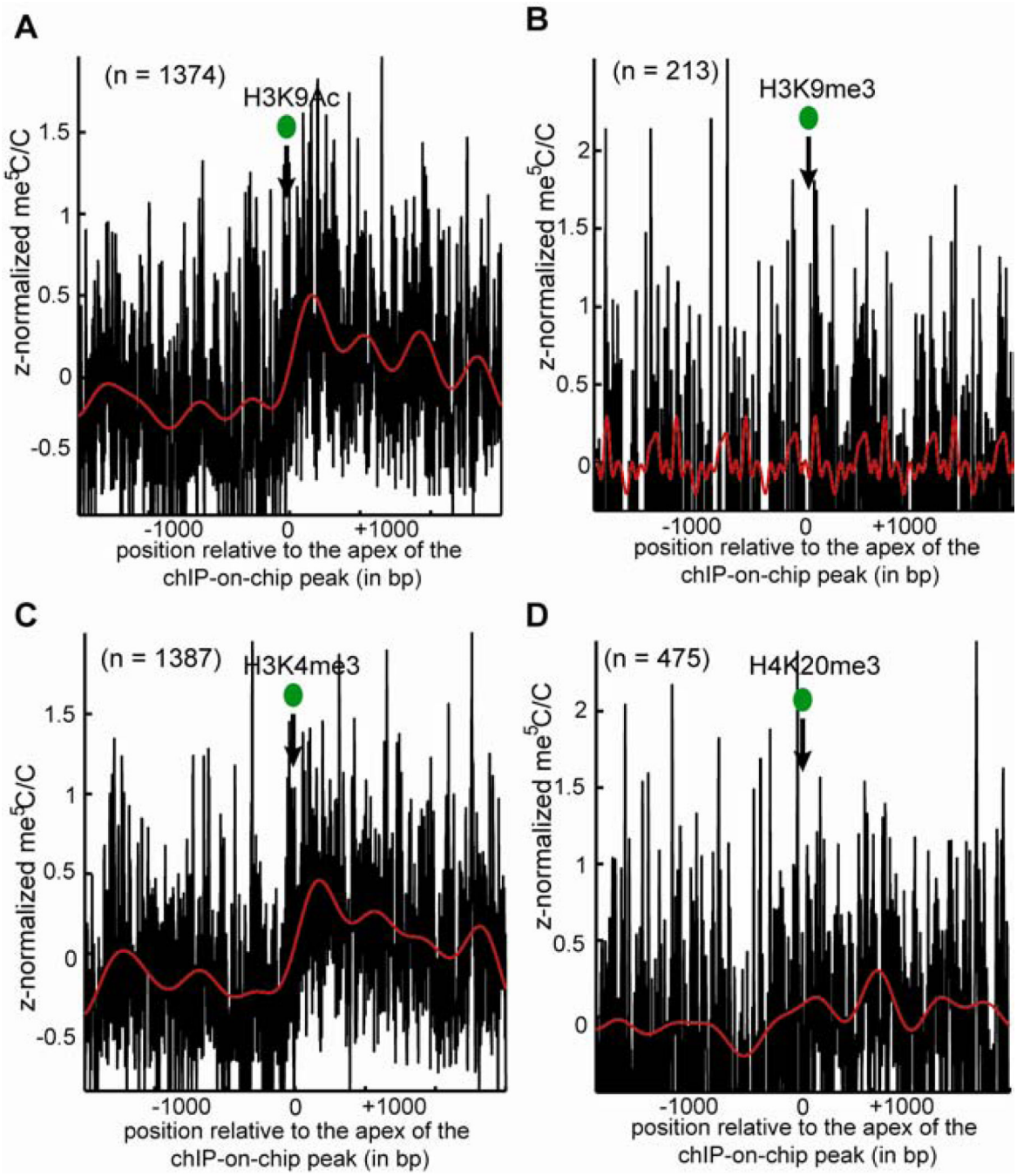

Figure 7. DNA methylation and histone modifications

(A) Methylation status of regions containing H3K9Ac. For each position in the region spanning 1600bp centered on the apex of the chIP-on-chip peak for H3K9Ac (Lopez-Rubio et al., 2009), me^5C/C are averaged and z-normalized (black curve; red curve = Fourier transform). The green dot is a scaled representation of a nucleosome. (B) Methylation status of regions containing H3K9me3. (C) Methylation status of regions containing H3K4me3. (D) Methylation status of regions containing H4K20me3. For each modification, similar profiles were obtained considering three biological replicates independently (data not shown). See also Figure S7.