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

Ayelen Lizarraa, Zach Klapholz O’Brownb,c, Konstantinos Bouliasb,c, Lara Roachb,c, Eric Lieberman Greeb,c, Patricia J. Johnsond,1, Pablo H. Strobl-Mazzullae,1,1, Natalia de Miogea,1

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 ultra-high-performance liquid chromatography (UHPLC), we analyzed the abundance of DNA methylation in strains with different 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.

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 eubacteria 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.


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).

1To whom correspondence may be addressed. Email: johnsonp@uc.edu or ndmiguel@intech.gov.ar.

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

www.pnas.org/cgi/doi/10.1073/pnas.1917286117

PNAS Latest Articles | 1 of 11
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 5-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 ultra-high-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 method (MeDIP-seq) and enrichments of 6mA in the immunoprecipitated fraction (McDIP-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. 1A). Unexpectedly, we failed to detect 5mC in either strain, contrary to our chicken gDNA control (Fig. 1A). Likewise, although we were able to detect 5mC in the RNA, this mark was undetectable by immunofluorescence in the nucleus of the parasite (Fig. 1B). 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. 1C) using antibodies that specifically recognize methylated adenosines (SI Appendix, Fig. S1A). 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 adenosines in the GATC context (Fig. 1D). In order to determine whether parasite adherence could be correlated with 6mA abundance, we performed dot blot assays (Fig. 1E) and UHPLC followed by mass spectrometry (Fig. 1F) 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 adenosines (Fig. 1F). 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 adenosines, whereas extremely low levels of 5mC were detected (0.004 to 0.0008% of all cytosines) by UHPLC followed by mass spectrometry (Fig. 1F). 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. S1B). These data demonstrate the lack of T. vaginalis genome 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. 2A). 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. 2B).

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. 2C 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. S3A). The most prevalent motif, found in 4,462 peaks, was CAATACCC (P value = 1e-1624) (SI Appendix, Fig. S3B). Remarkably, peak visualization in a genome browser suggested a clear enrichment of peaks in intergenic regions (Fig. 2D). Indeed, peak annotation revealed that 94% of all MeDIP-seq peaks (9,746 6mA peaks) were intergenic (Fig. 2E). Importantly, a shuffled distribution...
of peaks across *T. vaginalis* genome would result in an even distribution between genes and intergenic regions (Fig. 3A), 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. 3B). 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. 3B). 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* (Dataset S2), indicating a preference of intergenic 6mA for certain types of repeats element DNA7 and TE superfamilies P and Kolobok (Fig. 3C). The remaining families 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. 3C). 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 located within TEs (Fig. 3D). How-

### 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...
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 (19, 42). 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 methylated 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 6mA peaks distance and expression level of associated genes. We observed that intergenic 6mA is primarily located between nonexpressed genes and +1,000 bp of the nearest genes regardless of their expression level (Fig. 4D). Intriguingly, nonexpressed genes are found significantly closer to their nearest upstream or downstream intergenic 6mA 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. 5A). These intervals varied in the number of genes they contained (from 2 up to 36 genes) (Fig. 5B) as well as in length, with some spanning over 20 kb (Dataset S4). As can be observed in Fig. 5B, intervals containing a greater number of genes tend to be a sign of repressed expression when found on genes.

![Image](https://www.pnas.org/cgi/doi/10.1073/pnas.1917286117) Lizarraga et al.

**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.
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. 6A) to determine if the observed intervals, delimited by intergenic 6mA,
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...
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. 6B, 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. S4I), 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. S4I), 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. 6B). 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. 6B). 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. 7A). 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. 7B). 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), 5mA in T. vaginalis was nearly undetectable. Unfortunately, detection methods that rely on the use of antibodies can give false positives when 5mA levels are low (48), which could explain the low-quality reads obtained from the 5mA MeDIP-seq experiment. These results support our conclusion that 6mA, and not 5mA, 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 synthesized 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 gene that is specific in T. vaginalis (6mA 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
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* Tcl/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 on 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
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 trypticate-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 previously described (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 (BD Biosciences) with appropriate filter sets. Data were analyzed using FlowJo 7.6 software.

Genomic DNA Extraction. T. vaginalis cultures were lysed using a solution of 8 M urea, 2% sarcosyl, 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 crosslinked 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/L MboI or DpnI restriction enzyme (Thermo Scientific) at 37 °C overnight. Samples 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). 6mA abundance was quantified as previously described (28, 52). A 1.5 μg of genomic DNA in 30 mL double-distilled water (ddH2O) 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 for the same time. Measurement of 6mA and deoxyadenosine (dA) in the DNA degradation 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 degradation 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.

Flow Cytometry. Parasites (5 × 10^6 cells per milliliter) were incubated overnight at 37 °C in the presence of 100 μM ETOH. The following day, cells were resuspended in fresh medium without ETOH, 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.

Cell Cycle Synchronization. Parasites (5 × 10^6 cells per milliliter) were incubated overnight at 37 °C in the presence of 100 μM ETOH. The following day, 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 FACSAurora; 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 1x immunoprecipitation (IP) buffer (50 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 1x 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 --rc-filter parameter and all other parameters at default settings. The BAM files were used as input for MACS2 (70), which was run with the --gc parameter set to 1.768 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)
and fold enrichment (>2). Only peaks with more than 75% overlap between replicates were considered for further analysis. HOMER2 (71) was used for peak calling. Peptide sequence information was obtained using Scaffold (27). Scaffold 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 μL (<1 × 10⁶ cells per milliliter) of T. vaginalis strain 8722-10-10-5 with 1% formaldehyde for 15 min at room temperature. Cross-linking was stopped by the addition of 125 μL of glycine and incubated 5 min at room temperature, then on ice for 15 min. Cells were harvested, washed in 10 mL PBS, 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 E.coli DNA Ligase (Promega) were added and incubated overnight at 65 °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/chloroform/isoamyl alcohol extraction followed by EtOH precipitation. Concentration was determined with a NanoDrop ND-1000 spectrophotometer. The primers used were synthesized at Integrated DNA Technologies (IDT) followed by verification of the cross-ligation product.

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 × 10⁷ 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 nM of a complementary DNA (cDNA) in a 15-μL reaction volume using a Stratagene MX3005P 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 lab for helpful discussions. This research was supported with Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) Grants 2015-2119 (to M.D.L.) and 2015-2121 (to N.O.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.O.M. are scientists 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.