Adenine DNA methylation, 3D genome organization, and gene expression in the parasite *Trichomonas vaginalis*

Ayelen Lizarraga[a], Zach Klapholz O'Brown[b,c], Konstantinos Boulias[b,c], Lara Roach[b,c], Eric Lieberman Greer[b,c], Patricia J. Johnson[d,1], Pablo H. Strobl-Mazzulla[e], and Natalia de Miguel[a,1]

[a]Laboratorio de Parásitos Anaerobios, Instituto Tecnológico Chascomús, National Scientific and Technical Research Council-National University of San Martin (CONICET-UNSAM), B7130IWA Chascomús, Argentina; [b]Division of Newborn Medicine, Children's Hospital Boston, Boston, MA 02115; [c]Department of Pediatrics, Harvard Medical School, Boston, MA 02115; [d]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, CA 90095; and [e]Laboratorio de Biología del Desarrollo, Instituto Tecnológico Chascomús, CONICET-UNSAM, B7130IWA Chascomús, Argentina

Contributed by Patricia J. Johnson, March 9, 2020 (sent for review October 4, 2019; reviewed by Robert P. Hirt and Alvaro Rada-Iglesias)

Trichomonas vaginalis **is a common sexually transmitted parasite that colonizes the human urogenital tract causing infections that range from asymptomatic to highly inflammatory. Recent works have highlighted the importance of histone modifications in the regulation of transcription and parasite pathogenesis. However, the nature of DNA methylation in the parasite remains unexplored. Using a combination of immunological techniques and ultrahigh-performance liquid chromatography (UHPLC), we analyzed the abundance of DNA methylation in strains with differential pathogenicity demonstrating that N6-methyladenine (6mA), and not 5-methylcytosine (5mC), is the main DNA methylation mark in *T. vaginalis*. Genome-wide distribution of 6mA reveals that this mark is enriched at intergenic regions, with a preference for certain superfamilies of DNA transposable elements. We show that 6mA in *T. vaginalis* is associated with silencing when present on genes. Interestingly, bioinformatics analysis revealed the presence of transcriptionally active or repressive intervals flanked by 6mA-enriched regions, and results from chromatin conformation capture (3C) experiments suggest these 6mA flanked regions are in close spatial proximity. These associations were disrupted when parasites were treated with the demethylation activator ascorbic acid. This finding revealed a role for 6mA in modulating three-dimensional (3D) chromatin structure and gene expression in this divergent member of the Excavata.**

adenine methylation | *Trichomonas* | gene expression

Trichomonas vaginalis is a flagellated protozoan parasite responsible for trichomoniasis, the most frequent nonviral sexually transmitted infection. Although generally asymptomatic (1), *T. vaginalis* infection can cause numerous pathologies like vaginitis, urethritis, prostatitis, and various complications when infection is acquired during pregnancy (2, 3). Chronic infection has been associated with an increased risk of acquiring and transmitting HIV (4) and higher susceptibility to developing cervical or prostate cancer (5, 6).

With ∼160 Mb distributed over six chromosomes, the genome of *T. vaginalis* is the largest protozoan parasite genome sequenced to date (7). To date, the highly repetitive nature of the genome (up to 65% composed of repetitive elements) prevented the complete assembly of the scaffolds and subsequent studies of *T. vaginalis* chromosome architecture (7). Analysis of the age distribution of gene families coupled with the large genome size and low repeat polymorphism indicates the genome underwent one or more large-scale genome duplication events (7). Intriguingly, only half of the ∼46,000 protein-coding genes appear to be expressed (8, 9), and individual members within gene families show differential regulation in response to different stimuli (8–10). These results indicate that extrinsic and intrinsic factors accurately control the fine tuning of gene expression in this extracellular parasite.

Studies on *T. vaginalis* transcriptional regulation mostly focused on identifying cis-regulatory elements and their cognate transcription factors. Few core promoter elements or transcription factors have been identified to date (11). The metazoan-like Initiator (Inr) element (12) present in ∼75% of all protein-coding genes (7) was found to be the main core promoter element responsible for directing basal transcription in *T. vaginalis*. A novel initiator binding protein, IBP39, was identified and demonstrated to specifically recognize *T. vaginalis* Inr and directly interact with RNAPII (12, 13). Additionally, several conserved motifs that resemble metazoan Myb recognition elements (14) as well as proteins containing an Myb DNA binding domain, like novel transcription factor M3BP (14), have also been found in *T. vaginalis* (11). These Myb recognition elements can act as cis-regulatory elements directing the transcription start site (TSS) selection (14) and transcription of specific genes (15). However, little is known about other mechanisms of gene regulation in this parasite (11). Epigenetic mechanisms regulate gene expression by modifying

Significance

Trichomonas vaginalis **is a common sexually transmitted parasite, yet little is known about the regulation of gene expression in this parasite. We demonstrate that N6-methyladenine (6mA) is the main methylation mark in the *T. vaginalis* genome. 6mA is widespread in DNA of eubacterial genera but uncommon in genomes of most eukaryotes. Examination of the genome-wide distribution of 6mA reveals a preference for transposable elements and intergenic regions. Transcriptionally active or repressive intervals are found to be flanked by 6mA-enriched regions, and data suggest that 6mA flanked regions are in close three-dimensional (3D) spatial proximity. Our findings describe the presence of DNA methylation in *T. vaginalis* and reveal a role for 6mA in modulating 3D chromatin architecture.**

Author contributions: A.L., P.J.J., P.H.S.-M., and N.d.M. designed research; A.L., Z.K.O., K.B., L.R., and N.d.M. performed research; E.L.G. and P.J.J. contributed new reagents/analytic tools; A.L., Z.K.O., E.L.G., P.J.J., P.H.S.-M., and N.d.M. analyzed data; and A.L. and N.d.M. wrote the paper.

Reviewers: R.P.H., Newcastle University; and A.R.-I., Consejo Superior de Investigaciones Científicas–Universidad de Cantabria–Sociedad para al Desarollo de Cantabria.

The authors declare no competing interest.

Published under the PNAS license.

Data deposition: All sequencing data that support the findings of this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (accession no. PRJNA526331).

[1]To whom correspondence may be addressed. Email: johnsonp@ucla.edu or ndemiguel@intech.gov.ar.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1917286117/-/DCSupplemental.

Downloaded at Akademiska Sjukhuset on May 28, 2020

chromatin accessibility to transcription factors and other components of the transcriptional machinery (16). Specifically, histone modifications have been shown to be of importance in the regulation of gene expression and antigenic variation of other unicellular parasites (17). Recent studies have demonstrated that these modifications are present in *T. vaginalis* and play an essential role in transcriptional regulation and parasite pathogenesis (18, 19).

DNA methylation is one of the major epigenetic mechanisms involved in gene regulation. The most common DNA modification in eukaryotes is C5-methylcytosine (5mC), an epigenetic mark usually associated with gene silencing (20) that participates in a wide variety of processes such as imprinting, X chromosome inactivation, and silencing of transposable elements (TEs) (21). Although common among higher eukaryotes, 5mC is not universally present in unicellular organisms (17), and thus, few protozoan parasites with detectable levels of 5mC in their genome have been identified and its function characterized (17, 22–24). In contrast, N6-methyladenine (6mA) is the most prevalent DNA modification in prokaryotes (25), and early reports debated the existence and abundance of this mark in eukaryotes (26). However, in recent years numerous reports have greatly expanded the list of organisms with 6mA to include flies (27), nematodes (28), green algae (29), and even vertebrates (30–32). The abundance of this mark varies dramatically between species, with levels of 6mA that range from 6 to 7 parts per million of total adenines in mouse embryonic stem cells (30) up to 2.8% of total adenines in early-diverging fungi (33). Likewise, this variation between organisms was also evidenced in their genomic distribution, suggesting that 6mA might have species-specific functions (34). Intriguingly, in some cases the role of 6mA has been associated with active transcription in *Arabidopsis* (35), *Drosophila* (27), *Chlamydomonas* (29), and early-diverging fungi (33). Only two reports describe 6mA as a repressive mark similar to 5mC (30, 36).

Despite its importance in other organisms, the methylation status of *T. vaginalis* DNA remains unexplored. In this work, using a combination of immunological techniques and ultrahigh-performance liquid chromatography coupled with mass spectrometry (UHPLC-ms/ms), we identify the presence of 6mA and 5mC in the genomic DNA (gDNA) of *T. vaginalis*. We interrogate the distribution of 6mA across the genome using an adapted methylated DNA immunoprecipitation assay followed by high-throughput sequencing (MeDIP-seq). Interestingly, we found that 6mA is enriched at intergenic regions, with a preference for repetitive elements. Comparison with published RNA sequencing (RNA-seq) data of the same strain (37) suggests that 6mA is associated with silenced transcription when found on gene bodies. Finally, results from chromatin conformation capture (3C) experiments suggest that 6mA is associated with chromatin loop formation, raising the intriguing possibility that 6mA could play a role in modulating three-dimensional (3D) genome architecture. DNA methylation is described in *T. vaginalis*, and our data highlight a role for 6mA in epigenetic gene regulation in this deep-branching eukaryote.

Results

Identification of 6mA and 5mC in *T. vaginalis* Genomic DNA. Given the importance of 5mC and 6mA in other organisms, we set out to investigate the presence of these marks in *T. vaginalis* genomic DNA. We first evaluated the presence of 5mC in *T. vaginalis* by a dot blot assay using genomic DNA extracted from two *T. vaginalis* strains, B7268 and G3, with high and low adherence capacity, respectively (Fig. 1*A*). Unexpectedly, we failed to detect 5mC in either strain, contrary to our chicken gDNA control (Fig. 1*A*). Likewise, although we were able to detect m5C in the RNA, this mark was undetectable by immunofluorescence in the nucleus of the parasite (Fig. 1*B*). These data suggest the absence, or very low abundance, of 5mC mark in *T. vaginalis* DNA. On the other hand, we revealed the presence of 6mA mark in both the nucleus and cytoplasm by immunofluorescence (Fig. 1*C*) using antibodies that specifically recognize methylated adenines (*SI Appendix*, Fig. S1*A*). Importantly, parasites treated with RNase revealed 6mA exclusively detected in the nucleus, demonstrating that this epigenetic mark is present in the DNA of *T. vaginalis*. Similarly, using restriction enzymes MboI and DpnI, which digest unmethylated and methylated GATC sequence, respectively (38), we found that *T. vaginalis* genome possesses both methylated and unmethylated adenines in the GATC context (Fig. 1*D*). In order to determine whether parasite adherence could be correlated with 6mA abundance, we performed dot blot assays (Fig. 1*E*) and UHPLC followed by mass spectrometry (Fig. 1*F*) using gDNA from an adherent (B7268) and a less adherent (G3) strain (39). The UHPLC-ms/ms experiments confirmed, in an antibody-independent manner, that gDNA was methylated on the N6 position of adenines (Fig. 1*F*). However, we found no difference on the abundance of 6mA levels among them. Importantly, we demonstrated that 6mA is present in 2.5% of all adenines, whereas extremely low levels of 5mC were detected (0.004 to 0.008% of all cytosines) by UHPLC followed by mass spectrometry (Fig. 1*F*). To exclude the possibility of signal contamination from *Mycoplasma hominis*, a common bacterial symbiont of *T. vaginalis* (40), parasite cultures were subjected to antibiotic treatment and checked for the elimination of *M. hominis* prior to the experiments (*SI Appendix*, Fig. S1*B*). These data demonstrate the lack of the bacterial symbiont and confirm that the detected 5mC and 6mA are in *T. vaginalis* genomic DNA. Taken together, our results demonstrated that 6mA, but not 5mC, is the predominant DNA methylation mark in *T. vaginalis*.

To be considered an epigenetic mark, 6mA needs to be maintained after DNA replication. Our observations using hydroxyurea (HU) synchronized cultures revealed the expected decrease in 6mA levels in S phase due to DNA replication, which remained during the G2/M phase of the cell cycle, and restored original abundance after cell division (*SI Appendix*, Fig. S2). These observations support the presence of active methyltransferase(s) responsible for maintaining 6mA levels after cell division.

Genome-Wide Mapping of 6mA in *T. vaginalis*. To determine the function of DNA methylation in *T. vaginalis*, it is essential to identify its genomic distribution. To this end, we performed an adapted MeDIP-seq (41) on the adherent strain B7268 (Fig. 2*A*). Unfortunately, due to the extremely low levels of m5C in *T. vaginalis*, we were unable to get good-quality reads from two independent MeDIP-seq experiments. Therefore, we decided to focus exclusively on m6A for further analysis. To this end, two independent high-throughput sequencing experiments, each containing pooled DNA from three independent immunoprecipitation assays, were performed. In order to evaluate specificity and enrichments of 6mA in the immunoprecipitated fraction when using a specific anti-6mA antibody, a dot blot assay was performed using an immunoglobulin G (IgG) control antibody (Fig. 2*B*).

After quality filtering (q value <0.01, fold enrichment [FE] > 2), we obtained a total of 10,388 6mA peaks found in both replicates, corresponding to 92.1 and 75.6% of all peaks present in each independent MeDIP-seq experiment (Fig. 2*C* and Dataset S1). In order to validate these results, four genes selected from the MeDIP-seq were analyzed by MeDIP-qPCR, demonstrating an enrichment in 6mA immunoprecipitated sample compared with sample immunoprecipitated with IgG control antibody (*SI Appendix*, Fig. S3*A*). The most prevalent motif, found in 4,692 peaks, was CAATACCC (P value =1e-1642) (*SI Appendix*, Fig. S3*B*). Remarkably, peak visualization in a genome browser suggested a clear enrichment of peaks in intergenic regions (Fig. 2*D*). Indeed, peak annotation revealed that 94% of all MeDIP-seq peaks (9,746 6mA peaks) were intergenic (Fig. 3*A*). Importantly, a shuffled distribution

Downloaded at Akademiska Sjukhuset on May 28, 2020

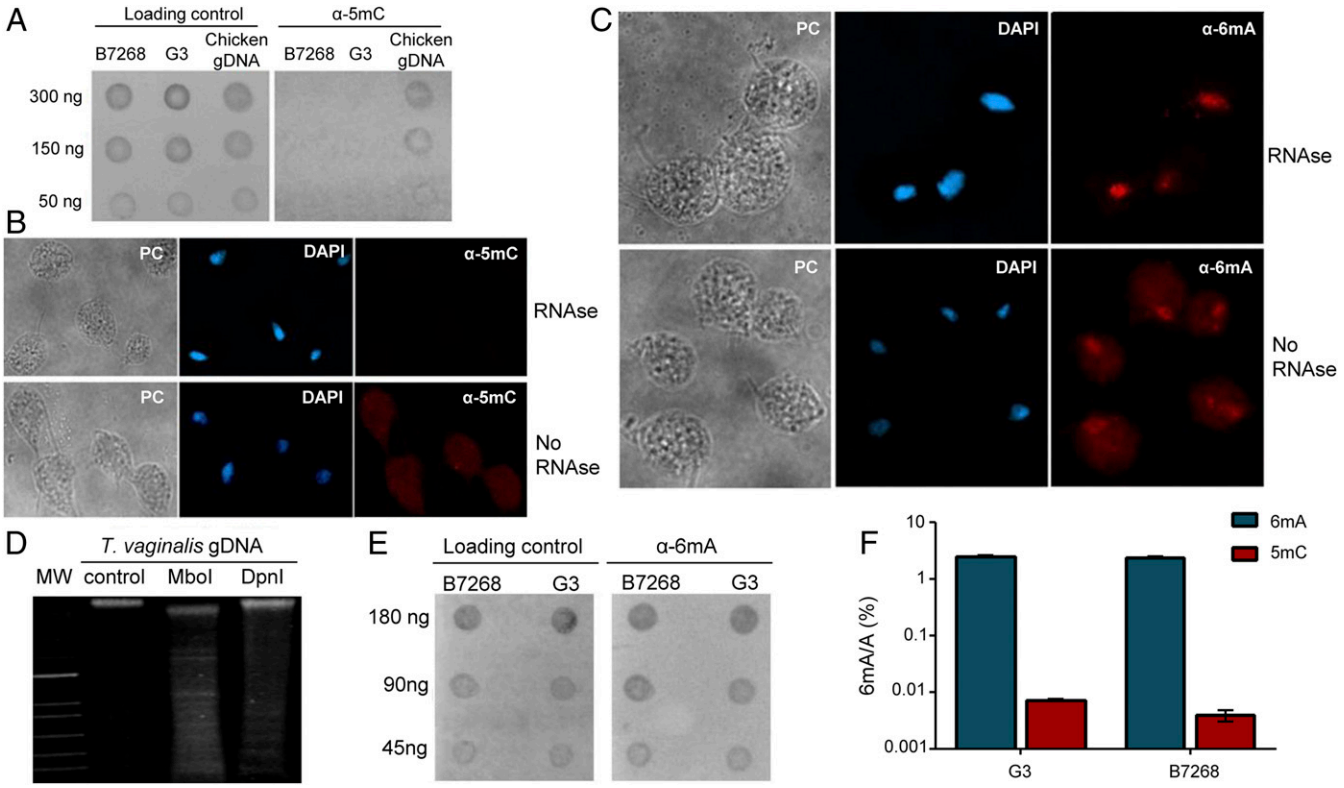


Fig. 1. DNA methylation in *T. vaginalis* genome. (*A*) Dot blot assay to detect 5mC shows lack of signal in the genomic DNA of highly adherent strain B7268 and less adherent strain G3. (*Left*) gDNA dots stained with methylene blue for loading control. (*Right*) 5mC signal. Chicken gDNA was used as positive control. (*B*) Immunofluorescence assay using specific antibodies detects 5mC in the cytoplasm (RNA) but fails to detect 5mC in *T. vaginalis* nucleus of RNase-treated parasites. Nucleus was stained with DAPI. PC, phase contrast. (*C*) Immunofluorescence assay using anti-6mA–specific antibody evidenced this modification in the cytoplasm (RNA) and nucleus (gDNA) or exclusively in the nucleus of RNase-treated parasites. Nucleus was stained with DAPI. (*D*) Methylation-sensitive restriction enzyme assay shows the presence of methylated and unmethylated adenines in *T. vaginalis* gDNA. Lane 1: molecular weight size marker (MW). Lane 2: undigested gDNA. Lane 3: MboI-treated gDNA (cuts GATC sites). Lane 4: DpnI-treated gDNA (cuts G6mATC sites). (*E*) Anti-6mA dot blot assay of RNA-free gDNA shows similar abundance of 6mA in the genomic DNA of strain B7268 and strain G3. (*Left*) gDNA dots stained with methylene blue for loading control. (*Right*) 6mA signal. (*F*) Quantification of methylated bases in the gDNA of adherent strain B7268 and less adherent strain G3 using UHPLC followed by mass spectrometry shows high levels of 6mA and low levels of 5mC but not differences among stains for both modifications. Each column represents the mean and SD of three independent experiments per group. Graphic is shown in logarithmic scale.

of peaks across *T. vaginalis* genome would result in an even distribution between genes and intergenic regions (Fig. 3*A*), confirming that 6mA is remarkably enriched in intergenic regions.

Among the 642 6mA peaks located in genes, most were distributed between the coding region (48%) and the transcription termination sites (TTSs; 43%), with only 9% found in the TSSs (Fig. 3*B*). Comparing those numbers with the percentages that would occur if the peaks had a shuffled distribution (22% in the TTS, 60% in the coding region, and 12% in the TSS), our results revealed an enrichment of 6mA in the TTS when located in genes (Fig. 3*B*). In order to understand the functional role of 6mA, we retrieved the gene IDs of the methylated genes (Dataset S2) and found that 70% belong to different families of TEs, with the majority (61%) belonging to families previously described in *T. vaginalis* according to Repbase (Fig. 3*C*). The remaining genes could be divided into ones with conserved domains that allow the assignment of a predicted role (8%) and hypothetical proteins with unknown functions (22%) found exclusively in *T. vaginalis* (Fig. 3*C*). Intriguingly, the observed enrichment of 6mA peaks in the TTS was found exclusively in TE genes, suggesting a particular role for methylation at the 3′ of these TE genes. Gene ontology (GO) analysis of the methylated genes (excluding genes from TEs) revealed an enrichment of genes involved in protein modification and phosphorylation (P value <0.05) (*SI Appendix*, Table S1). However, it is important to take into account that most of the non-TE genes have no identifiable domain (hypothetical proteins) and therefore, were not considered in the GO enrichment analysis. Nonetheless, it is interesting to note that 13% of the methylated hypothetical proteins possess predicted transmembrane domains (Dataset S2); however, whether these genes play any role in parasite pathogenesis remains to be determined.

Considering that 6mA has been previously described as an epigenetic mark of TEs in other organisms (27, 30) and that 70% of *T. vaginalis* methylated genes belong to TEs, we analyzed if intergenic peaks were also located within TEs. Analysis using RepeatMasker revealed that of 94% of intergenic peaks, 57.4% were found in *T. vaginalis* TEs previously described in Repbase. Most of the peaks were found in Mutator, Kolobok, and Polinton superfamilies of *T. vaginalis* DNA transposons (Fig. 3*D*). However, if we consider the numbers of TEs within the genome, repeat element DNA7 and TE superfamilies P and Kolobok were the most enriched in the MeDIP-seq (*SI Appendix*, Table S2). Surprisingly, no methylation was found in NesL, the only retrotransposon family found in the parasite (*SI Appendix*, Table S2), indicating a preference of intergenic 6mA for certain types of TEs and raising the question of whether 6mA could act as a regulator of specific TEs.

6mA Methylation and Gene Expression. 6mA has been found to be an epigenetic mark of active genes in *Arabidopsis* (35), *Drosophila* (27), and *Chlamydomonas* (29) or repressed genes in mouse embryonic stem cells (30) and rice (36). To examine whether this mark

Downloaded at Akademiska Sjukhuset on May 28, 2020

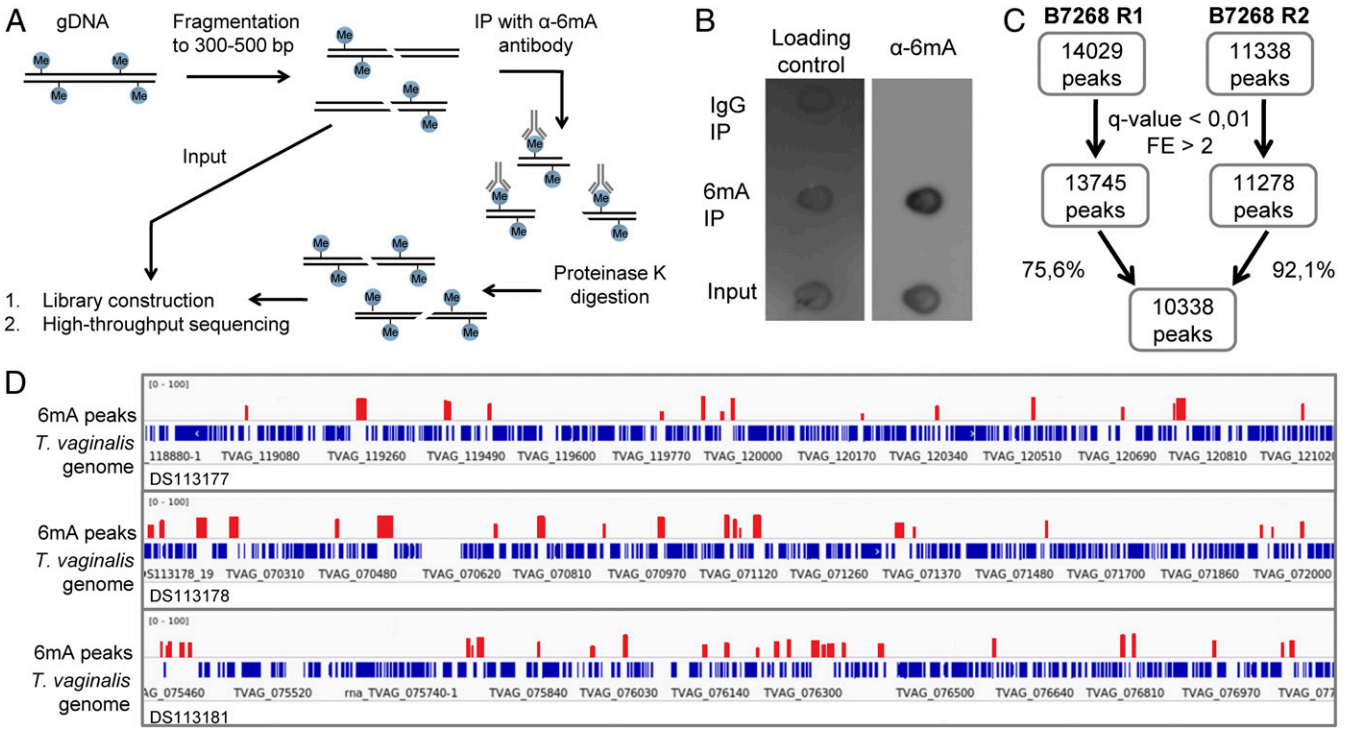


Fig. 2. 6mA MeDIP-seq of adherent strain B7268. (*A*) Schematic representation of the adapted MeDIP-seq protocol. Me, methylated adenine. (*B*) Dot blot assay of the different MeDIP fractions shows enrichment of 6mA-containing fragments in the 6mA–IP fraction compared with the input control. Lack of signal in the IgG–IP fraction shows the specificity of 6mA antibody. (*C*) Workflow showing final peak count obtained from two independent 6mA MeDIP-seq experiments. Peaks were filtered by *q* value (<0.01) and FE >2. Only peaks with at least 75% overlap between experiments were considered for further study. (*D*) Genome browser visualization of 6mA peaks. Representative image of three different contigs (DS113177, DS113178, and DS113181) shows that 6mA possesses a primarily intergenic localization. Red bars: 6mA peaks. Blue bars: *T. vaginalis* genes.

has a role in transcription in *T. vaginalis*, we made use of publicly available RNA-seq data from B7268 strain (37). To test if these data were applicable for our analysis, we divided the genes into low (reads per kilobase of exon model per million mapped reads [RPKM] ≤ 1), moderate (1 < RPKM ≤ 20), and high (RPKM > 20) expression and selected some well-characterized genes from each group to measure their relative expression levels via qRT-PCR. Our results showed that the relative expression of each gene was in accordance to what had been described previously (19, 42), and more importantly, we could rank the genes based on their relative expression obtained by qRT-PCR similarly to ranking obtained by gene expression RPKM (*SI Appendix*, Table S3), validating the use of the RNA-seq in our analysis.

We next looked into the expression levels of the methylated genes using RNA-seq data available and maintaining the same criteria for low, moderate, and high expression. Considering that the majority of those genes belong to *T. vaginalis* TEs (Fig. 3*C*), it was not surprising that gene expression analysis revealed that over 80% of all methylated genes are poorly expressed, with 14% possessing moderate expression and only a small percentage (4%) highly expressed (Fig. 4*A*). However, when we considered the expression levels of the methylated genes without taking into account TE genes, poorly expressed genes were still the most abundant group (over 50%) (Dataset S3), suggesting 6mA could be a sign of repressed expression when found on genes.

As described previously, most of the MeDIP-seq peaks have an intergenic localization (Fig. 3*A*). In order to evaluate if 6mA could be affecting the expression of adjacent genes as has been described for 6mA in mouse embryonic stem cells (30), we analyzed the expression level of the genes closest to intergenic methylation (Fig. 4*B*). We obtained a total of 6,000 genes found near 6,714 peaks. However, due to the fragmented nature of *T. vaginalis* reference genome, not all of the intergenic peaks could be considered, and it is likely that the real number of genes near intergenic 6mA could be higher. Surprisingly, in contrast with our observations regarding the expression levels of methylated genes (Fig. 4*A*), we found a higher percentage of expressed genes (13.5% with moderate and 16.4% with high expression) near intergenic methylation compared with poorly expressed genes (Fig. 4*B*). In fact, genes with intergenic 6mA peaks nearby were expressed at significantly higher levels than genes in unmethylated regions (Fig. 4*C*). Finally, we examined the association between the 6mA peaks distance and expression level of associated genes. We observed that intergenic 6mA is primarily located between −1,500 and +1,000 bp of the nearest genes regardless of their expression level (Fig. 4*D*). Intriguingly, nonexpressed genes are found significantly closer to their nearest upstream or downstream intergenic peak than expressed genes (*SI Appendix*, Fig. S3 *B* and *C*). These observations suggest that the effect of 6mA on gene expression in *T. vaginalis* depends on its relative position to the corresponding gene.

6mA Delimits Chromatin Loops. During the analysis of the expression levels of genes near intergenic methylation regions, we noticed various instances of intergenic 6mA peaks flanking sections of the genome where all of the genes contained within had similar expression levels. A closer look revealed the presence of 1,046 of 2,937 such intervals with more than one gene: 452 containing genes that were not expressed (repressive intervals) and 594 containing expressed genes (active intervals) (Fig. 5*A*). These intervals varied in the number of genes they contained (from 2 to up to 36 genes) (Fig. 5*B*) as well as in length, with some spanning over 20 kb (Dataset S4). As can be observed in Fig. 5*B*, intervals containing a greater number of genes tend to

Downloaded at Akademiska Sjukhuset on May 28, 2020

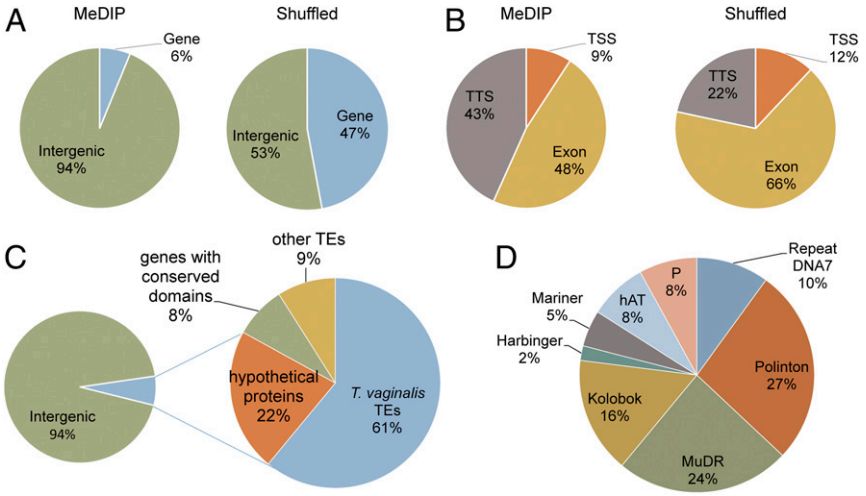
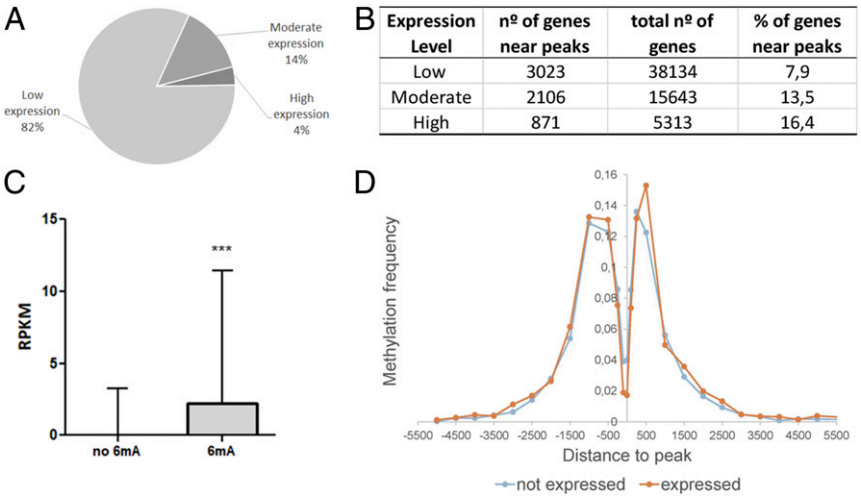


Fig. 3. Genome distribution of 6mA peaks. (*A*) 6mA is enriched at intergenic regions. Annotated peaks were grouped according to their genomic localization: peaks that fall within the TSSs, coding region, and TTSs were considered gene peaks and grouped together, and the rest of the peaks were considered intergenic. MeDIP: Percentages of intergenic and gene peaks according MeDIP-seq analysis. Shuffled: MeDIP-seq peaks were randomly distributed across the genome, and the percentage of gene and intergenic peaks was calculated. (*B*) Enrichment of 6mA at the TTS region of methylated genes. MeDIP: Percentages of peaks within the TSS, exon, and TTS regions of methylated genes according MeDIP-seq analysis. Shuffled: Peaks randomly distributed across the genome. (*C*) Identity of *T. vaginalis* methylated genes. A total of 642 MeDIP peaks fall within 512 genes: 61% belong to TEs described in the parasite (*T. vaginalis* TEs), 9% of genes possess sequence similarity to repetitive elements found in other organisms (other TEs), 22% are *T. vaginalis*-specific hypothetical proteins with unknown functions (hypothetical proteins), and 8% are genes with sequence similarity to genes from other organisms (genes with known domains). (*D*) Identity of *T. vaginalis* methylated repetitive elements. Only intergenic peaks with more than 50% overlap with a *T. vaginalis* TE were considered, obtaining a total of 5,596 peaks. Most represented DNA transposon superfamilies included Mutator (MuDR) with 24% of all peaks, Polinton (27%), and Kolobok (16%). Less represented TE superfamilies included hAT (8%), Mariner (5%), and P (8%); 10% of the peaks were found in *T. vaginalis* DNA repeat DNA7.

be silenced. Conversely, most of the intervals containing fewer genes are actively transcribed.

In other organisms, it is well known that the 3D genome organization within the nucleus is an important element in the regulation of gene expression (43). In particular, chromatin looping is a type of intrachromosomal interaction that has been shown to influence gene expression (44). Recent work in mammalian cells demonstrated that 5mC was capable of regulating the expression of two genes 10 kbp apart via chromatin loop formation (45). Based on these antecedents and our bioinformatics observations, we performed a 3C assay (Fig. 6*A*) to determine if the observed intervals, delimited by intergenic 6mA,



Expression Level	nº of genes near peaks	total nº of genes	% of genes near peaks
Low	3023	38134	7,9
Moderate	2106	15643	13,5
High	871	5313	16,4

Fig. 4. Expression levels of genes closest to 6mA DNA methylation. (*A*) Intragenic 6mA methylation is associated with low expression; 82% of the 512 identified methylated genes are poorly expressed (RPKM ≤ 1), 14% are moderately expressed (1 < RPKM ≤ 20), and 4% are highly expressed (RPKM > 20). (*B*) Percentage of genes with different expression levels near intergenic 6mA peaks. Of the 6,000 genes closest to intergenic methylation, 3,023 have low expression, 2,106 have moderate expression, and 871 genes have high expression, corresponding to 7.9, 13.5, and 16.4% of all *T. vaginalis* poorly, moderately, and highly expressed genes, respectively. (*C*) Bar plot comparing the median and interquartile range of RPKM expression of genes near intergenic 6mA and genes in nonmethylated regions shows that genes near intergenic methylation possess a significantly higher expression than genes in nonmethylated regions. The *P* value was calculated by a two-sided Mann–Whitney test (***$P < 0.0001$). (*D*) Distribution of methylation frequency shows that intergenic 6mA is primarily located between −1,500 and +1,000 bp of the nearest not expressed (RPKM ≤ 1) or expressed (RPKM > 1) genes. Methylation frequency of intergenic peaks was calculated and plotted against the distance to the nearest gene, with zero corresponding to peaks that fall within genes.

Downloaded at Akademiska Sjukhuset on May 28, 2020

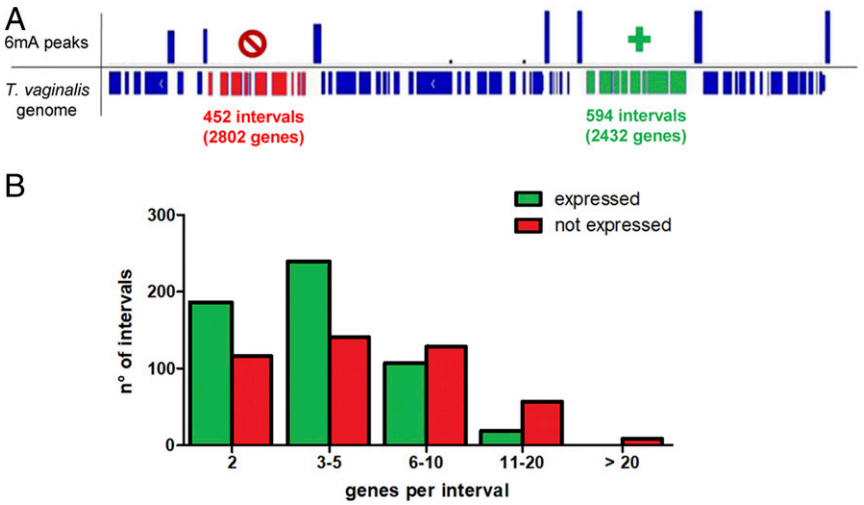


Fig. 5. Gene blocks flanked by intergenic 6mA. (*A*) Representation of intervals delimited by 6mA peaks composed of genes with the similar level of expression; 1,046 intervals were found, 452 containing not expressed genes (red) and 592 containing expressed genes (green). (*B*) Quantification of intervals flanked by 6mA containing different numbers of expressed (green) and nonexpressed (red) genes.

could be forming this type of chromatin architecture. In order to select intervals to analyze by 3C assay, we filtered out all of the intervals with fewer than five genes, obtaining a list of 187 active and 233 repressive intervals. Surprisingly, we found several intervals containing BspA family members: two of the repressive intervals contained four BspA genes each, while six active intervals contained one BspA gene per interval (Dataset S5). In bacteria, the BspA family of cell surface proteins has been shown to be involved in the colonization of the oral mucosa and triggering of host immune response (46). In *T. vaginalis*, recent work suggests that BspA proteins might have a similar function (46). Therefore, we decided to focus primarily on those active and repressive intervals that contained BspA genes. Unfortunately, the repetitive nature of *T. vaginalis* genome presented a challenge for the design of specific 3C PCR primers, limiting the number of BspA-containing intervals to be considered for the

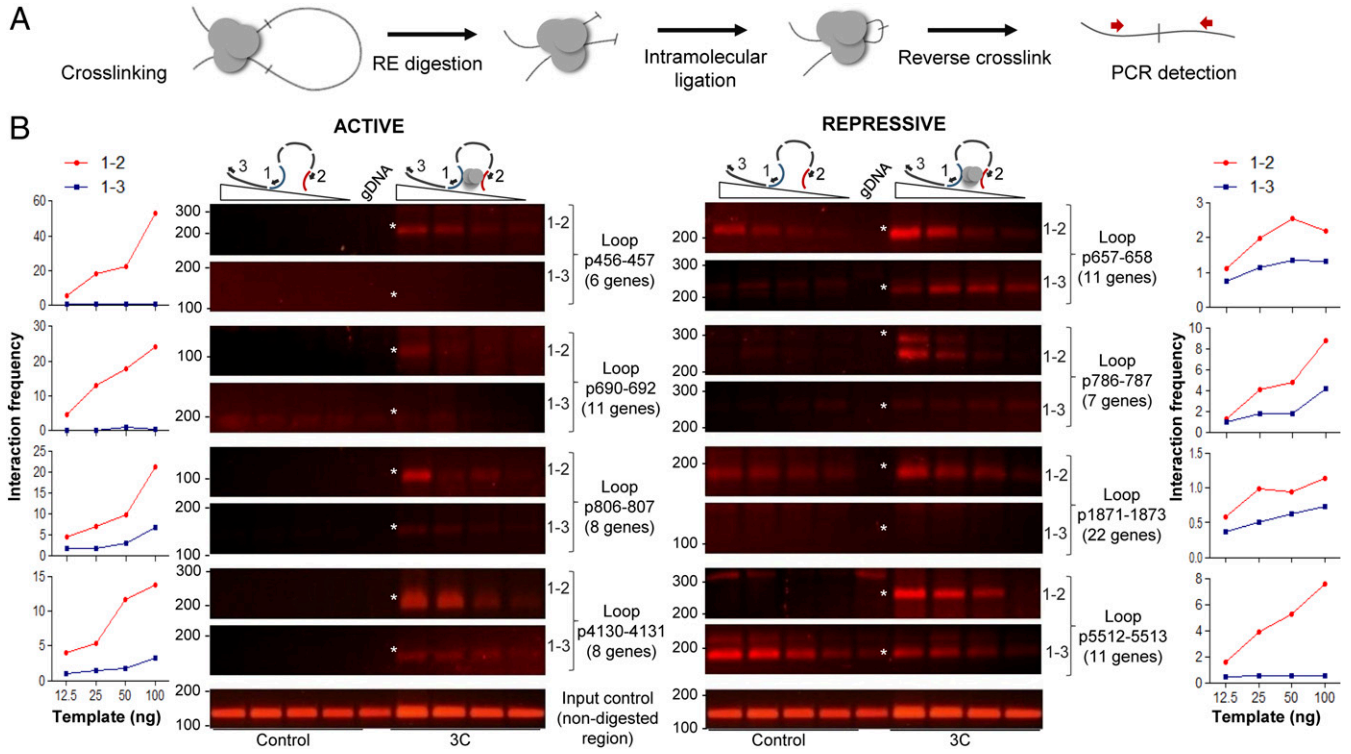


Fig. 6. The 3C assay evidence loop formation in *T. vaginalis* genome. (*A*) Schematic representation of the 3C assay. RE, restriction enzyme. (*B*) The 3C assay shows evidence of loop formation in *T. vaginalis* genome. Lanes 1 to 4: serial dilutions of uncross-linked control template. Lanes 5: undigested gDNA. Lanes 6 to 9: serial dilutions of 3C template. Active: loops containing expressed genes. Repressive: loops containing not expressed genes. Nondigested control: amplification of a region without a restriction enzyme site in between the utilized primers. Line graphs show the interaction frequency of the methylated regions (primers pairs 1 to 2; red lines) and control region (primers pairs 1 to 3; blue lines) of each interval tested. Asterisks mark the bands with the expected fragment size.

Downloaded at Akademiska Sjukhuset on May 28, 2020

experiment to repressive intervals p5512 to p5513 and p2296 to p2297 and active intervals p4130 to p4131, p560 to p561, and p271 to p272. Consequently, we decided to include intervals lacking BspA genes, resulting in four additional active and repressive intervals to be analyzed by 3C assay (*SI Appendix*, Table S4). Among this set of regions, we were unable to amplify only the BspA-containing intervals p560 to p561 and p271 to p272 and intervals p1065 to p1066 and p228 to p230. On the other hand, we successfully amplified active intervals p456 to p457, p690 to p692, p806 to p807, and p4130 to p4131, as well as repressive intervals p657 to p658, p786 to p787, p1871 to p1873, and p5512 to p5513. PCR amplification of 3C product showed bands of the expected sizes that were undetected on undigested control gDNA (Fig. 6*B*, primer pair 1 to 2). Importantly, sequencing of each 3C band verified the identity of the expected cross-ligation products (*SI Appendix*, Fig. S4). Additionally, we amplified BspA-containing interval p2296 to p2297 (*SI Appendix*, Fig. S4*I*), obtaining a band of the expected size. Although we were not able to achieve reproducible results with this primer pair, band sequencing confirmed that the identity of the fragment was the expected cross-ligation product (*SI Appendix*, Fig. S4*J*), suggesting that these distant methylated regions delimiting intervals (*SI Appendix*, Table S4) were in close spatial proximity to each other in the cell. To further validate these results, we designed PCR primers to evaluate the interaction between one extreme of each interval and an adjacent nonmethylated region. To account for distance bias during ligation, primers were designed so that the length of each adjacent control region was similar to its corresponding interval (*SI Appendix*, Table S5). When using the adjacent control primer pairs (1 to 3), PCR amplification of 3C product of each interval showed weaker or absent bands of the expected size. These bands were completely absent in all the cases when undigested control gDNA was used as template (Fig. 6*B*). After considering the PCR amplification efficiency of each primer set, quantitative comparison of signal intensities of PCR products demonstrated that the interaction frequency of the interval fragments was consistently higher than the interaction frequency of their respective adjacent control fragments (line graphs in Fig. 6*B*). Furthermore, the intensity of PCR amplification of 3C product was also compared with the one obtained using an alternative set of primers designed between one extreme of each interval and a flanking region located at a greater distance (1 to 4) (*SI Appendix*, Fig. S5). Taken together, our results suggest the presence of chromatin loops at these regions.

To examine whether 6mA might be responsible for regulating the formation of these structures, we performed a 3C assay of parasites treated with ascorbic acid (AA), which has been previously shown to favor DNA demethylation in mammalian cells (47). Dot blot assay of parasites treated with AA confirmed a diminution of 6mA mark of parasites treated with 100 mM AA compared with nontreated parasites (Fig. 7*A*). Unfortunately, due to the repetitive nature of the regions where 6mA is located, we were unable to design specific primers to test for the loss of methylation at particular loop regions through MeDIP-qPCR. However, loop formation, as determined by 3C analysis, was diminished in parasites treated with AA when compared with the nontreated control, while no such change was observed for the adjacent control regions (Fig. 7*B*). In order to analyze if loop formation could be affecting the expression of genes located within the loops, we selected two active (p456 to p457 and p690 to p692) and two repressive (p786 to p787 and p5512 to p5513) intervals whose formation was affected by AA and performed RT-qPCR of genes contained in these loops. Interestingly, the expression of all genes located within active or repressive intervals was up-regulated upon treatment with AA. Our observations suggest that 6mA could indeed be a mark associated with chromatin looping and may regulate the transcriptional activity of gene-associated blocks in *T. vaginalis*.

Discussion

Epigenetics studies and their relationship with gene expression in *T. vaginalis* are still in their infancy. To date, only the contribution of histone modifications to the regulation of gene expression in the parasite has been described (18, 19). In the present study, we describe the presence of 6mA modification in the genome of this unicellular parasite. Here, we present a whole-genome analysis of 6mA distribution that provides insight into the possible contributions of this modification to 3D genome architecture and gene expression.

In recent years, the role of 6mA has been expanded from a primarily prokaryotic DNA modification to an important epigenetic mark present in several multicellular and unicellular eukaryotic organisms (34). In this study, we use antibody-dependent and -independent methods to demonstrate that 6mA is an abundant DNA modification in *T. vaginalis* genome; 6mA levels in the parasite are comparable with those found in early-diverging fungi (33), making *T. vaginalis* one of the eukaryotic organisms with the highest levels of 6mA reported to date. It is interesting to note that just like early-diverging fungi (33), 5mC in *T. vaginalis* was nearly undetectable. Unfortunately, detection methods that rely on the use of antibodies can give false positives when 5mC levels are low (48), which could explain the low-quality reads obtained from the 5mC MeDIP-seq experiment. These results support our conclusion that 6mA, and not 5mC, is the main DNA modification in *T. vaginalis* genomic DNA.

To be considered an epigenetic mark, 6mA needs to be maintained after DNA replication. In *Tetrahymena*, maintenance methylation occurs quickly after DNA replication (49). Similarly, 6mA in *Chlamydomonas* is installed shortly after DNA replication and is stably maintained during cell proliferation (29). Although the precise moment of 6mA deposition in the newly synthetized DNA strand in the parasite remains elusive, our observations indicate that maintenance of 6mA methylation occurs after cell division and supports the presence of active 6mA methyltransferase(s) in the parasite. It should be noted that, although phylogenetic analysis revealed the presence of a group of putative parasite-specific 6mA DNA methyltransferases encoded in the genome database (TrichDB) (50), functional analyses have not been performed to validate these candidates yet. On the other hand, while DNA 6mA demethylases have been identified in several organisms (27, 28, 30, 36), no classical TET or AlkB homologs were identified in *T. vaginalis* genome (50). This could be indicative of the presence of nonconserved enzyme, a lack of active 6mA demethylation, or the use of an alternative pathway, such as the base excision–repair DNA demethylation pathway found in plants (51). It will be of interest to take a closer look at the DNA glycosylases found in the *T. vaginalis* genome database and evaluate if they could be involved in active 6mA DNA demethylation. Identifying the proteins responsible for the regulation of 6mA levels in the parasite will further our understanding of the mechanisms and functions of this mark.

Genome-wide distribution of 6mA varies between species (27–30, 52). An evolutionary conservation for 6mA in unicellular organisms has been proposed due to the similarities in 6mA distribution pattern and function between green algae and *Tetrahymena* (49). However, using an MeDIP-seq approach we found that the distribution pattern in *T. vaginalis* resembles more that of 6mA in *Drosophila* and mouse than other unicellular eukaryotes. Methylation peaks in *T. vaginalis* are enriched in intergenic regions and seem to have a preference for certain types of DNA TE superfamilies, reminiscent of the preference for young LINE-1 transposons described in mouse (30). In

Downloaded at Akademiska Sjukhuset on May 28, 2020

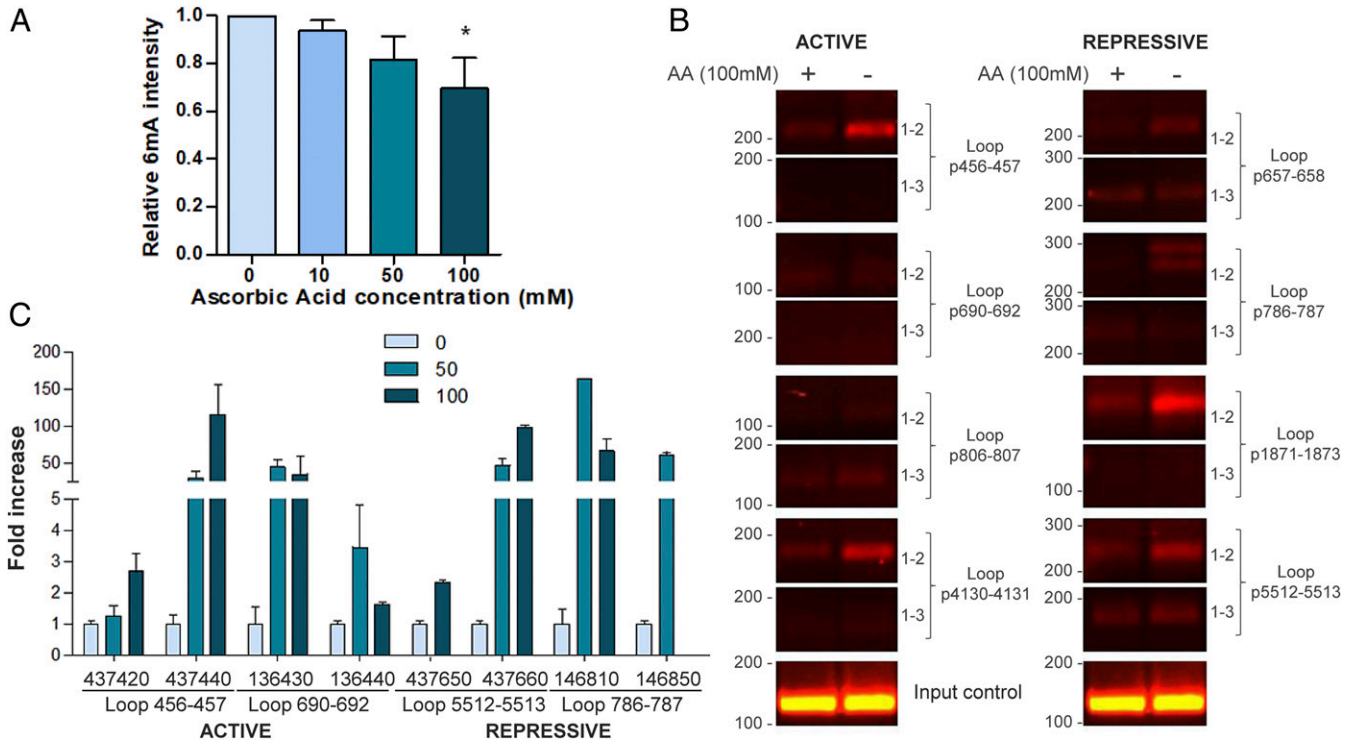


Fig. 7. The 3C assay on parasites treated with AA. (*A*) Quantification of 6mA levels of parasites treated with different concentrations of AA shows diminution of 6mA abundance in parasites treated with 100 mM AA compared with untreated parasites. Each column represents the mean and SEM of two independent experiments. (*$P < 0.05$). (*B*) The 3C assay of parasites exposed to 100 mM AA (+) shows diminished loop formation compared with nontreated parasites (–); 25 ng of 3C template was used for PCR amplification using primers for each active and repressive loop and adjacent control regions. 1 to 2: PCR using interval primers 1 to 2. 1 to 3: PCR using interval primer 1 and adjacent control region primer 3. Nondigested control: amplification of a region without a restriction enzyme site in between the utilized primers. (*C*) Expression analysis of genes located within loops upon treatment with AA. Parasites were exposed to increasing concentrations of AA, and expression was analyzed by qRT-PCR. Expression levels of eight transcripts found within four different loops, two active and two repressive, were examined. Expression of all tested genes changed after AA treatment. Data are expressed as fold increase compared with 0 mM AA. Each column represents the mean and SD of two independent experiments per group.

Drosophila, some 6mA peaks are enriched in the gene bodies of transposons (27). Similarly, we found that the majority of the methylated genes in the parasite belong to TEs, with 6mA peaks preferentially located on the CDS or TTS of these TE genes (*SI Appendix*, Table S6). This predilection for certain types of TEs and the specific localization of peaks in transposon genes raise the question of whether 6mA could act as a regulator of specific TEs in *T. vaginalis*. A characterization of some members of the T. *vaginalis* Tc1/mariner family by Bradic et al. (53) demonstrated that this family of TEs can influence the expression or neighboring genes. Consequently, 6mA on TEs could also have an indirect effect on the expression of nearby genes.

Like its genomic distribution, 6mA association with transcription is dependent of the organism in question. It has been demonstrated that 6mA acts as a mark of active transcription in *Arabidopsis* (35), *Drosophila* (27), green algae (29), and fungi (33), while it is a mark of transcriptional repression in mouse (30) and rice (36). In *T. vaginalis*, 6mA seems to be associated with silencing when found on genes. As it has been demonstrated that gene expression varies among different strains (37, 54), it would be of interest to analyze the association between gene expression and genome 6mA distribution in other *T. vaginalis* strains, particularly those that vary greatly in adherence and cytotoxicity to host cells. Such analysis would help define the role of 6mA in parasite pathogenesis. Importantly, we cannot rule out the possibility of 6mA acting in conjunction with other components of the epigenetic code to regulate gene expression in a context-dependent manner. It has been demonstrated that 6mA lowers the melting temperature and can cause severe unwinding and bending of DNA helix, increasing DNA curvature to variable degrees depending on sequence context (55). This characteristic could help in the recruitment of other members of the epigenetic machinery to regulate gene expression in a concert manner. In future studies, it will be interesting to determine whether 6mA binding proteins exist in *T. vaginalis* that could help regulate the response of chromatin to this modification. An integrative analysis is necessary to fully understand how epigenetics shapes the transcriptional landscape of *T. vaginalis*.

Chromatin looping is a common intrachromosomal protein-mediated interaction capable of influencing gene expression. The function of these interactions is determined by the nature of the sequences brought together and the proteins that interact with them (56). For example, promoter–enhancer loops are strongly associated with gene activation (57), and it has been proposed that the presence of chromatin loops may contribute to the coregulation of specific blocks of genes by increasing the local concentration of the transcription machinery, resulting in the formation of structures similar to transcription factories (58). Alternatively, chromatin looping has also been implicated in transcriptional repression in *Drosophila* (59) and mammals (60). Interestingly, our bioinformatics analysis revealed the presence of active and repressive intervals of varying length flanked by 6mA. A 3C assay confirmed that the methylated extremes in both transcriptionally active and repressive blocks of genes are in close spatial proximity and that disruption of these structures leads to altered gene expression. Notably, some contained one or more members of the BspA-like gene family of surface proteins, which might be involved in parasite pathogenesis (61). In

Downloaded at Akademiska Sjukhuset on May 28, 2020

Plasmodium falciparum, chromosomes containing virulence genes of the *var* family form loops that permit the perinuclear colocalization of the var genes (62), and a recent analysis demonstrated that this clustering of virulent genes is a common feature among other *Plasmodium* species (63). Since clustered organization of virulence genes allows coordination of gene expression in *Plasmodium* (63), future studies are warranted to evaluate if the same is true for *T. vaginalis*. Additionally, it would be interesting to evaluate if DNA methylation and chromatin looping play a role in *T. vaginalis* transcriptional regulation induced by contact with host cells (8) as well as the changes in gene expression observed when parasites are exposed to different stimuli (9).

The association of 6mA with chromatin looping is noteworthy because it points to a function for this epigenetic mark in eukaryotes with scarce 5mC levels. Given that this work demonstrated that chromatin structure may regulate, positively or negatively, specific blocks of genes, these findings open avenues for understanding the role of 6mA writers, erasers, and readers that may play an important role in the transcriptional regulation of *T. vaginalis*.

Materials and Methods

Parasites, Cell Cultures, and Media. *T. vaginalis* strains B7268 (64) and G3 (ATCC PRA-98) were cultured at 37 °C in trypticase-yeast extract-maltose (TYM) medium supplemented with 10% horse serum, penicillin, and streptomycin (Invitrogen) (65). To obtain *Mycoplasma hominis*-free strains, parasite cultures were treated for 2 wk with the recommended working concentration of BM-Cyclin (Roche). *M. hominis* clearance was confirmed by PCR as described previously (66).

6mA and 5mC Immunolocalization Experiments. Parasites in the absence of host cells were incubated at 37 °C on glass coverslips as previously described (54) for 4 h and then incubated in cold methanol for 10 min for fixation and permeabilization of both cellular and nuclear membranes. Cells were washed with phosphate-buffered saline (PBS) and incubated at 37 °C with a 1:100 dilution of RNase A (QIAGEN) in PBS for 3 h. Non-RNase A-treated control parasites were incubated for 3 h in PBS. Parasites were washed once again, blocked with 5% fetal bovine serum (FBS) in PBS for 30 min, and incubated overnight with a 1:100 dilution of anti-6mA or anti-5mC (Abcam) primary antibody in PBS containing 2% FBS. Parasites were washed and then incubated with a 1:5,000 dilution of Alexa Fluor conjugated secondary antibody (Molecular Probes). Nuclei was stained with DAPI (2 mg/mL), and coverslips were mounted onto microscope slips using Fluoromount Aqueous Mounting Medium (Sigma-Aldrich). All observations were performed on a Zeiss Axio Observer 7 epifluorescence microscope. Adobe Photoshop (Adobe Systems) was used for image processing.

Genomic DNA Extraction. *T. vaginalis* cultures were lysed using a solution of 8 M urea, 2% sarkosyl, 0.15 M NaCl, 0.001 M ethylenediaminetetraacetic acid (EDTA), and 0.1 M Tris HCl, pH 7.5 (67). RNase A was added to a final concentration of 0.1 mg/mL, and lysates were incubated at 37 °C for 2 to 3 h. Total DNA was extracted using phenol/chloroform/isoamyl alcohol (Sigma-Aldrich), precipitated with isopropanol, and resuspended in nuclease-free water.

Dot Blot Assays. RNA free genomic DNA samples were diluted to the desired concentration, and 1 μL of each sample was spotted on Amersham Hybond-N+ membranes (GE Healthcare). For anti-5mC blots, DNA was denatured for 10 min at 95 °C. Membranes were allowed to air dry, and then, DNA was cross-linked for 3 min twice in a Hoefer UVC 500 Crosslinker (Amersham Biosciences). Membranes were stained with 0.02% methylene blue in 0.3 M NaOAc, pH 5.2, solution for DNA loading control and then washed with distilled water. The membranes were blocked for 1 h in 5% milk mixture of tris-buffered saline and Tween 20 (TBS-T) and incubated overnight at 4 °C with 1:1,000 anti-6mA primary antibody dilution or 1:500 anti-5mC primary antibody dilution in 5% milk TBS-T. Membranes were washed three times for 10 min with TBS-T and incubated with a 1:10,000 dilution of alkaline phosphatase (AP)-conjugated secondary antibody in 5% milk TBS-T for 1 h at room temperature. After three washes of 10 min with TBS-T, blots were developed using 4-Nitro blue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate, 4-toluidine salt (BCIP) (Roche).

Restriction Enzyme Digestion Assay. Restriction enzyme digestion was performed by treating 1 μg of genomic DNA with 1 μL of 10 U/μL MboI or DpnI restriction enzyme (Thermo Scientific) at 25 °C overnight. Sample with no restriction enzyme was used as control. Enzyme was heat inactivated, and the product mixture was separated by agarose gel electrophoresis. Gels were visualized with an ultraviolet transilluminator.

Quantification of 6mA by ultra high-performance liquid chromatography coupled with triple-quadrupole mass spectrometry (UHPLC-QQQ-MS/MS) Analysis. 6mA abundance was quantified as previously described (28, 52); 1.5 μg of genomic DNA in 30 mL double-distilled water (ddH_2O) was digested to free nucleosides using 5 U of DNA Degradase Plus (Zymo Research) in 25-μL reactions incubated for 2 h at 37 °C. The diluted solution was filtered through a 0.22-mm filter, and 10 mL solution was injected into liquid chromatography–mass spectrometry (LC-MS/MS). The nucleosides were separated by reverse-phase UHPLC on a C18 column (Agilent), with online mass spectrometry detection using Agilent 6460 QQQ triple-quadrupole tandem mass spectrometer set to multiple reaction monitoring in positive electrospray ionization mode. Nucleosides were quantified using the nucleoside precursor ion to base ion mass transitions of 266.1 to 150.0 for 6mA and 252.1 to 136.0 for adenosine (A). Quantification of the ratio 6mA/A was performed using the calibration curves obtained from nucleoside standards running at the same time. Measurement of 6mA and deoxyadenosine (dA) levels in the DNA degradase enzyme mix alone served as an additional control for background levels of nucleic acid contributed by the digestion enzymes. The low levels of 6mA and dA in the DNA degradase enzyme mix were subtracted from the final measurements. Because *T. vaginalis* has relatively high levels of 6mA, the low levels of methylated and unmethylated DNA in the reaction mix were a negligible fraction of the total digested DNA samples.

Cell Cycle Synchronization. Parasites (5×10^5 cells per milliliter) were incubated overnight at 37 °C in the presence of 100 μM HU. The following day, cells were resuspended in fresh medium without HU, divided into 8-mL aliquots, and left to grow at 37 °C. An aliquot was taken every 2 h and was used for flow cytometry analysis to determine cell cycle stage. The remaining sample was used for genomic DNA extraction to determine abundance of 6mA.

Flow Cytometry. Parasites were treated as described previously (68). Parasites were fixed in 5 mL of ice-cold 100% EtOH and incubated at 4 °C overnight. Each sample was washed once in 1 mL PBS containing 2% vol/vol bovine serum (BS) and resuspended in 1 mL PBS containing 180 μg/mL RNase A and 2% vol/vol BS. After a 30-min incubation at 37 °C, samples were stained with 25 μg/mL propidium iodide (PI) solution and incubated an additional 30 min at 37 °C. Samples were analyzed using a fluorescence-activated cell sorter (BD FACSCalibur; BD Biosciences) with appropriate filter sets. Data were analyzed using FlowJo 7.6 software.

MeDIP-seq. Twenty micrograms of purified genomic DNA from *T. vaginalis* strain B7268 (64) was diluted in 400 μL 1× immunoprecipitation (IP) buffer (10 mM Na-Phosphate, pH 7, 140 mM NaCl, 0.05% Triton X-100) and fragmented to 100 to 500 bp using a water bath sonicator (amp: 50%, cycle: 30 s on and 30 s off, time: 20 min). Sample was divided into aliquots containing 6 to 7 μg of sonicated DNA and heat denatured; 1 μg was saved and stored at −20 °C to use as input control, and the rest was immunoprecipitated overnight at 4 °C using 5 μg of anti-6mA antibody (Abcam) bound to protein A magnetic beads (Invitrogen). After three washes with 1× IP buffer, beads containing bound methylated DNA were resuspended in 250 μL digestion buffer (50 mM Tris, pH 8, 10 mM EDTA, 0.5% sodium dodecyl sulfate [SDS]) with 100 μg Proteinase K (Roche) and incubated at 65 °C for 4 h. Eluted DNA and input control samples were purified using phenol/chloroform/isoamyl alcohol extraction followed by EtOH precipitation. MeDIP samples were sent to Applied Biological Materials Inc. for library preparation and paired-end sequencing. Two independent high-throughput sequencing experiments, each containing pooled DNA from three independent immunoprecipitation, were performed.

Bioinformatics Analysis of MeDIP-Seq Data. After the adaptors were trimmed, alignment of high-quality fastq reads to the reference genome (GCF 000002825.2 ASM282v1) was performed using Bowtie2 (69) with –qc-filter parameter and all other parameters at default settings. The BAM files were used as input for MACS2 (70), which was run with the –g parameter set at 1.76e8 and all other parameters at default settings. Input reads were used as control sample for peak calling. Peaks were then filtered by *q* value (<0.01)

Downloaded at Akademiska Sjukhuset on May 28, 2020

and fold enrichment (>2). Only peaks with more than 75% overlap between replicates were considered for further analysis. HOMER2 (71) was used for peak annotation. For repetitive element annotation, RepeatMasker v4.0 (72) was used with a custom library of *T. vaginalis*-specific TE taken from Repbase (73). GO enrichment analysis was performed using the Results Analysis feature at TrichDB with default settings.

Chromatin Conformation Capture Assay. Cross-linking of DNA was achieved by incubation of 50 mL (~1×10^6 cells per milliliter) of *T. vaginalis* strain B7268 (64) culture with 1% formaldehyde for 15 min at room temperature. Cross-linking was stopped by the addition of 125 mM glycine and incubated 5 min at room temperature, then on ice for 15 min. Cells were harvested, washed in 10 mL PBS/PI, and resuspended in 1.4 mL of cold lysis buffer (10 mM Tris·HCl, pH 8, 10 mM NaCl, 0.2% Nonidet P-40, 1× PI) followed by homogenization with a dounce homogenizer pestle A. After centrifugation, the pellet was washed once with 1.25× Buffer H (Promega) and resuspended in 500 μL of 1.25× Buffer H (Promega). A total of 7.5 μL of SDS 20% was added and incubated for 40 min at 65 °C and then for 20 min at 37 °C. SDS was quenched with 50 μL of 20% Triton X-100 at 37 °C for 1 h; 400 U of EcoRI (Promega) was added and incubated overnight at 37 °C followed by enzyme heat inactivation with 1% SDS at 65 °C for 20 min. Approximately 30 μL of digested sample was set aside to test restriction enzyme efficiency by electrophoresis and PCR. Digestion product was cooled and transferred to a prechilled 15-mL tube containing 4 mL 1.1× Ligase Buffer (Promega) and 1% Triton X-100. After a 1-h incubation at 37 °C, the sample was cooled down, mixed with 30 Weiss Units of T4 DNA Ligase (Promega), and incubated for 4 h in a water bath at 16 °C. Cross-linking was reversed overnight at 65 °C in the presence of 300 μg of Proteinase K (Roche). DNA was purified using phenol/choloroform/isoamyl alcohol extraction followed by EtOH precipitation. Twofold serial dilutions starting with 100 ng/μL were used for PCR amplification. The sequences of primers used in 3C assays are shown in *SI Appendix*, Table S7. Primer efficiency was determined using a mix of equimolar amounts of 3C DNA fragments, and interaction frequencies between cross-linked and noncross-linked product were calculated as previously described (74); 3C PCR fragments were purified and sequenced to verify the identity of the cross-ligation product.

qPCR. Total RNA was extracted from ~4×10^6 *T. vaginalis* using the RNAspin Mini RNA Isolation Kit (GE Healthcare Life Sciences) following the manufacturer's instructions. Total RNA was treated with amplification-grade DNase I (Invitrogen) and reverse transcribed using SuperScript II reverse transcriptase and random primers (Invitrogen). Real-time PCRs were performed using Brilliant SYBR Green qPCR Master Mix (Roche), a 150 to 450 nM concentration of each primer, and 200 to 500 ng of complementary DNA (cDNA) in a 15-μL reaction volume using a Stratagene Mx3005PTM system (Agilent Technologies). Using data from the exponential phase of the qPCR, threshold cycle baselines were set according to manufacturer protocols. Data from different samples were interpolated from standard curves established for each primer set and then normalized against housekeeping gene alpha-tubulin. Every experimental and standard curve sample was tested in triplicate in two independent experiments. The primers used are provided in *SI Appendix*, Table S8.

Accession Numbers. All sequencing data that support the findings of this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) and are accessible through SRA accession number PRJNA526331.

ACKNOWLEDGMENTS. We thank Dr. Pablo Manavella and our colleagues in the laboratory for helpful discussions. This research was supported with Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) Grants Banco Interamericano de Desarrollo (BID) Proyectos de Investigación Científica y Tecnológica (PICT) 2015-2118 (to N.d.M.) and PICT-2018-01892 (to N.d.M.). Work in the laboratory of E.L.G. was supported by NIH Grants DP2AG055947 (to E.L.G.) and R21HG010066 (to E.L.G.). P.H.S.-M. and N.d.M. are researchers from the National Council of Research (CONICET) and National University of San Martin (UNSAM). A.L. is a PhD fellow from CONICET. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

1. P. Kissinger, *Trichomonas vaginalis*: A review of epidemiologic, clinical and treatment issues. *BMC Infect. Dis.* **15**, 307 (2015).
2. B. J. Silver *et al.*; STRIVE investigators, Incidence of curable sexually transmissible infections among adolescents and young adults in remote Australian Aboriginal communities: Analysis of longitudinal clinical service data. *Sex. Transm. Infect.* **91**, 135–141 (2015).
3. H. Swygard, A. C. Seña, M. M. Hobbs, M. S. Cohen, Trichomoniasis: Clinical manifestations, diagnosis and management. *Sex. Transm. Infect.* **80**, 91–95 (2004).
4. S. M. Graham *et al.*, Initiation of antiretroviral therapy leads to a rapid decline in cervical and vaginal HIV-1 shedding. *AIDS* **21**, 501–507 (2007).
5. S. Gander, V. Scholten, I. Osswald, M. Sutton, R. van Wylick, Cervical dysplasia and associated risk factors in a juvenile detainee population. *J. Pediatr. Adolesc. Gynecol.* **22**, 351–355 (2009).
6. O. Twu *et al.*, *Trichomonas vaginalis* homolog of macrophage migration inhibitory factor induces prostate cell growth, invasiveness, and inflammatory responses. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8179–8184 (2014).
7. J. M. Carlton *et al.*, Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* **315**, 207–212 (2007).
8. S. B. Gould *et al.*, Deep sequencing of *Trichomonas vaginalis* during the early infection of vaginal epithelial cells and amoeboid transition. *Int. J. Parasitol.* **43**, 707–719 (2013).
9. K. Y. Huang *et al.*, Adaptive responses to glucose restriction enhance cell survival, antioxidant capability, and autophagy of the protozoan parasite *Trichomonas vaginalis*. *Biochim. Biophys. Acta* **1840**, 53–64 (2014).
10. L. Horváthová *et al.*, Transcriptomic identification of iron-regulated and iron-independent gene copies within the heavily duplicated *Trichomonas vaginalis* genome. *Genome Biol. Evol.* **4**, 1017–1029 (2012).
11. A. Smith, P. Johnson, Gene expression in the unicellular eukaryote *Trichomonas vaginalis*. *Res. Microbiol.* **162**, 646–654 (2011).
12. D. R. Liston, A. O. Lau, D. Ortiz, S. T. Smale, P. J. Johnson, Initiator recognition in a primitive eukaryote: IBP39, an initiator-binding protein from *Trichomonas vaginalis*. *Mol. Cell. Biol.* **21**, 7872–7882 (2001).
13. M. A. Schumacher, A. O. Lau, P. J. Johnson, Structural basis of core promoter recognition in a primitive eukaryote. *Cell* **115**, 413–424 (2003).
14. A. J. Smith *et al.*, Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis*. *Mol. Cell. Biol.* **31**, 1444–1458 (2011).
15. S. J. Ong, S. C. Huang, H. W. Liu, J. H. Tai, Involvement of multiple DNA elements in iron-inducible transcription of the ap65-1 gene in the protozoan parasite *Trichomonas vaginalis*. *Mol. Microbiol.* **52**, 1721–1730 (2004).
16. S. Golbabapour, M. A. Abdulla, M. Hajrezaei, A concise review on epigenetic regulation: Insight into molecular mechanisms. *Int. J. Mol. Sci.* **12**, 8661–8694 (2011).
17. M. M. Croken, S. C. Nardelli, K. Kim, Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives. *Trends Parasitol.* **28**, 202–213 (2012).
18. M. J. Song *et al.*, Epigenome mapping highlights chromatin-mediated gene regulation in the protozoan parasite *Trichomonas vaginalis*. *Sci. Rep.* **7**, 45365 (2017).
19. T. Pachano *et al.*, Epigenetics regulates transcription and pathogenesis in the parasite *Trichomonas vaginalis*. *Cell. Microbiol.* **19**, e12716 (2017).
20. K. Chen, B. S. Zhao, C. He, Nucleic acid modifications in regulation of gene expression. *Cell Chem. Biol.* **23**, 74–85 (2016).
21. P. A. Jones, D. Takai, The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068–1070 (2001).
22. H. Wei *et al.*, Characterization of cytosine methylation and the DNA methyltransferases of *Toxoplasma gondii*. *Int. J. Biol. Sci.* **13**, 458–470 (2017).
23. N. Ponts *et al.*, Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum*. *Cell Host Microbe* **14**, 696–706 (2013).
24. O. Fisher, R. Siman-Tov, S. Ankri, Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res.* **32**, 287–297 (2004).
25. G. Z. Luo, M. A. Blanco, E. L. Greer, C. He, Y. Shi, DNA N(6)-methyladenine: A new epigenetic mark in eukaryotes? *Nat. Rev. Mol. Cell Biol.* **16**, 705–710 (2015).
26. B. F. Vanyushin, S. G. Tkacheva, A. N. Belozersky, Rare bases in animal DNA. *Nature* **225**, 948–949 (1970).
27. G. Zhang *et al.*, N6-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906 (2015).
28. E. L. Greer *et al.*, DNA methylation on N6-adenine in *C. elegans*. *Cell* **161**, 868–878 (2015).
29. Y. Fu *et al.*, N6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879–892 (2015).
30. T. P. Wu *et al.*, DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
31. J. Liu *et al.*, Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **7**, 13052 (2016).
32. M. J. Koziol *et al.*, Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2016).
33. S. J. Mondo *et al.*, Widespread adenine N6-methylation of active genes in fungi. *Nat. Genet.* **49**, 964–968 (2017).
34. K. D. Meyer, S. R. Jaffrey, Expanding the diversity of DNA base modifications with N^6-methyldeoxyadenosine. *Genome Biol.* **17**, 5 (2016).
35. Z. Liang *et al.*, DNA N(6)-adenine methylation in *Arabidopsis thaliana*. *Dev. Cell* **45**, 406–416.e3 (2018).
36. C. Zhou *et al.*, Identification and analysis of adenine N^6-methylation sites in the rice genome. *Nat. Plants* **4**, 554–563 (2018).
37. M. Bradic *et al.*, Genetic indicators of drug resistance in the highly repetitive genome of *Trichomonas vaginalis*. *Genome Biol. Evol.* **9**, 1658–1672 (2017).
38. M. Piknova, P. Pristas, P. Javorsky, A. Kasperowic, T. Michalowski, GATC-specific restriction and modification systems in treponemes. *Lett. Appl. Microbiol.* **38**, 311–314 (2004).

Downloaded at Akademiska Sjukhuset on May 28, 2020

39. G. Lustig, C. M. Ryan, W. E. Secor, P. J. Johnson, *Trichomonas vaginalis* contact-dependent cytolysis of epithelial cells. *Infect. Immun.* **81**, 1411–1419 (2013).
40. R. Fichorova, J. Fraga, P. Rappelli, P. L. Fiori, *Trichomonas vaginalis* infection in symbiosis with Trichomonasvirus and Mycoplasma. *Res. Microbiol.* **168**, 882–891 (2017).
41. M. Weber *et al*., Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853–862 (2005).
42. D. Leitsch, B. D. Janssen, D. Kolarich, P. J. Johnson, M. Duchêne, *Trichomonas vaginalis* flavin reductase 1 and its role in metronidazole resistance. *Mol. Microbiol.* **91**, 198–208 (2014).
43. A. Denker, W. de Laat, The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
44. S. S. Rao *et al*., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
45. O. Oyinlade *et al*., Analysis of KLF4 regulated genes in cancer cells reveals a role of DNA methylation in promoter-enhancer interactions. *Epigenetics* **13**, 751–768 (2018).
46. C. J. Noël *et al*., *Trichomonas vaginalis* vast BspA-like gene family: Evidence for functional diversity from structural organisation and transcriptomics. *BMC Genom.* **11**, 99 (2010).
47. X. X. Yu *et al*., Ascorbic acid induces global epigenetic reprogramming to promote meiotic maturation and developmental competence of porcine oocytes. *Sci. Rep.* **8**, 6132 (2018).
48. A. Singh *et al*., Determination of the presence of 5-methylcytosine in *Parameciumtetraurelia*. *PLoS One* **13**, e0206667 (2018).
49. Y. Wang, X. Chen, Y. Sheng, Y. Liu, S. Gao, N6-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in Pol II-transcribed genes in *Tetrahymena*. *Nucleic Acids Res.* **45**, 11594–11606 (2017).
50. L. M. Iyer, D. Zhang, L. Aravind, Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays* **38**, 27–40 (2016).
51. J. K. Zhu, Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143–166 (2009).
52. Z. K. O'Brown *et al*., Sources of artifact in measurements of 6mA and 4mC abundance in eukaryotic genomic DNA. *BMC Genom.* **20**, 445 (2019).
53. M. Bradic, S. D. Warring, V. Low, J. M. Carlton, The Tc1/mariner transposable element family shapes genetic variation and gene expression in the protist *Trichomonas vaginalis*. *Mob. DNA* **5**, 12 (2014).
54. N. de Miguel *et al*., Proteome analysis of the surface of *Trichomonas vaginalis* reveals novel proteins and strain-dependent differential expression. *Mol. Cell. Proteomics* **9**, 1554–1566 (2010).
55. Z. K. O'Brown, E. L. Greer, N6-methyladenine: A conserved and dynamic DNA mark. *Adv. Exp. Med. Biol.* **945**, 213–246 (2016).
56. C. T. Ong, V. G. Corces, CTCF: An architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
57. S. S. P. Rao *et al*., Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
58. M. J. Rowley, V. G. Corces, Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
59. K. P. Eagen, E. L. Aiden, R. D. Kornberg, Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8764–8769 (2017).
60. Y. Guo *et al*., CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. *Genome Biol.* **19**, 160 (2018).
61. R. P. Hirt, *Trichomonas vaginalis* virulence factors: An integrative overview. *Sex. Transm. Infect.* **89**, 439–443 (2013).
62. F. Ay *et al*., Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* **24**, 974–988 (2014).
63. E. M. Bunnik *et al*., Changes in genome organization of parasite-specific gene families during the Plasmodium transmission stages. *Nat. Commun.* **9**, 1910 (2018).
64. J. A. Upcroft, P. Upcroft, Drug susceptibility testing of anaerobic protozoa. *Antimicrob. Agents Chemother.* **45**, 1810–1814 (2001).
65. C. G. Clark, L. S. Diamond, Methods for cultivation of luminal parasitic protists of clinical importance. *Clin. Microbiol. Rev.* **15**, 329–341 (2002).
66. A. Blanchard *et al*., Evaluation of intraspecies genetic variation within the 16S rRNA gene of *Mycoplasma hominis* and detection by polymerase chain reaction. *J. Clin. Microbiol.* **31**, 1358–1361 (1993).
67. F. Mercer *et al*., Leukocyte lysis and cytokine induction by the human sexually transmitted parasite *Trichomonas vaginalis*. *PLoS Negl. Trop. Dis.* **10**, e0004913 (2016).
68. L. S. Iriarte *et al*., TfVPS32 regulates cell division in the parasite *Tritrichomonas foetus*. *J. Eukaryot. Microbiol.* **65**, 28–37 (2018).
69. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
70. Y. Zhang *et al*., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
71. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
72. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0 (2013–2015). http://www.repeatmasker.org/faq.html. Accessed 14 June 2018.
73. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
74. E. Splinter, F. Grosveld, W. de Laat, 3C technology: Analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol.* **375**, 493–507 (2004).

Downloaded at Akademiska Sjukhuset on May 28, 2020