Putative SET-domain methyltransferases in *Cryptosporidium parvum* and histone methylation during infection

Manasi Sawant¹, Sadia Benamrouz-Vanneste^{1, 2}, Dionigia Meloni¹, Nausicaa Gantois¹, Gaël Even^{3,4}, Karine Guyot¹, Colette Creusy⁵, Erika Duval⁵, René Wintjens⁶, Jonathan Weitzman⁷, Magali Chabe¹, Eric Viscogliosi¹, Gabriela Certad^{1, 8}

¹Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 – UMR 9017 – CIIL – Centre d’Infection et d’Immunité de Lille, F-59000 Lille, France

²Equipe Ecologie et Biodiversité, Unité de Recherche Smart and Sustainable Cities, Faculté de Gestion, Economie et Sciences, Institut Catholique de Lille, France

³Gènes Diffusion, F-59501 Douai, France;

⁴PEGASE-Biosicences

(Plateforme d’Expertises Génomiques Appliquées aux Sciences Expérimentales), Institut Pasteur de Lille, F-59000 Lille, France

⁵Service d'Anatomie et de Cytologie Pathologiques, Groupement des Hôpitaux de l'Institut Catholique de Lille (GHICL), F-59000 Lille, France

⁶Unit of Microbiology, Bioorganic and Macromolecular Chemistry, Department of Research in Drug Development (RD3), Faculté de Pharmacie, Université Libre de Bruxelles, 1050 Brussels, Belgium

⁷Université de Paris, Epigenetics and Cell Fate, CNRS, Paris, France.

⁸Délégation à la Recherche Clinique et à l'Innovation, Groupement des Hôpitaux de l'Institut Catholique de Lille, F-59462 Lomme, France

* Correspondence: gabriela.certad@pasteur-lille.fr

Abstract:

Cryptosporidium parvum is a major cause of an intestinal pathology called cryptosporidiosis which affects humans and other vertebrates. Despite being declared as a public health problem by World Health Organization (WHO) since 2006, pathogenesis caused by this parasite remains poorly understood. More recently, *C. parvum* has been linked with oncogenesis. In particular, the mechanisms involved in the processes of gene expression regulation are completely unexplored in *Cryptosporidium*. In the current study, we took the opportunity to investigate a dynamic epigenetic modification called histone lysine methylation during the life cycle of the parasite. We successfully identified putative SET-domain containing proteins, lysine methyltransferases (KMTs), which catalyze the methylation of different lysine residues. Phylogenetic analysis classified them into distinct subfamilies namely CpSET1, CpSET2, CpSET8, CpKMTox and CpAKMT. Structural analysis further characterized CpSET1, CpSET2 and CpSET8 to be histone lysine methyltransferases (HKMTs). Their functional significance was predicted by using site-specific methyl-lysine antibodies during development of the parasite (CpSET1:H3K4; CpSET2:H3K36; CpSET8:H4K20). In particular, the SET domain of CpSET8 showcased methyltransferase activity confirming the existence of functional HKMTs in *Cryptosporidium*. Moreover, the consequence of *C. parvum* infection on the host lysine methylation events highlights the inherent potential of the parasite to exploit the host epigenetic regulation to its advantage. Thus, this study is the first one to provide insights on epigenetics mechanisms occurring throughout the parasite's life cycle and during the interaction with its host. As *Cryptosporidium* is a protozoan that significantly affects the health of both humans and animals, a better understanding of its developmental processes within the definitive host may highlight novel infection control strategies.

Author Summary

Cryptosporidium species have a very compact genome (~9.2 Mb) unlike its apicomplexan homologs such as *Toxoplasma* (~63 Mb). Moreover, the lack of large families of transcriptional factors requires them to heavily rely on chromatin remodeling components for its gene regulation. Thus, study and identification of novel elements which contribute to chromatin dynamics could assist a better understanding of the biology of this parasite. In the current study we investigated histone lysine methylation, a dynamic epigenetic modification which regulates gene activation as well as repression. More importantly, characterizing the enzymes which bring about this regulation, provides potential new druggable targets to attack the parasite.

1. Introduction

Cryptosporidium belong to the eukaryotic phylum Apicomplexa, distinct relatives of the parasites that cause malaria and toxoplasmosis (1). *Cryptosporidium* infection, also termed as cryptosporidiosis, is responsible for self-limited diarrhea in healthy immunocompetent individuals but is capable of causing life-threatening disease among those who are immunocompromised (2). Most recently, a series of epidemiological studies has reported the diarrheal disease caused by the parasite to be one of the leading causes of early childhood morbidity and mortality in developing countries, together with rotavirus, enterotoxigenic *Escherichia coli* and *Shigella* (3,4). The environmentally resilient nature of the *Cryptosporidium* oocysts allows the parasite to withstand common water treatments such as chlorination (5), and therefore to remain a major cause of waterborne outbreaks in industrialized countries as well (6,7). Even though being identified to have a significant impact on public health, no vaccine or chemoprophylactic drugs to prevent *Cryptosporidium* infection and very few chemotherapeutic options for its treatment are currently available (8).

Majority of human infections by this protozoan are caused by *Cryptosporidium hominis* and *Cryptosporidium parvum*. Thus, new opportunities to uncover targets for drugs are attributed to the availability and *in silico* analysis of genome sequences of *C. hominis* (9) and *C. parvum* (10). These two genomes are small in size (approximately 9 Mb), adenosine- and thymidine-rich (approximately 70%) and differ only at nucleotide level by 3% -5% (9,10). In parallel, *C. parvum* transcriptomes at the oocyst, excysted sporozoites and intracellular stages have been widely investigated to understand *Cryptosporidium* life cycle development (11,12) even though the asynchronous nature of the parasite life cycle has made the transcriptome interpretations challenging. Interestingly, with the availability of the technologies to genetically manipulate the parasite (13), it became possible to analyze transcriptome at specific

intracellular stages *in vitro* as well as *in vivo* (14), leading to the discovery of novel targets for therapeutic interventions.

Strikingly, the compact genome of *Cryptosporidium* compared to other apicomplexan homologs (e.g approximately 63 Mb for *T. gondii*) accompanied by the lack of large families of recognizable transcription factors typically encountered in eukaryotic organisms (15), suggest major differences in the mechanisms of apicomplexan gene regulation. Accordingly, these mechanisms may depend on epigenetic events in order to control gene expression and cellular differentiation (16).

In eukaryotes, epigenetic changes include DNA methylation (17) and histone modifications such as lysine methylation (18) and acetylation (19). Apicomplexan parasites such as *Toxoplasma* and *Cryptosporidium* are known to encode putative DNA methylation enzymes but they lack detectable DNA cytosine methylation events (20). On the other hand, novel drugs targeting the enzyme histone deacetylase (HDAC) which results in hyperacetylation of their genomes causing arrest of parasite differentiation has been reported in *Toxoplasma* (21), *Plasmodium* (22) and *Cryptosporidium* (23). Along with acetylation, histone lysine methylation, a rather more sophisticated and dynamic post-translational modification previously identified to be restricted to metazoans has been extensively studied in *Toxoplasma* and *Plasmodium* (16). Lysine methyltransferases (KMTs) known to methylate Histone 4 lysine 20 (H4K20), were identified to regulate cell cycle progression in these two apicomplexan parasites (24). In *Plasmodium*, lysine methylation markers such as Histone 3 lysine 4 (H3K4) and Histone 3 lysine 9 (H3K9) have been reported to be involved in regulatory mechanisms of switching between variant surface antigens, enabling the parasite to evade the host immune response (25,26). Histone methylation is a reversible event catalyzed by histone demethylases. Jumonji-C-terminal (JmjC) domain-containing putative histone demethylases have been identified in *T. gondii*, *P. falciparum*, *Babesia bovis*, and *Theileria annulata* (27). Databases

have also identified two lysine-specific demethylases (LSD)-like proteins in *T. gondii* genome (28). However, these epigenetic modifications have not been previously investigated in *Cryptosporidium* parasites.

On the other hand, pathogens usually induce alterations of their hosts employing several strategies to target cellular processes during their complex interactions with host cells (29). In this way, they can evade the barriers imposed by checkpoint responses, and can manipulate various pathways activated for cell protection in order to increase their survival and transmission. Then, epigenetic mechanisms could play a fundamental role in the dynamic of host–parasite interactions (29). As an example, lysine methylation is emerging as a versatile and dynamic post-translational modification that contributes critically to cellular differentiation programs being a pivotal dynamic event in host-pathogen interactions (30). Interestingly, *Theileria*, which causes a lymphoproliferative disease in cow manipulates the leukocyte epigenetic mechanisms favoring parasite replication and persistence leading to the host cell transformation (30). Strikingly, epidemiological and experimental studies suggest a potential link between *Cryptosporidium* infection and digestive cancer (31). Of all apicomplexan parasites, *Cryptosporidium* and *Theileria* are able to transform their mammalian hosts. However, little is known about the significance of epigenetic variations in *Cryptosporidium* development and in the parasite interactions with its host.

In the current study, we thus aim to characterize the KMTs of *Cryptosporidium* in order to identify lysine methylation events which might be involved in regulating gene expression during the life cycle of the parasite and to evaluate host epigenetic modulators potentially involved in pathogenicity and parasite-induced transformation. In parallel, this study also highlights the potential of lysine methylation modifications to be considered as targets in the development of therapeutic strategies against *Cryptosporidium* infection.

2. Results

2.1. *In silico* identification of putative KMTs of *C. parvum* and sequence comparison of their SET domains

Queries were run to retrieve KMTs encoding genes from *C. parvum* containing SET domain consensus. As a result, ten putative KMTs with a recognisable SET domain were predicted (Table 1).

Table 1 - List of putative KMTs identified in the *C. parvum* genome.

Gene ID	Name	Chromosomal location	Protein size (amino acids)	Domains in addition to SET domain	Family representative	Predicted Substrate Specificity
cgd8_2730	CpSET1	CM000436: 718307..725208 (+ strand)	2244	PHD, bromo, post-SET	SET1	H3K4Me1/2/3
cgd5_400	CpSET2	CM000433: 103870..106923 (+ strand)	1004	Zinc finger, PHD, AWS, post-SET	SET2	H3K36Me1/2/3
cgd4_370	CpSET8	CM000432: 90500..92170 (+ strand)	556	-	SET8	H4K20Me1/2/3
cgd4_2090	CpAKMT	CM000432: 502573..504324 (- strand)	583	TPR	AKMT	Unknown
cgd5_2340	CpKMTox	CM000433: 503754..505532 (+ strand)	524	-	KMTox	H2A/H4
cgd7_5090	CpSET UNK1	CM000435: 1177390..1179244 (- strand)	574	Zinc finger - MNYD	SMYD3 ?	H3K4 ?
cgd1_2170	CpSET UNK2	CM000429: 494029..495242 (- strand)	366	-	SMYD3 ?	H3K4 ?
cgd6_1470	CpSET UNK3	CM000434: 350153..352792 (- strand)	879	AWS	Ez11p ?	H3K27 ?
cgd8_3840	CpSET UNK4	CM000434: 894260..896419 (+ strand)	719	-	?	?
cgd6_980	CpSET UNK5	CM000434: 248189..249592 (- strand)	467	-	?	?

*PHD – Plant homeodomain

*AWS – Associated with SET

*TPR – Tetrapeptide repeat

*UNK – Unknown

The corresponding genes are distributed on 6 chromosomes and the size of the full-length proteins range from 467 to 2244 amino acids. *In silico* analysis also showed the presence of CpSET proteins orthologues in all *Cryptosporidium* species including *C. muris*, *C. andersoni*, *C. hominis*, *C. meleagridis* and *C. ubiquitum*. Among the known methylated lysines, only H3K79 is known to be methylated by DOT1, a non-SET domain histone methyltransferase in mammals. Unlike SET domain proteins, DOT1 domain containing proteins were observed only in *C. muris* and *C. andersoni*. Analysis of domain organisation of so-called CpKMTs identified additional domains, including the bromodomain, PHD (plant homeodomain) zinc finger domain, associated with SET (AWS) domain, tetratricopeptide repeat (TPR) domain and MYND (Myeloid translocation protein 8, Nervy, and DEAF-1). In contrast, JmJC-domain present in lysine demethylases (KDMs) was not identified in *Cryptosporidium* genomes.

The SET domain proteins are usually classified into seven families including SUV39, SET1, SET2, EZ, RIZ, SMYD and SUV4-20 as well as a few orphan members SET7/9 and SET8 (also called as Pr-SET7). Sequences of CpSET domains were aligned with the representatives of these families. Sequence analysis of SET domain of KMTs revealed the presence of four signature motifs namely motif I (GxG), motif II (YxG), motif III (RFINHxCxPN) and motif IV (ELxFDY). As a result, 8 out of 10 predicted CpSETs exhibited at different levels of similarity the three catalytically essential motifs GxG, RFINHxCxPN and ELxFDY of the SET domain (Figure 1) which was not the case for the two remaining CpSETs (cgd6_3840 and cgd6_980), which were therefore excluded from our analysis. The motif II was found to be very well conserved in cgd8_2730, cgd5_400 and cgd4_370 compared to other CpSETs. The sequence between motif II and motif III was observed to be the most variable region of the SET domain. The canonical preSET domain was not reported preceding any of the CpSETs according to InterPro database. However, cysteine (Cys) rich cluster was observed flanking the N-terminal extremity of cgd5_400 and cgd6_1470 SET domains. This region has been identified as AWS

domain by InterPro database. Regarding the C-terminal flanking region of the CpSET domain, it is composed of the post-SET domain which contains the CXCX₂₋₄C motif, well conserved in cgd8_2730, cgd5_400, cgd1_2170 and cgd6_1470. This region has also been identified as post-domain in InterPro database. All the CpSETs exhibiting the post-SET domain also conserved the Cys residue in the motif III. Indeed, the Cys residues from post-SET domain and motif III together form a channel to accommodate target lysine side chain. However, cgd7_5090 showed sequence variation in the post-SET motif (CXCX₂C) similar to SMYD (SET-and MYND-domain containing) and Suv4-20 KMT families. With a distinct residues conserved in motif III, cgd5_2340 is recognized to be part of separate family of KMT called as KMT_{ox}. Finally, cgd4_2090 was identified to conserve the post-SET cysteine cluster (CXCX₂CX₁₁CX₂C) which has been previously described to be a feature of Apical lysine methyltransferases (AKMT), a cluster of KMTs only including apicomplexan homologues (Figure 1).

2.2. Phylogenetic analysis of putative CpSETs

The 8 *C. parvum* amino acid sequences identified as putative CpSETs were subjected to phylogenetic analysis along with representatives of different substrate specific SET domains (Figure 2). cgd8_2730 clustered with representatives of the SET1 family of HKMTs such as *S. cerevisiae* SET1 and *H. sapiens* SET1 and was thus identified as CpSET1. This distribution was strongly supported by bootstrap resampling in NJ (86%) and ML (97%) methods. The clustering of CpSET1 together with previously identified apicomplexan homologues from *T. gondii*, *P. falciparum* and *T. annulata* was also moderately supported by bootstrap values (51% and 65% of the replicates under NJ and ML, respectively). The substrate specificity of this family is at the 4th lysine residue of the histone 3 protein (H3K4). Thus, we predicted H3K4 to be the substrate of CpSET1 (Figure 2) (Table 1).

Regarding cgd5_400, this sequence fell within the paraphyletic SET2 family of KMTs including *Saccharomyces cerevisiae* and *Homo sapiens* as well as apicomplexan enzymes from *T. gondii*, *T. annulata* and *P. falciparum*. Our phylogenetic tree suggested that cgd5_400 belongs to SET2 family of HKMTs which have been reported to methylate H3K36 and is so-called CpSET2 (Figure 2) (Table 1).

cgd4_370 was observed to be highly homologous to human SET8 and therefore named CpSET8. This homology was strongly supported by bootstrap values of 87% and 94% according to NJ and ML methods, respectively, in our phylogenetic analysis, proposing that CpSET8 methylate H4K20 as described for human SET8. Another putative KMT of *C. parvum* cgd5_2340 grouped together with KMTox from the Apicomplexa *T. gondii* and *Besnoitia besnoiti* with strong bootstrap values of 99% (NJ method) and 98% (ML method) and was thus reported as CpKMTox. KMTox has been defined as a new family of nuclear KMTs specifically found in Apicomplexa and thus including *Cryptosporidium*. Similar to KMTox, AKMTs also form a distinct clade from other known KMTs only found in apicomplexan species. cgd4_2090

or CpAKMT represents *Cryptosporidium* AKMT as it clustered with high bootstrap support (bootstrap values of 99% and 99% according to NJ and ML methods, respectively) with other apicomplexan homologues.

The phylogenetic emergence of the three remaining CpSETs so-called CpSET Unk1 (cgd7_5090), CpSET Unk2 (cgd1_2170) and CpSET Unk3 (cgd6_1470) remained uncertain according to our present study (Table 1). Indeed, CpSET Unk3 branched with unsupported bootstrap support at the base of a large group including KMT of the protozoan ciliate *Tetrahymena* and the KMTox of Apicomplexa. Since the KMT of *Tetrahymena thermophila* is known to methylate H3K27, CpSET Unk3 could be involved in the same methylation event even if no sequence has been yet reported to be associated with H3K27 substrate specificity in Apicomplexa (Figure 2).

The two others CpSET Unk1 and Unk2 are representative of a paraphyletic group that includes in particular the human KMT called SMYD3 known to methylate H3K4. Consequently, the weak grouping of CpSET Unk1 and CpSET Unk2 with SMYD3 speculates a possible methylation of H3K4 by these two KMTs of *Cryptosporidium* (Figure 2).

To complete this overview, unlike other apicomplexan parasites, no *C. parvum* sequence was found to be associated with the SUV39 and EZ families mediating H3K9 and H3K27 methylation, respectively (Figure 2).

0.50

Figure 2. Phylogenetic analysis of SET domain proteins of *C. parvum*. All the putative *Cryptosporidium* sequences are highlighted in red. Numbers near the individual nodes indicate bootstrap values given by NJ (left of the slash) and Maximum Likelihood (right of the slash). Asterisks indicate nodes with bootstrap values below 50%. Branch lengths are proportional to sequence divergence and can be measured relative to the scale bar. The scale bar indicates the branch length corresponding to 0.50 substitutions per site. The predicted substrate specificities of KMTs are also indicated on the right of the

figure. The substrate specificities of characterized KMTs clustering with the Cp SET domain proteins are also indicated on the right of the figure. Except KMTox and AKMT are not “substrate specificities”

2.3 Structural characteristics of CpSETs

Homology modelling of the structures of CpSET1, CpSET2 and CpSET8 using publicly available X-ray crystal structures of homologous enzymes showed that the overall architecture of the SET domains belonging to different subfamilies of KMTs were nearly identical. The MolProbity scores evaluating the quality of the models were good with values of 1.62 (92nd percentile), 1.73 (88th percentile) and 1.57 (93rd percentile) for CpSET1, CpSET2 and CpSET8, respectively (Table 2).

Table 2 - MolProbity statistics for 3D models of CpSETs

3D model	Clashscore ^a		Ramchandran favoured	MolProbity Score ^b	
CpSET1	2.24	99 th percentile	95.04%	1.62	92 nd percentile
CpSET2	2.91	98 th percentile	89.22%	1.73	88 th percentile
CpSET8	4.61	96 th percentile	94.52%	1.57	93 rd percentile

^a percentile score established with N =1784, considering all resolutions

^b Percentile score established with N=27675, considering structures at all resolution ranges

Pairwise structure comparisons with DALI program was used to check whether conserved residues line up between CpSETs and the templates. As described in Table 3, CpSET1 shares highest structural identity with the templates (52%) compared to CpSET2 (43%) and CpSET8 (44%). The visualization of these superimposed structures was performed using ChimeraX software (Figure 3).

Table 3. Pairwise structure comparison between generated models and templates

Structure Parameters	CpSET1 onto 5F6L	CpSET2 onto 6J9J	CpSET8 onto 5TEG
RMSD (Å) ^a	0.3	0.6	0.4
Number of superimposed residues	136	139	144
Total number of residues	143	169	148
Dali z-score ^b	24.4	22.2	25.8
Percentage identity	52 %	43 %	44 %

^a RMSD is the root-mean-square deviation computed over the Cα atoms of superimposed residues.

^b A DALI z-score greater than 3 is considered as reflecting similar structures. Higher z-scores indicate more similar structures.

^c the percentage identity is calculated between structurally superimposed residues.

Briefly, all the structures contained the specific β fold identified only in KMTs, not in any other previously characterized AdoMet-dependent methyltransferases. The fold has several series of curved β strands forming several small sheets that define the core of the SET domain. This β fold is followed by knot-like structure which is also observed in all the 3D models of CpSETs. The knot involves a C-terminal β strand threading through a hoop consisting of two β strands and their connecting region. This represents an archetypal feature of SET domain which

consists of residues from the motif II, motif III and post-SET region that enclose the lysine residue and holds it in the appropriate chemical environment and position for methyl transfer by motif I (Figure 3, boxed in black dotted line). The essential residues have similar arrangement in all the SET domains (Figure 3, enlarged images of active site).

CpSET1 (Tyr 2129 and Tyr 2217) and CpSET2 (Tyr 574 and Tyr 659) conserved the key tyrosine residues (Figure 3A; 3B). These Tyr residue are expected to form an intricate network of hydrogen bonds which would place the methyl group in direct line with the N_ε of the lysine residue of the histone tail. Structural alignment revealed that in CpSET8 one of these tyrosine residues is replaced by phenylalanine residue (Tyr 445 and Phe 537) (Figure 3C). This represents the F/Y switch which determines the KMT can mono-, di- or tri- methylate the histone.

The C-terminal flanking region of different SET domain families is often divergent but CpSET1 and CpSET2 exhibited a classical post-SET domain. The prominent feature of this domain is a zinc-binding cage formed by three Cys residues from the C-terminal region whereas the fourth tetrahedral Cys ligand (CpSET1 Cys 2183 ; CpSET2 Cys 654) is provided by the loop linking motif II and motif III of SET domain. The narrow channel formed as a result of Cys interactions accommodates the target lysine and brings the N_ε in close proximity of the donor at the opposite end of the channel (Figure 3A, B). In CpSET8, the C-flanking domain consists of a helix and the presence of the Trp 552 residue is responsible for interactions with the cofactor (Figure 3C).

315

316

Figure 3 – Structural modelling of SET domain regions of *C. parvum* KMTs. Superimposition of 3D homology models of SET domains of [sandy brown] CpSET1 (A), CpSET2 (B) and CpSET8 (C), coloured in sandy brown, with, in cyan, those of resolved crystal structures of SET domains of *Homo sapiens*, HsSET1, HsSET2 and HsSET8, respectively. The superimposition of the catalytic site has been highlighted in the black dotted box and depicted in an enlarged view with side chains showed in stick representation (nitrogen and oxygen atoms in blue and red, respectively) and labelled in orange and blue for *C. parvum* and *H. sapiens* structures, respectively.

2.4 Enzymatic activity of CpSET8

To check whether CpSETs represent active KMTs, we expressed and produced SET domain regions of CpSET1 and CpSET8 in a bacterial system. Western blots analysis showed detectable amounts of CpSET8 in the soluble fraction after induction and lysis of bacteria compared to CpSET1 (Figure 4A). Therefore, KMT activity assay was performed on the induced soluble protein fraction of bacterial lysate containing SET8. As a result, CpSET8 showed HKMT activity against H4 which evidenced by a relatively strong band seen below 15kDa (Figure 4B, Lane 2) compared to the controls which did not contain either SAM (Figure 4B, Lane 1) or H4 (Figure 4B, Lane 3). The specificity of H4K20 methylation band was also confirmed by the detection of H4 at the same molecular weight. Moreover, since anti-his tag antibody detects the 6X histidine tag present in the recombinant CpSET8 and, equal intensity of anti-his tag bands corresponds to equal amount of protein loaded in each well, this indicated that the high intensity band detected (red arrow) was the result of enzymatic activity of CpSET8.

Figure 4. Expression of putative CpSETs and enzymatic activity of CpSET8. A. Western blot analysis of 6x-histidine tagged CpSET domains using bacterial expression system. Anti-6x-His tag antibody was used to detect the CpSETs. M – Molecular marker, Un – Uninduced soluble protein fraction, In – Induced soluble fraction. B. Ponceau S staining of the western blot subjected to chemiluminescent detection of H4 methylation by recombinant 6x-histidine tagged CpSET8. C. Chemiluminescent signals showing CpSET8 to tri-methylate H4 using anti-H4K20Me3 antibody (red arrow). Recombinant human H4 was detected using anti-H4 antibody. CpSET8 was detected using anti-6x-His tag antibody. H4 – Histone 4. SAM – S-adenosyl methionine.

2.5 Functional characterization of identified KMTs during *C. parvum* life cycle

To determine the expression pattern of KMTs during the *C. parvum* biological cycle *in vitro*, RT-qPCR analysis of the eight *CpSET* genes was performed. The expression of all *CpSET* genes was determined at 3 time points i.e. 6 h, 24 h and 55 h PI, relative to their expression at 2 h PI. *CpSET1* showed 14-fold increase at 6 h PI, corresponding to the time at which trophozoite development is predominant. Thus, *CpSET1* is the highest expressed KMT gene, particularly during initial stages of the infection. *CpAKMT* expression was 10-fold increased during meront development at 24 h PI. *CpSET2*, *CpSET8*, and *CpKMTox* genes also showed high expression during trophozoite stage (6 h PI) followed by asexual (24 h PI) and sexual stages (55 h PI). At 55 h PI, wherein the sexual stages were predominant, all the identified putative *CpKMTs* were constitutively expressed. The gene expression of putative unknown *CpKMTs* such as *CpSET Unk1*, *CpSET Unk2* and *CpSET Unk3* during intracellular development was observed to be lower compared to that observed during the extracellular sporozoite stage at 2 h PI (Figure 5A). Out of the five reported putative *CpKMTs*, *CpSET1*, *CpSET2* and *CpSET8* have been identified as histone lysine methyltransferases (HKMTs). In order to functionally characterize these HKMTs, the histone lysine methylation events associated with their activity during the life cycle of *C. parvum* were investigated *in vitro* in relation with the different parasite developmental stages. As previously described, *CpSET1*, *CpSET2* and *CpSET8* are associated with H3K4, H3K36 and H4K20 methylation. Immunofluorescence analysis revealed that H3K4, H3K36 and H4K20 methylations occur during all the intracellular stages of the parasite. It can be observed that anti-tri-methyl H3K4 (H3K4Me3) antibody recognized the *C. parvum* chromatin with a broader nuclear distribution through all the developmental stages. Regarding anti-tri-methyl H3K36 (H3K36Me3) and anti-tri-methyl H4K20 (H4K20Me3) antibodies staining, it showed a punctate pattern spread throughout the nucleus, most likely the pericentric heterochromatin during the intracellular stages compared to extracellular sporozoite stage

(Figure 5B). By quantifying the labelling of these histone lysine methylations by western blotting, we were able to confirm that H3K4 methylation levels remain consistent during meronts development at 24 h PI and microgamonts and macrogamonts development at 55 h PI. On the other hand, H3K36 and H4K20 methylations levels have been shown to fluctuate from asexual (24 h PI) to sexual stage 55 h PI (Supplementary Figure 2A). H3K36 and H4K20 methylations increase by 2 and 3 fold respectively, when there is predominance of sexual stages of the parasite in the culture (Supplementary Figure 2B).

Figure 5 – Characterization of KMTs during *C. parvum* infection in *in vitro* culture. A. RT qPCR analysis illustrating expression of *CpSET* genes during *C. parvum* development *in vitro*. The expression levels were analysed in triplicates and normalized with 18S rRNA gene as internal control. The ΔC_t values at the sporozoite stage were used as the calibrator B. Immunofluorescence analysis of histone lysine modifications in different stages of *C. parvum*. Co-staining with anti-histone antibodies such as

H3 and H4 and anti-histone methylation antibodies such as H3K4Me3, H3K36Me3 and H4K20Me3 (green), anti-*Cryptosporidium* antibody (red) and DAPI (blue).

2.6 Effect of *C. parvum* infection on methylation of the host histone lysine proteins

The well-documented model of *C. parvum*-induced colon cancer in SCID mice treated with dexamethasone was used. The detection of *C. parvum* infection in these animals was confirmed by quantification of the oocyst shedding for the entire duration of the experiment. Upon histological examination of the ileo-caecal region of infected animals, the presence of well-differentiated adenocarcinomas invading the submucosae through the *muscularis mucosae* was confirmed after 60 days PI. The availability of antibodies to detect histone lysine methylations in the parasite allowed us to decipher the effect of *C. parvum* infection on the host epigenome. Immunofluorescence analysis of H3K9 revealed a strong immunofluorescence signal in the epithelium of the ileo-caecal region of *C. parvum* infected animals at 60 days PI, while a significant decrease in the H3K4, H3K36, H3K27 and H4K20 methylations was observed. H3K18 methylation levels were shown to be unaffected (Figure 6A). The endogenous levels of H3 protein were used as control. Immunofluorescence staining revealed strong signals in parasite nuclei with all antibodies except H3K18. Quantification of fluorescence intensity signals is available in supplementary figure 3. Further, western blotting analysis was performed to determine which stages of the parasite could affect the host methylation events during *C. parvum* infection *in vitro*. As a result, we observed downregulation of all the methylation marks except for H3K18 to be downregulated by 0.5 fold in infected HCT-8 cells relative to non-infected HCT-8 cells at 55h PI (Figure 6B; supplementary Figure 4C). H3K36 and H4K20 methylations were also relatively downregulated by 0.5 fold in infected host cells during the peak of meront stage development of the parasite at 24h PI (Figure 6B, supplementary figure 4C). H3K4, H3K36, H3K27 and H3K9 methylations were relatively upregulated in the host

414 during the trophozoite stage development of the parasite at 6h PI (Figure 6B, supplementary
415 figure 4C).

416

A

B

Figure 6 – Effect of *C. parvum* infection on the host methylation epigenome. A) Immunofluorescence analysis of histone methylation events during *C. parvum* infection *in vivo* at day 60 PI. Co-staining with anti- histone methylation antibodies such as H3K4Me3, H3K36Me3, H3K27Me3, H3K18Me1 and H4K20Me3 (green), anti-*Cryptosporidium* antibody (red) and DAPI (blue) of the ileo-caecal region of *C. parvum* infected SCID mice. The difference in the fluorescence intensity was observed to be significant ($p < 0.05$). B) Western blotting analysis of histone methylation events during *C. parvum* infection *in vitro* at 6 h PI (trophozoite stage), 24 h PI (asexual stage), 55 h PI (sexual stages) after purification of histones from host cell. NI – Non-infected HCT8 cells. I – Infected HCT8 cells. Scale bar – 20 μ m. Results are representative of three independent experiments.

4. Discussion

In the current study we investigated the potential significance of an epigenetic modification known as histone lysine methylation during *C. parvum* infection. Identification of structurally active KMTs followed by their functional characterization in the form of lysine methylation events makes this study the first one to recognize the importance of epigenetics in the development and potential pathogenicity of *Cryptosporidium*.

Several proteins including KMTs which are responsible for the methylation of specific residues have been characterized in higher eukaryotes, and all but one of these enzymes possess a SET domain. A first screening using InterPro database identified 10 SET-domain containing proteins in *C. parvum* so-called CpSET proteins. A further analysis emphasized their conservation in all other sequenced *Cryptosporidium* genomes suggesting the parasite ancestor had acquired these genes before speciation and divergence within this genus. However, KMTs belonging to the DOT1 family were detected in *C. muris* and *C. andersoni* but not in *C. parvum*. The absence of DOT1 domain containing lysine methyltransferases in some *Cryptosporidium* species as in related genera including *Toxoplasma* (32) and *Plasmodium* (27) suggests a secondary loss of the corresponding genes during the evolution of the phylum Apicomplexa. Along with the SET

domain, the presence of PHD zinc finger domain (33) and bromodomain (34) regards the CpSETs proteins as capable of forming protein complexes and interact with the chromatin. Therefore, it can be suspected that PHD zinc finger domain and bromodomain facilitates binding of the CpSET proteins chromatin and assist in lysine methylation of histones which is known to be catalyzed by SET domain.

The histone modifications catalyzed by KMTs possess narrow substrate specificities, often targeting a single lysine within the respective substrates. These KMTs also differ in their preference for different methylation states (mono-, di-, or tri- methylation) of lysine residues. In spite of the conserved overall structural plasticity, the variations at the active sites have been identified to contribute to their varying substrate specificities (35). Hence, we decided to align the primary sequence of the SET domain of identified CpSET proteins with different families of known KMTs. The SET domain consists of four signature motifs; motif I to IV belong to the pre-SET region of the SET domain, out of which the motifs III and IV are highly conserved and form a pseudoknot structure. The structural significance of this knot is to bring two conserved motifs together in order to form an active site immediately close to the motif I which is a binding site for the methyl donor S-adenosylmethionine (AdoMet). Thus, identification of these catalytically essential motifs involved in transfer of methyl group in 8 out of 10 putative CpSETs allows speculating that *C. parvum* KMTs are functionally active.

Out of all the putative CpKMTs, CpSET1, CpSET2 and CpSET8 were identified as potential HKMTs. Especially, all the residues of the signature motifs of SET and post-SET domains were conserved in the CpSET1 primary amino acid sequence. Moreover, CpSET1 clustered together with other SET1 family homologues with high bootstrap support in the present phylogenetic analysis. ScSET1, one of the homologues grouping together with apicomplexan enzymes has been associated with tri-methylation of H3K4 in the regions of genes which are transcribed early, thus, making them a mark of transcription activation (36). Moreover, domain

organization of CpSET1 consists of bromodomain and PHD domain which have been reported to interact with trimethylated H3K4 (37), suggesting that CpSET1 employs these interacting domains to target euchromatin and methylate H3K4. In addition, superimposition of CpSET1 3-D model with the available crystal structure of human MLL1, one of the members of the SET1 family, strongly suggest that CpSET1 can be considered as a structurally active HKMT. MLL1 has been shown to catalyze multiple methylations. The superimposition study has shown that CpSET1 conserves all the active site residues (Phe 2159, Tyr 2217, Tyr 2219, Phe 2221, Cys 2156 to Phe 2158) found in SET domain of MLL1. Moreover, MLL1 SET domain has similar arrangement to other SET domains (38). CpSET1 also conserved key tyrosine residues (Tyr 2129 Tyr 2217) required for the transfer of methyl group. Even though, sequence analysis indicates that these tyrosine residues constrain HKMTs such as SET8 to be mono-methylases. It can be speculated that certain modulations in the configuration of the active site of CpSET1 could allow multiple methylations which is the case for MLL1. Superposed structure of MLL1 SET domain with other HKMTs (SET7/9, SET8 and Dim5) revealed that MLL1 has a more spacious active site (38). This spacious active site is attributed to shift in the orientation of SET-I region and C-terminal flanking region in MLL1. This feature is also evident in CpSET1 after superimposition with MLL1. The CpSET1 residues (Cys 2156 to Phe 2158) are conserved which are responsible for this change in orientation and allowing free movement of the lysine side chain. Thus, CpSET1 can be predicted to mono, di or tri-methylate H3K4. Interestingly, MLL family members (MLL1-4, SET1A and SET1B) are known to methylate H3K4 and have pivotal roles in the regulation of the transcription of genes involved in development and hematopoiesis (38). In particular, cell cycle progression (39). In this context, the immunofluorescence labeling observed during the development of the parasite using anti-H3K4Me3 antibody could be attributed to the existence of a functional CpSET1 representing a MLL1 member of the HKMT family *Cryptosporidium*. Moreover, the consistent high

expression of *CpSET1* gene during the parasite development suggests that H3K4 methylation is not a dynamic but a rather stable post-translational modification necessary during all the stages of the parasite life cycle. In addition, H3K4Me3 has been shown to mark the promoter of actively transcribed genes in different apicomplexan parasites such as *T. gondii* (40), *T. annulata* (41) and *P. falciparum* (25). Hence, the stable nature of this mark can be attributed to the fact that this modification is involved in expression of actively transcribing genes during each stage of the parasite. However, the open conformation of the SET domain of MLL1 has been predicted to be not efficient to transfer the methyl group from the cofactor S-adenosyl-L-methionine (AdoMet) to the target lysine (38). Recently, other components such as RBBP5-ASH2L have been identified to bind and activate MLL family methyltransferases through a conserved mechanism (42). As the SET domain of CpSET1 is structurally identical to MLL1 SET domain, the involvement of other components needed by CpSET1 to maintain its HKMT activity has to be investigated in further studies. Even though, we made preliminary attempts to clone and purify the CpSET1 domain in a bacterial system, western blots analysis did not show detectable amounts of CpSET1 in the soluble fraction after induction and lysis of bacteria. Further studies should work on standardizing the solubility protocols for the concerned domain.

CpSET2 and CpSET8 were the other two HKMTs of *Cryptosporidium* identified to be structurally active. Even though, the phylogenetic tree reconstructed in the present study do not strongly support the grouping of CpSET2 with homologous enzymes from Apicomplexa and others eukaryotes, analysis of its primary sequence revealed the presence of a typical post-SET domain and a cysteine rich N-terminal region preceding the SET domain which both represent peculiarities of proteins belonging to the SET2 family of HKMTs (43). Moreover, the superimposition of 3-D homology model of SET domain of CpSET2 with a template SET domain of SETD2 (protein data bank: 6J9J.A) revealed 43 % identity between the two

structures. Interestingly, SETD2, is the only HKMT known to mediate tri-methylation of H3K36, whereas other H3K36 methyltransferases can only mono or di-methylate H3.

Strikingly, H3K36me3 is known to be mediated by a single HKMT such as SETD2, whereas other H3K36 methyltransferases can only mono- and di-methylate H3K36 (44). Intriguingly, mutating arginine residue (Arg1625Cys) present in the SET domain results in an enzymatically inactive SETD2 failing to tri-methylate H3K36 (45). This arginine residue (Arg 648) is conserved in CpSET2. Thus, based on the structurally conserved residue it can be predicted to tri-methylate H3K36 in *Cryptosporidium*. In addition, the use of anti-H3K36Me3 antibody, detected this methylation mark to be distributed throughout the life cycle stages of the parasite. Thus, functionally validating the speculation of active CpSET2 to be present in *Cryptosporidium*. However, H3K36Me3 has been reported to be a mark of gene activation in higher eukaryotes (46). But methylation of H3K36 has been linked to repression of *var* genes in *Plasmodium* parasites (Jiang L et al. 2013). In contrast, *Theileria* genome did not show any regions of strikingly high H3K36Me3 enrichment (41). Thus, the significance of this methylation mark in *Cryptosporidium* remains to be explored.

SET8 family of HKMTs, have been characterized to mono-methylate H4 in humans (47). The identification and characterization of SET8 related homologs in Apicomplexa such as *Plasmodium* (27) and *Toxoplasma* (48), overruled the thought that this enzyme is restricted to metazoans (35). CpSET8 identified in our study was also predicted to methylate H4K20 according to our phylogenetic analysis. Thus, reinforcing the importance of histone methylations in chromatin structure and function in Apicomplexa. The superimposition of 3-D homology model of SET domain of CpSET8 with a template SET domain of SET8 (protein data bank: 5teg.A) revealed 44 % identity between the two structures. The conserved tyrosine residues (Tyr 245 and Tyr 334) within the active site of human SET8 are responsible for maintaining an intricate network of hydrogen bonds to position the side chain of only mono-

543 methylated lysine residue (47). On the contrary, Dim-5, a HKMT which can tri-methylate its
544 target lysine can accommodate mono-, di- and tri-methylated lysine in its active site . This
545 characteristic of Dim-5 has been attributed to the replacement of one of the tyrosine residues to
546 phenylalanine residue (Tyr 178 and Phe 281) (49). Interestingly, structural alignment revealed
547 that in CpSET8 as well one of these tyrosine residues is replaced by phenylalanine residue (Tyr
548 445 and Phe 537). Thus, suggesting that CpSET8 is capable of adding multiple methyl group
549 to its target lysine.

550 Thus, based on the analysis of structural elements, it can be hypothesized that CpSET8 is
551 capable to methylate H4K20Me1, H4K20Me2 and H4K20Me3 like SET8 of *T. gondi* (48). This
552 structural analysis was confirmed herein by showcasing HKMT activity of CpSET8 domain for
553 H4K20Me3. Along with the suspected dynamic nature of CpSET8, the methylation events
554 predicted for this HKMT have also been observed to be regulated during parasite
555 differentiation. This phenomenon can be considered as a characteristic of apicomplexan
556 parasites as H4K20Me3 mark in *T. gondii* has been reported to be cell cycle dependent (48).
557 Even then, the possibility of CpSET8 methylating other targets cannot be excluded. However,
558 unlike other HKMTs such as SMYD3 which are known to have multiple targets (50,51). SET8
559 family of HKMTs have till date only been shown to have a single target that is H4K20 (43).
560 Hence we strongly propose that CpSET8 is capable of methylating H4K20. Indeed, considering
561 these as preliminary results, further studies need to be carried out by mutating the predicted
562 catalytic residues of the whole protein to see if the protein folding has any effect on the function
563 of the SET domain

564 CpAKMT was another identified KMT which differed in the C terminal region wherein it
565 retains two extra cysteines in addition to the post-SET domain. Phylogenetic analysis revealed
566 that CpAKMT and its homologues from other Apicomplexa clustered together and represented
567 a sister-group of HsSMYD3, as described in previous evolutionary studies including

Plasmodium (27) and *Toxoplasma* KMTs (52). The lack of MYND zinc finger domain, a fundamental feature of SMYDs allows to group the AKMT homologues as a distinct family of KMTs. Structural comparison between the AKMT of *T. gondii* and SMYD proteins revealed that SMYD1-3 do not seem to possess the necessary features that are compatible for dimer formation whereas AKMT are dimeric enzymes (52). In *T. gondii*, a functional AKMT has been reported to be localized at the apical complex and associated with parasite motility and egress (53). Moreover, lack of a MYND domain in AKMT might correlate with its specific function outside the nucleus (54). Interestingly, immunofluorescence observations detected the labeling of histone lysine methylations at the apical region of *C. parvum* sporozoites and merozoites. In addition, the *CpAKMT* gene had a relatively high expression during the merozoite development and egress in *C. parvum*. In *P. falciparum*, histones released from the parasite exert a disruptive effect on the endothelial barrier function through a charge-related mechanism in order to induce pro-inflammatory responses (55). Along with the extra-nuclear localization of histones, histone modification such as H3K9Me1 in *P. falciparum* has been recognized as a new function linked to the parasitophorous vacuole and the interaction of the parasite with the host (56). Therefore, we hypothesize that *CpAKMT* is a functional KMT localized at the apical region capable to methylate extra-nuclear histones of *C. parvum*. However, possibility of *CpAKMT* to methylate non-histone proteins at the apical region to assist in motility cannot be overlooked. Similar to SMYD3 which has been recently characterized as a non-histone methyltransferase considering its cytoplasmic localization (57). Further studies have to be carried out to address this functional aspect.

KMTox was another identified KMT in *C. parvum*. First identified in *T. gondii*, the presence of High Mobility Group (HMG) domain in KMTox allows it to recognize the bent DNA with the SET domain involved in the methylation of histones H4 and H2A *in vitro* (58). Even if the HMG domain was not identified in *CpKMTox*, the phylogenetic analysis based on SET domain

sequences clustered together all apicomplexan KMTox with a high bootstrap value suggesting that CpKMTox could be predicted as a novel histone H4- and H2A-specific methyltransferase. Even though it was previously reported that TgKMTox formed a distinct clade with no obvious homologues in the Apicomplexa lineage (58), our phylogenetic analysis identified other apicomplexan parasites retaining KMTox such as *C. parvum* and *B. besnoiti* that could represent a new clade of KMTs only found in this group of protozoa.

The nine cysteines residues of the pre-SET domain usually found in the SUV39 family of KMTs (59) were not identified in CpSETs. In parallel, none of the CpSETs clustered in our phylogenetic tree with the known homologues belonging to this family as previously reported (24). Thus, it can be speculated that either H3K9 methylation is not a crucial post-translational modification required for the survival of *C. parvum* or that a yet unknown CpSET may carry out this function. The post-SET region of KMTs also consists of conserved cysteine residues (CXCX₄C). These three cysteines coordinate a zinc ion tetrahedrally together with cysteine of motif III of SET domain to form a narrow channel to accommodate the target lysine side chain. Consequently, the post-SET region is extremely crucial for the activity of a functional KMT. We observed that in CpSET1, CpSET2, CpSET Unk2 and CpSET Unk3 the cysteines of the post-SET region along with cysteine of motif III were conserved. However, CpSET Unk2 and CpSET Unk3 lacked the presence of the signature of motif I and could not be grouped together with known families of KMTs. On the other hand, CpSET Unk1 presented a variant of post-SET domain which is observed in the SMYD or SUV4-20 families of KMTs. The presence of a MYND zinc finger motif in CpSET Unk1 and its weak association with *H. sapiens* SMYD3 in our phylogenetic tree suggests that this CpSET could belong to the SMYD subfamily, which is composed of H3K4-specific methylase (50). However, due to their high variability in terms of primary sequence, the functions associated with these CpSETs cannot be affirmatively predicted and thus these enzymes were considered as unknowns. Unlike the other characterized

KMTs based on phylogenetic and structural analysis which show dynamic expression during the intracellular stages of the parasite, *CpSET Unk1*, *CpSET Unk 2* and *CpSET Unk 3* expression levels were relatively unaffected. Thus, these putative KMTs are predicted to not play a role in parasite development.

To complete this overview, KDMs containing JmjC-domain were not identified in *Cryptosporidium*, highlighting that mechanism for histone demethylation is not present in this parasite as reported in previous studies (27,60).

Strikingly, new insights into the effect of *C. parvum* infection on the host histone lysine methylation events were provided in the present study. Indeed, the modulation of different lysine methylation marks in *C. parvum* infected mice was described with the upregulation of H3K9me3 and the downregulation of H3K27Me3, H4K20Me3, H3K4Me3 and H3K36Me3. Till date, one single study identified *C. parvum* infection to induce transcriptional suppression of host genes by upregulation of H3K9Me3 methylation (61). In addition, H3K9Me3 and H3K4Me3 are considered to be opposing and mutually exclusive chromatin modifications (62). Accordingly, it can be proposed that during *C. parvum* infection, H3K9Me3 and H3K4Me3 are functionally connected. Interestingly, the neoplasia developed as a result of chronic infection has been speculated to induce massive loss in the methylation levels. The PI3K/AKT signaling pathway known to be activated during *Cryptosporidium* infection (63) can downregulate H3K27 and H3K4 methylations and consequently activate Epithelial Mesenchymal Transition (EMT) in gastric cancers (64) explaining in part the transformation process that takes place in *Cryptosporidium* infected epithelial cells. This result is consistent with a recent microarray study reporting that EMT takes place within the tumor microenvironment induced by *C. parvum* infection in a rodent model (65). Moreover, *in vitro* analysis, revealed that both these methylations decreased over time predominantly during the development of sexual stages. Down regulation of H3K4Me3 mark has been reported on the promoter of NF- κ B-related pro-

inflammatory genes in *Leishmania* infected macrophages as a survival strategy of the parasite (66). Unlike *C. parvum*, *T. gondii* infection induces upregulation of H3K27Me3 mark. This upregulated level of methylation represses pro-inflammatory cytokines genes to allow the persistence of the infection (67). Thus, regulation of histone lysine methylations levels could be one of the immune escape strategies of *C. parvum*. Particularly, during the development of the sexual stage, the parasite can be expected to target the host's innate immune response genes. In fact, through recent transcriptomics analysis, persistent *C. parvum* infection has been reported to evade host innate immune response to develop tumor microenvironment (65). Further studies have to be conducted to better elucidate this aspect

In conclusion, lysine methylases were successfully characterized in *C. parvum*. Since these enzymes are involved in the development of the parasite life cycle, they may be considered as potential targets to curb the infection opening new avenues for anti-parasite drug discovery. On the other hand, our work contributes to the growing evidence suggesting that protozoan parasites are able to manipulate host cells via epigenetic modifications of the host genome altering transcription and signaling pathways. This unexplored territory of epigenetic modulations in *C. parvum* infection would require further investigation for instance by ChIP-sequencing to identify all regulated genes.

Materials and Methods

In silico analysis

The protein sequences of the SET domains of several representative KMTs including *Saccharomyces cerevisiae* SET1 (GenBank Accession number EDN62358) and SET2 (NP012367), *Homo sapiens* EZH2 (NP004447), SUV39H1 (BAD96791), SET8 (NP065115), SMYD3 (NP001161212), *Toxoplasma gondii* KMTox (XP002371399) and AKMT (XP002370918) were retrieved from databases and used as queries to search *C. parvum* homologs by performing BLASTp analysis on the database CryptoDB (<http://Cryptodb.org>). In parallel,

JumonjiC (JmJC)-domain was also used for the search of lysine demethylases (KDMs) in the *C. parvum* genome. The presence of the conserved SET domain within the putative KMTs of *C. parvum* was confirmed by analyzing the identified sequences using the InterPro program (<http://www.ebi.ac.uk>) which integrates the signatures provided from 13 different databases (CATH, CDD, HAMAP, MobiDB Lite, PANTHER, Pfam, PIRSF, PRINTS, PROSITE, SFLD, SMART, SUPERFAMILY and TIGRFAMs). Furthermore, multiple sequence alignment was performed to compare the putative SET domain sequences of *C. parvum* with those of a panel of 31 representative KMTs using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) software under manual supervision. Simultaneously, a phylogenetic analysis was performed from the same set of SET domain sequences. All positions containing gaps and regions of ambiguous alignment were removed, yielding 116 sites for phylogenetic inference. Full-length alignment and boundaries can be available upon request to the corresponding author. Briefly, phylogenetic trees were constructed using the Neighbour-joining (NJ) and Maximum Likelihood (ML) methods implemented in Mega X (68) using the Jones-Taylor-Thornton (JTT) substitution model. The relative stability of topological elements was assessed using 1000 bootstrap replicates for both NJ and ML.

Homology modelling

Three-dimensional (3D) models of SET domains identified in the *C. parvum* genome (CpSETs) were built with the automated comparative modeling program Swiss Model Interactive Workspace (<https://swissmodel.expasy.org/interactive>) using as homologous protein templates, highly resolved X-ray crystal structures of human SET1 (MLL1) (Protein Data Bank (PDB) code: 5F6L; X-ray resolution 1.90 Å, SET2 (SETD2) (6J9J; 1.78 Å), and SET8 (SETD8) (5TEG; 1.30 Å). For each CpSET model developed, the quality of the structure was evaluated by the MolProbity web server (69). The MolProbity score is a combination of the clash score,

rotamer and geometric parameters, and the Ramachandran evaluations into a single score (70). Lower MolProbity scores are better, meaning good quality structures. The MolProbity server reports also a percentile relative to the score distribution for crystal structures near the resolution of the submitted structure. In case of a modelled structure, the distribution is established covering all resolutions (range of 0Å-99Å). Distance matrix alignment (Dali) server (71) was used to perform pairwise structural alignment between the template and the newly generated CpSET models. The secondary structures were assigned using DSSP algorithm and the ChimeraX software was used to visualize the superimposition of templates and CpSET models (72).

***Cryptosporidium* oocysts**

Oocysts of *C. parvum* strain Iowa (purchased from Waterborne™, New Orleans, LA, USA) were stored in phosphate-buffered saline (PBS) with penicillin, streptomycin, gentamycin, amphotericin B and 0.001% Tween 20 at 4°C until use. Absence of bacteria and fungi was assured by testing the oocyst suspensions on both Plate Count Agar and Sabouraud plates at 37°C for 1 week. Oocysts viability was determined as previously described (73).

***In vitro* culture**

Human ileocecal adenocarcinoma cells (HCT-8; ATCC CCL-244) were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 2 mM L-glutamine, 10% Fetal Bovine Serum (FBS) and antibiotics (100 U/ml penicillin, 100 µg/ml streptomycin). Cells were cultured at 37°C in humidified incubator supplemented with 5% CO₂. Oocysts excystation was triggered as described previously (74). After infecting the cells with the parasite, the culture was maintained in RPMI 1640 medium supplemented with 2 mM L-glutamine, 15 mM HEPES buffer, 23 mM sodium bicarbonate, 5 mM glucose, 0.5 µM folic acid, 7 µM 4-aminobenzoic acid, 0.1 µM calcium pantothenate, 50 nM ascorbic acid, 1% (vol/vol) heat-inactivated fetal

calf serum, 210 μ M gentamycin, 170 μ M streptomycin and penicillin (105 U/mL). For negative controls that received no parasites, only maintenance medium was applied onto the monolayers.

Animal experiment

A total of 10 seven-week-old CB17-SCID mice were obtained from a colony bred at the Pasteur Institute of Lille (France). Mice were administered with 4 mg/L of dexamethasone (Merck, Lyon, France) through drinking water. Infective doses of *C. parvum* (10^5 oocysts / mouse) were prepared as described previously (75) and inoculated by oral-gastric gavage. In order to quantify parasite shedding, mice faeces were collected and treated as described previously (65). At 60 days post-infection (PI) or when clinical signs of imminent death appeared, mice were euthanized by carbon dioxide inhalation. Experiments were conducted in the animal facility at the Institute Pasteur of Lille (research accreditation number, D 59 350 009). Animal protocols were approved by the French regional ethical committee with the number APAFIS#9621.

Histopathology

Ileo-caecal regions were removed from each mouse, fixed in 4 % neutral formalin and embedded in paraffin. Sections of 4 μ m thick were stained by hematoxylin-eosin-saffron (Leica Autostainer-XL, Rueil-Malmaison, France). Histological sections were analysed using a Leica DMRB microscope equipped with a Leica digital camera connected to an Imaging Research MCID analysis system (MCID software, Cambridge, United Kingdom). Neoplastic lesions at different sites were scored as previously described (75)

Immunofluorescence assay

Sporozoites were fixed in 4% paraformaldehyde (PAF) for 10 min. After a wash with 1X PBS, the sporozoites were incubated in permeabilization solution (0.2 % Triton X-100 in 1X PBS) for 5 min then treated 10 min with blocking solution (0.3 M glycine, 1 % BSA, 0.1 % Tween

20 in 1X PBS). Finally, the sporozoites were incubated in primary antibody solutions for respective histone lysine methylations (Supplementary Table 1) overnight at 4 °C in a humidified chamber. The primary antibody solution was washed away with 1X PBS and the sporozoites were incubated with secondary antibody solution (Supplementary Table 1) for 1 h at room temperature. Following another wash with PBS, the sporozoites were incubated with the antibody anti-*Cryptosporidium* (Sporoglo, Waterborne™, New Orleans, LA, USA) for 45 min. After a final incubation with DAPI (1 µg/ml) for 15 min, the slides were mounted using Mowiol mounting medium (Mowiol ® 4-88, Sigma, USA). For the *in vitro* staining, HCT-8 cells grown on coverslips in 24 well-plates were infected with 30,000 excysted oocysts per well and fixed at different time points PI: 6 h and 24 h to detect asexual stages and 55 h to detect sexual stages. The staining procedure was similar to that described above for sporozoites. For the *in vivo* staining, ileo-cecal sections of 5 µm thickness were obtained from formalin- fixed and paraffin-embedded specimens and placed on glass slides. The progressive rehydration was followed by an antigen retrieval step using citrate buffer pH 6.5 in a microwave oven for 15 min. After 1 h incubation in blocking buffer (2.5% BSA in 0.1 % Tween-20 1X PBS), the primary antibodies, diluted in blocking buffer, were applied for 1 h at 37°C. After three washes of 5 min with 1X PBS supplemented with 0.1 % Tween-20, the slides were incubated in the secondary antibodies for 1 h at 37°C. After a final wash, the slides were counterstained with DAPI and mounted with Mowiol mounting medium. The images were acquired using Zeiss LSM880 confocal microscope and analyzed using the ZEN lite Digital Imaging software.

RNA extraction, cDNA synthesis and Real-Time quantitative PCR (RT-qPCR)

Total RNA was extracted from infected and non-infected HCT8 cells using NucleoSpin RNA Kit (Macherey-Nagel, Germany). An on-column DNase digestion with a RNase-free DNase was included in the process described by the fabricant to remove any genomic DNA contamination in RNA samples. RNA quality and quantity were determined using Agilent

RNA6000 Nano kit by capillary electrophoresis (Agilent 2100 bioanalyzer, Agilent Technologies, Santa Clara, CA, USA). cDNA was synthesized from 1 µg of total RNA using oligo-dT primer and Superscript III reverse transcriptase (RT) in a 20 µl reaction (Invitrogen) according to the manufacturer's instruction. Each amplification was performed in a volume of 20 µl containing 1 µl of cDNA, 200 nM of each primer and 1X Brilliant III Ultra-Fast SybrGreen qPCR Master Mix (Agilent Technologies). The RT-qPCR reactions were performed on a QIAGEN Rotor-Gene Q instrument (Corbett Research, Qiagen) and included an initial denaturation at 95°C for 3 min followed by a two-step cycling protocol consisting of 45 cycles of denaturation at 95°C during 10 s and annealing/extension at 60°C during 10 s. The PCR cycling program was followed by a standard melt step, stepwise increasing temperature each 5 s by 1°C, ranging from 65°C to 95°C. Primers used for RT-qPCR of putative CpKMTs are listed in supplementary Table 2. The $2^{-\Delta\Delta C_t}$ method was used to calculate the relative expression levels of *KMT* genes with the constitutively expressed 18S rRNA gene as the internal reference and the ΔC_t value of the sporozoite stage as the calibrator.

Purification of histones and Western blot

Parasite histones were enriched by performing a fractionation protocol. Briefly, HCT-8 cells infected and non-infected at different PI time points were incubated in ice cold fractionation buffer (25mM Tris-HCl pH 8.5, 50 mM NaCl, 0.1 % Triton-X100, 1 mM EDTA, 1x protease inhibitor (cOmplete™ Protease Inhibitor Cocktail, Roche, USA). After dislocating the cells, the lysate was centrifuged at 2,000 g for 10 min at 4°C. Respective cellular fractions (pellet and supernatant) were subjected to histone purification using the EpiQuik™ Total Histones Extraction Kit (Epigentek, OP-0006-100, USA). The purified histone concentration was determined using micro BCA protein assay kit (Pierce, Thermofischer Scientific, USA). Approximately equal amounts of purified histones and parasite histones were separated by 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS- PAGE) and transferred to

nitrocellulose membranes (Millipore, USA). Chemiluminescent detection of bands was carried out by using Super Signal West Femto Maximum Sensitivity Substrate (Thermo Scientific, USA).

Cloning and protein expression of CpSETs

To express the SET domains proteins, putative active domains of CpSET1 (aa residues 2100 – 2244) and CpSET8 (aa residues 402 – 556) were *de novo* synthesized and cloned into the bacterial expression vector pET15b at the NdeI and BamHI cloning sites which follows N-terminal 6x histidine tag sequence (Gencust, Boynes, France). The expression plasmid was amplified in *E. coli* BL21 (DE3) cells. After induction with 0.1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 16°C for 16 h, the cells were collected by centrifugation. The cell pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM Imidazole, 1x protease inhibitor) then broken using bead beating method (Mixer Mill MM 400, Retsch, Germany). After centrifugation, his-tagged SET domains were recovered in the supernatant as soluble protein fraction.

***In vitro* histone methyltransferase assay**

To detect histone methyltransferase activity, 1 μ g of recombinant H4 (New England Biolabs) and 100 μ M S-adenosyl methionine (SAM) (Sigma, USA) were mixed with recombinant CpSET domains (CpSET8) in methyltransferase buffer (50 mM Tris-HCl (pH 8.5), 5 mM $MgCl_2$, 4 mM DTT) at 30°C for 1 h. The reaction was stopped by adding 2x SDS sample buffer, and the reaction mixture was analyzed on 15% SDS-PAGE, followed by western blotting with antibodies against tri-methylation of lysine 20 of histone 4 (H4K20Me3) at a dilution of 1:1000 (Supplementary table 1).

Signal quantification

Fluorescence intensity signals between infected and non-infected host tissue were quantified using ImageJ software version 1.52a (NIH, USA). Signal intensity was measured from individual nucleus of infected vs non-infected epithelial cell from intestinal crypts. For the statistical analysis, a mixed model was used to test the relationship between fluorescence intensity markers and group condition taking in account sample repetition. A mixed regression model was created considering fluorescence quantification as the main outcomes and sample identifier as random effect. The general significance level was set at a p-value below 0.05. All analyses were performed using packages nlme from the R statistical computing program (Version 4.1.1, date of release 8 October 2021; R Development Core Team, <http://www.R-project.org>, accessed on 12 January 2022) The data is represented using GraphPad Prism 9.1 (San Diego, California, USA)

Acknowledgments:

We wish to thank the Bioimaging Center Lille, Institut Pasteur de Lille, France for access to the microscopy instruments.

Author contributions:

Conceptualization: Manasi Sawant, Sadia Benamrouz-Vanneste, Jonathan Weitzman, Magali Chabé, Eric Viscogliosi, Gabriela Certad

Data curation: Manasi Sawant, Dionigia Meloni

Formal analysis: Manasi Sawant, Gaël Even

Funding acquisition: Jonathan Weitzman, Gabriela Certad

Investigation: Manasi Sawant, Sadia Benamrouz-Vanneste, Dionigia Meloni, Nausicaa

Gantois, Gaël Even, Karine Guyot, Colette Creusy, Erika Duval, Magali Chabé

Methodology: Manasi Sawant, René Wintjens, Jonathan Weitzman, Magali Chabé, Eric Viscogliosi, Gabriela Certad

Project administration: Jonathan Weitzman, Eric Viscogliosi, Gabriela Certad,

Resources: Jonathan Weitzman, Eric Viscogliosi, Gabriela Certad,

Supervision: Sadia Benamrouz-Vanneste, Jonathan Weitzman, Eric Viscogliosi, Gabriela Certad

Visualization: Manasi Sawant, Dionigia Meloni

Writing– original draft: Manasi Sawant, Sadia Benamrouz-Vanneste, Magali Chabé, Eric Viscogliosi, Gabriela Certad

Writing– review & editing: Manasi Sawant, Sadia Benamrouz-Vanneste, Dionigia Meloni, René Wintjens, Jonathan Weitzman, Magali Chabé, Eric Viscogliosi, Gabriela Certad

Funding:

This study was supported by the Plan Cancer “Epigénétique et cancer” 2015 (PARA-CAN #PARA-15-RCA), the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Institute Pasteur of Lille, the University of Lille and the Centre Hospitalier Regional Universitaire (CHRU) de Lille. M.S. was supported by a PhD fellowship from the University of Lille. R.W. is a Research Associate with the Belgian National Funds for Scientific Research (FRS-FNRS).

References

1. Arisue N, Hashimoto T. Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol Int.* 2015;64:254–9.
2. Ramirez NE, Ward LA, Sreevatsan S. A Review of the Biology and Epidemiology of Cryptosporidiosis in Humans and animals. *Microbes Infect.* 2004;6:773–85.
3. Khalil IA, Troeger C, Rao PC, Blacker BF, Brown A, Brewer TG, et al. Morbidity, Mortality, and Long-Term Consequences Associated with Diarrhoea from Cryptosporidium Infection in Children Younger Than 5 Years: A Meta-Analyses Study. *Lancet Glob Healt.* 2018;6:e758–68.

4. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al. Burden and Aetiology of Diarrhoeal Disease in Infants and Young Children in Developing Countries (The Global Enteric Multicenter Study, GEMS): A Prospective, Case-Control Study. *Lancet*. 2013;382:209–22.
5. Korich DG, Mead JR, Madore MS, Sinclair NA, Sterling CR. Effects of ozone, chlorine dioxide, chlorine, and monochloramine on *Cryptosporidium parvum* oocyst viability. *Appl Env Microbiol*. 1990;56:1423–8.
6. Baldursson S, Karanis P. Waterborne Transmission of Protozoan Parasites: Review of Worldwide Outbreaks—An Update 2004–2010. *Water Res*. 2011;45:6603–14.
7. Efstratiou A, Ongerth JE, Karanis P. Waterborne Transmission of Protozoan Parasites: Review of Worldwide Outbreaks—An Update 2011–2016. *Water Res*. 2017;114:14–22.
8. Checkley W, White Jr AC, Jaganath D, Arrowood MJ, Chalmers RM, Chen XM, et al. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *Lancet Infect Dis*. 2015;15(1):85–94.
9. Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, et al. The genome of *Cryptosporidium hominis*. *Nature*. 2004;431(7012):1107–12.
10. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 2004;304(5669):441–5.
11. Matos LVS, McEvoy J, Tzipori S, Brescian KDS, Widmer G. the transcriptome of *Cryptosporidium* oocysts and intracellular stages. *Sci Rep*. 2019;9:7859.
12. Lippuner C, Ramakrishnan C, Basso WU, Schmid MW, Okoniewski M, Smith NC, et al. RNA-Seq analysis during the life cycle of *Cryptosporidium parvum* reveals significant differential gene expression between proliferating stages in the intestine and infectious sporozoites. *Int J Parasitol*. 2018;48(6):413–22.
13. Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, et al. Genetic modification of the diarrhoeal pathogen *Cryptosporidium parvum*. *Nature*. 2015;523:477–80.
14. Tandel J, English ED, Sateriale A, Gullicksrud JA, Beiting DP, Sullivan MC, et al. Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium parvum*. *Nat Microbiol*. 2019;4:2226–36.
15. Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, et al. Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res*. 2004;14(9):1686–95.
16. Hakimi MA, Deitsch KW. Epigenetics in Apicomplexa: control of gene expression during cell cycle progression, differentiation and antigenic variation. *Curr Opin Microbiol*. 2007;10(4):357–62.
17. Kumar S, Chinnusamy V, Mohapatra T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front Genet*. 2018;18(9):640.

- 906 18. Husmann D, Gozani O. Histone lysine methyltransferases in biology and disease. *Nat Struct Mol*
907 *Biol.* 2019;26:26, 880–9.
- 908 19. Barnes CE, English DM, Cowley SM. Acetylation & Co: an expanding repertoire of histone
909 acylations regulates chromatin and transcription. *Essays Biochem.* 2019;63(1):97–107.
- 910 20. Gissot M, Choi SW, Thompson RF, Greally JM, Kim K. *Toxoplasma gondii* and *Cryptosporidium*
911 *parvum* Lack Detectable DNA Cytosine Methylation. *Eukaryot Cell.* 2008;7(3):537–40.
- 912 21. Vanagas L, Jeffers V, Bogado SS, Dalmasso MC, Sullivan Jr WJ, Angel SO. *Toxoplasma* histone
913 acetylation remodelers as novel drug targets. *Expert Rev Anti Infect Ther.* 2012;10(10):1189–201.
- 914 22. Chua MJ, Arnold MSJ, Xu W, Lancelot J, Lamotte S, Spath GF, et al. Effect of clinically approved
915 HDAC inhibitors on *Plasmodium*, *Leishmania* and *Schistosoma* parasite growth. *Int J Parasitol*
916 *Drugs Drug Resist.* 2017;7(1):42-50.
- 917 23. Guo F, Zhang H, McNair NN, Mead JR, Zhu G. The Existing Drug Vorinostat as a New Lead Against
918 *Cryptosporidiosis* by Targeting the Parasite Histone Deacetylases. *J Infect Dis.* 2018;217(7):1110–
919 7.
- 920 24. Sautel CF, Cannella D, Bastien O, Kieffer S, Aldebert D, Garin J, et al. SET8-mediated methylations
921 of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes. *Mol Cell*
922 *Biol.* 2007;27(16):5711–24.
- 923 25. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Rivas RH, Scherf A. 5' flanking region of var genes
924 nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in
925 malaria parasite. *Mol Microbiol.* 2007;66:1296–305.
- 926 26. Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, et al. Epigenetic memory at malaria
927 virulence genes. *Proc Natl Acad Sci U A.* 2007;104(3):899–902.
- 928 27. Cui L, Fan Q, Cui L, Miao J. Histone lysine methyltransferases and demethylases in *Plasmodium*
929 *falciparum*. *Int J Parasitol.* 2008;38:1083-1097.
- 930 28. Saksouk N, Bhatti MM, Kieffer S, Smith AT, Musset K, Garin J, et al. Histone-modifying complexes
931 regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma*
932 *gondii*. *Mol Cell Biol.* 2005;25:10301–14.
- 933 29. Villares M, Berthelet J, Weitzman JB. The clever strategies used by intracellular parasites to
934 hijack host gene expression. *Semin Immunopathol.* 2020;42:215–26.
- 935 30. Cock-Rada AM, Medjkane S, Janski N, Yousfi N, Perichon M, Chaussepied M, et al. SMYD3
936 promotes cancer invasion by epigenetic upregulation of the metalloproteinase MMP-9. *Cancer*
937 *Res.* 2012;72:810–20.
- 938 31. Sawant M, Baydoun M, Creusy C, Chabé M, Viscogliosi E, Certad G, et al. *Cryptosporidium* and
939 *Colon Cancer*: Cause or Consequence? *Microorganisms.* 2020;8:1665.
- 940 32. Dixon SE, Stilger KL, Elias EV, Naguleswaran A, Sullivan WJ., Jr. A decade of epigenetic research in
941 *Toxoplasma gondii*. *Mol Biochem Parasitol.* 2010;173:1–9.

- 942 33. Jain K, Fraser CS, Marunde MR, Parker MM, Sagum C, Burg JM, et al. Characterization of the
943 plasnt homeodomain (PHD) reader family for their hisotne tails interactions. *Epigenetics*
944 *Chromatin*. 2020;13:3.
- 945 34. Winston F, Allis CD. The bromodomain: a chromatin-targeting module? *Nat Struct Biol*.
946 1999;6:601–4.
- 947 35. Nishioka K, Rice JC, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, et al. PR-Set7 is a
948 nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated
949 with silent chromatin. *Mol Cell*. 2002;9:1201–13.
- 950 36. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Tolga Emre NC, et al. Active
951 genes are tri-methylated at K4 of histone H3. *Nature*. 2002;419:407–11.
- 952 37. Zhang Y. It takes a PHD to interpret histone methylation. *Nat Struct Mol Biol*. 2006;13:572–4.
- 953 38. Southhall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural basis for the requirement of
954 additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell*.
955 2009;33:181–91.
- 956 39. Mishra BP, Ansari KI, Mandal SS. Dynamic association of MLL1, H3K4 trimethylation
957 withchromatin andHoxgene expression during the cell cycle. *FEBS J*. 2009;276:1629–40.
- 958 40. Gissot M, KIM K. How Epigenomics Contributes to the Understanding of Gene Regulation in
959 *Toxoplasma gondii*. *J Eukaryot Microbiol*. 2008;55:476–480.
- 960 41. Cheeseman, K, Jannot, G, Lourenço, N, Villares M, Berthelet J, Calegari-Silva T, et al. Dynamic
961 methylation of histone H3K18 in differentiating *Theileria* parasites. *Nat Commun*. 2021;12(3221).
- 962 42. Li Y, Han J, Zhang Y, Cao F, Liu Z, Li S, et al. Structural basis for activity regulation of MLL family
963 methyltransferases. *Nature*. 2016;530:447–52.
- 964 43. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine
965 methyltransferases. *Genome Biol*. 2005;6:227.
- 966 44. Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation of histone H3. *Nat*
967 *Rev Mol Cell Bio*. 2012;13:115–26.
- 968 45. Hacker KE, Fahey CC, Shinsky SA, Chiang YCJ, DiFiore JV, Jha DK, et al. Structure/Function Analysis
969 of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain
970 Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation. *J Biol Chem*.
971 2016;291:21283–95.
- 972 46. Sims RJ III, Nishioka K, Reinberg D. Histone lysine methylation: a signature for chromatin
973 function. *Trends Genet*. 2003;19:629–39.
- 974 47. Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, Frenkiel TA, et al. Specificity and mechanism of
975 the histone methyltransferase Pr-Set7. *Genes Dev*. 2005;19:1444–54.
- 976 48. Sautel CF, Cannella D, Bastien O, Kieffer S, Aldebert D, Garin J, et al. SET8-mediated methylations
977 of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes. *Mol Cell*
978 *Biol*. 2007;27:5711–24.

979 49. Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, Selker EU, et al. Structural Basis for the Product
980 Specificity of Histone Lysine Methyltransferases. *Mol Cell*. 2003;12:177–85.

981 50. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, et al. SMYD3 encodes a histone
982 methyltransferase involved in the proliferation of cancer cells. *Nat Cell Bio*. 2004;6:731–40.

983 51. Van Aller GS, Reynoird N, Barbash O, Huddleston M, Liu S, Zmoos AF, et al. Smyd3 regulates
984 cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation. *Epigenetics*. 2012;7:340–3.

985 52. Sivagurunathan S, Heaslip A, Liu J, Hu K. Identificaiton of functional modules of AKMT, a novel
986 lysine methyltransferase regulating the motility of *Toxoplasma gondii*. *Mol Biochem Parasitol*.
987 2013;189:43–53.

988 53. Pivovarova Y, Liu J, Lesigang J, Koldyka O, Rauschmeier R, Hu K, et al. Structure of a Novel
989 Dimeric SET Domain Methyltransferase that Regulates Cell Motility. *J Mol Biol*. 2018;430:4209–
990 29.

991 54. Heaslip AT, Nishi M, Stein B, Hu K. The motility of a human parasite, *Toxoplasma gondii*, is
992 regulated by a novel lysine methyltransferase. *PLoS Pathog*. 2011;7:e1002201.

993 55. Gillrie MR, Lee K, Gowda DC, Davis SP, Monestier M, Cui L, et al. *Plasmodium falciparum* histones
994 induce endothelial proinflammatory response and barrier dysfunction. *Am J Pathol*.
995 2012;180:1028–39.

996 56. Luah YH, Chahal BK, Ziying Ong E, Bozdech Z. A moonlighting function of *Plasmodium falciparum*
997 histone 3, mono-methylated at Lysine 9? *PLoS One*. 2010;5:e10252.

998 57. Pereira D, Richert A, Medjkane S, Hénon S, Weitzman JB. Cell geometry and the cytoskeleton
999 impact the nucleo-cytoplasmic localisation of the SMYD3 methyltransferase. *Sci Rep*.
1000 2020;10:20598.

1001 58. Sautel CF, Ortet P, Saksouk N, Kieffer S, Garin J, Bastien O, et al. The histone methylase KMTx
1002 interacts with the redox-sensor peroxiredoxin-1 and targets genes involved in *Toxoplasma gondii*
1003 antioxidant defences. *Mol Microbiol*. 2009;77:212–26.

1004 59. Rea S, Eisenhaber F, O’Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin
1005 structure by site-specific histone H3 methyltransferases. *Nature*. 2000;406:593–9.

1006 60. Rider SD Jr, Zhu G. *Cryptosporidium*: genomic and biochemical features. *Exp Parasitol*.
1007 2010;124:2–9.

1008 61. Wang Y, Gong A-Y, Ma S, Chen X, Strauss-Soukup JK, Chen X-M. Delivery of parasite
1009 Cdg7_Flc_0990 RNA transcript into intestinal epithelial cells during *Cryptosporidium parvum*
1010 infection suppresses host cell gene transcription through epigenetic mechanisms. *Cell Microbiol*.
1011 2017;19:e12760.

1012 62. Zhang T, Cooper S, Brockforff N. The interplay of histone modifications - writers that read. *EMBO*
1013 *Rep*. 2015;16:1467–81.

1014 63. Forney JR, Dewald DB, Yang S, Speer CA, Healey MC. A role for host phosphoinositide 3-kinase
1015 and cytoskeletal remodeling during *Cryptosporidium parvum* infection. *Infect Immun*.
1016 1999;67:844–52.

1017 64. Liu X, Li Z, Song Y, Wang R, Han L, Wang Q, et al. AURKA induces EMT by regulating histone
1018 modification through Wnt/ β -catenin and PI3K/Akt signaling pathway in gastric cancer.
1019 *Oncotarget*. 2016;7:33152–64.

1020 65. Sawant M, Benamrouz-Vanneste S, Mouray A, Bouquet P, Gantois N, Creusy C, et al. Persistent
1021 *Cryptosporidium parvum* Infection Leads to the Development of the Tumor Microenvironment in
1022 an Experimental Mouse Model: Results of a Microarray Approach. *Microorganisms*. 2021;9:2569.

1023 66. Lecoeur H, Prina E, Rosazza T, Kokou K, N'Diaye P, Aulner N, et al. Targeting macrophage histone
1024 H3 modification as a *Leishmania* strategy to dampen the NF- κ B/NLRP3-mediated inflammatory
1025 response. *Cell Rep*. 2020;30:1870-1882.e4.

1026 67. Braun L, Brenier-Pinchart MP, Hammoudi PM, Cannella D, Kieffer-Jaquinod S, Vollaie J, et al. The
1027 *Toxoplasma* effector TEEGR promotes parasite persistence by modulating NF- κ B signalling via
1028 EZH2. *Nat Microbiol*. 2019;4:1208–20.

1029 68. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis
1030 across computing platforms. *Mol Biol Evol*. 2018;35:1547–9.

1031 69. Williams CJ, Head JJ, Moriarty NW, Prisant MG, Videau LW, Deis LN, et al. MolProbity: More and
1032 better reference data for improved all-atom structure validation. *Protein Sci*. 2018;27:293-315.

1033 70. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-
1034 atom structure validation for macromolecular crystallography. *Acta Crystallogr Biol Crystallogr*.
1035 2009;66:12–21.

1036 71. Holm L. Using Dali for Protein Structure Comparison. *Methods Mol Biol*. 2020;2112:29–42.

1037 72. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX:
1038 Structure visualization for researchers, educators, and developers. *Protein Sci*. 2021;30:70–82.

1039 73. Dibao-Dina A, Follet J, Ibrahim M, Vlandas A, Senez V. Electrical impedance sensor for
1040 quantitative monitoring of infection processes on HCT-8 cells by the waterborne parasite
1041 *Cryptosporidium*. *Biosens Bioelectron*. 2015;66:69–76.

1042 74. Keegan AR, Fanok S, Monis PT, Saint CP. Cell culture-Taqman PCR assay for evaluation of
1043 *Cryptosporidium parvum* disinfection. *Appl Env Microbiol*. 2003;69:2505–11.

1044 75. Certad G, Ngouanesavanh T, Guyot K, Gantois N, Chassat T, Mouray A, et al. *Cryptosporidium*
1045 *parvum*, a Potential Cause of Colic Adenocarcinoma. *Infect Agents Cancer*. 2007;2:22.

1046