(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 January 2020 (23.01.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/018917 A1

(51) International Patent Classification:
G01N 33/68 (2006.01)

(21) International Application Number:
PCT/US2019/042625

(22) International Filing Date:
19 July 2019 (19.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/701,536 20 July 2018 (20.07.2018) US
62/848,414 15 May 2019 (15.05.2019) US

(71) Applicant: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, 535 W. 116th Street, New York, New York 10027 (US).

(72) Inventors: **LANDWEBER, Laura**; 3 Queenston Pl., Princeton, New Jersey 08540 (US). **BEH, Yee Ming Leslie**; 121C Westwood Avenue, Singapore 648301 (SG).

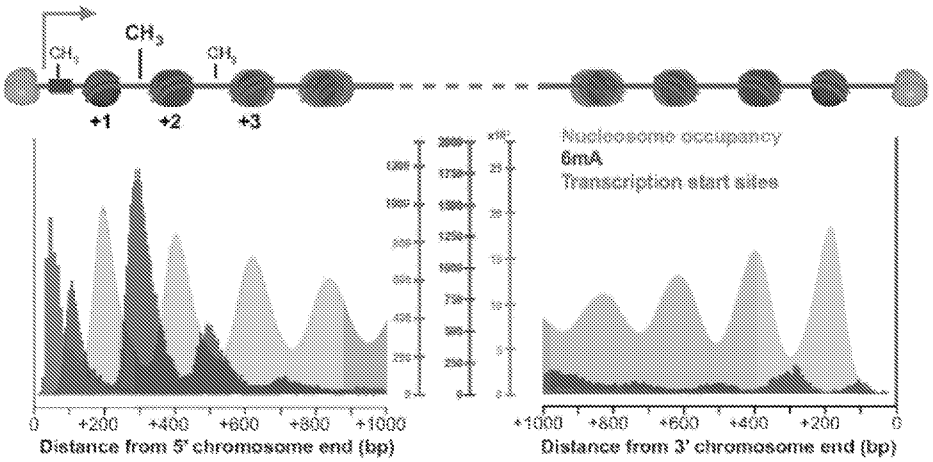
(74) Agent: **YANG, Ke**; BRYAN CAVE LEIGHTON PAISNER LLP, 1290 Avenue of the Americas, New York, New York 10104 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: NUCLEIC ACID MODIFICATION WITH TOOLS FROM OXYTRICHA

Fig. 1A



(57) Abstract: The present disclosure provides, *inter alia*, methods for treating a disease characterized by an abnormal level of m6dA in a subject, such as cancer, methods of modifying a nucleic acid from a cell, methods for identifying protein binding sites on DNA, methods of mediating DNA N6-adenine methylation, methods of modulating nucleosome organization and/or transcription in a cell, using MTA1c or any components thereof. The present disclosure also provides methods of generating a synthetic chromosome and synthetic chromosomes made by such methods. Pharmaceutical compositions comprising MTA1c or any components thereof and kits containing such compositions or for carrying out such processes are further provided. Eukaryotic cells, vectors and transgenic organisms comprising MTA1c or any components thereof are also provided. Synthetic chromosomes and methods of making same are also provided.

WO 2020/018917 A1

[Continued on next page]

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

NUCLEIC ACID MODIFICATION WITH TOOLS FROM OXYTRICHA

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims benefit of U.S. Provisional Patent Application Serial No. 62/701,536, filed on July 20, 2018, and U.S. Provisional Patent Application Serial No. 62/848,414, filed on May 15, 2019, which applications are incorporated by reference herein in their entireties.

FIELD OF DISCLOSURE

[0002] The present disclosure provides, *inter alia*, various methods, kits and compositions for modifying nucleic acid using MTA1c or any components thereof. Such embodiments may be used to treat disease and as research tools.

GOVERNMENT FUNDING

[0003] This invention was made with government support under GM059708 and GM122555 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE DISCLOSURE

[0004] Covalent modifications on DNA have long been recognized as a hallmark of epigenetic regulation. DNA N6-methyladenine (6mA) has recently come under scrutiny in eukaryotic systems, with proposed roles in retrotransposon or gene regulation, transgenerational epigenetic inheritance, and chromatin organization (Luo et al., 2015). 6mA exists at low levels in Arabidopsis thaliana (0.006%-0.138% 6mA/dA), rice (0.2%), C. elegans (0.01%-0.4%), Drosophila (0.001%-0.07%), Xenopus laevis (0.00009%), mouse embryonic stem cells (ESCs) (0.0006-0.007%), human cells (Greer et al., 2015; Koziol et al., 2016; Liang et al., 2018; Wu et al., 2016; Xiao et al., 2018; Zhang et al., 2015; Zhou et al., 2018), and the mouse brain (Yao et al., 2017), although it accumulates in abundance (0.1%-0.2%) during vertebrate embryogenesis (Liu et al., 2016). Disruption of DMAD, a 6mA demethylase, in the Drosophila brain leads to the accumulation of 6mA and Polycomb-mediated silencing (Yao et al., 2018). The existence of 6mA in mammals remains a subject of debate. Quantitative liquid chromatography-tandem mass

spectrometry (LC-MS/MS) analysis of HeLa and mouse ESCs failed to detect 6mA above background levels (Schiffers et al., 2017). A recent study, however, reported that loss of 6mA in human cells promotes tumor formation (Xiao et al., 2018), suggesting that 6mA is a biologically relevant epigenetic mark.

[0005] In contrast to metazoa, 6mA is abundant in various unicellular eukaryotes, including ciliates (0.18%-2.5%) (Ammermann et al., 1981; Cummings et al., 1974; Gorovsky et al., 1973; Rae and Spear, 1978), and the green algae Chlamydomonas (0.3%-0.5%) (Fu et al., 2015; Hattman et al., 1978). High levels of 6mA (up to 2.8%) were also recently reported in basal fungi (Mondo et al., 2017). Ciliates have long served as powerful models for the study of chromatin modifications (Brownell et al., 1996; Liu et al., 2007; Strahl et al., 1999; Taverna et al., 2002; Wei et al., 1998). They possess two structurally and functionally distinct nuclei within each cell (Bracht et al., 2013; Yerlici and Landweber, 2014). In the ciliate Oxytricha trifallax, the germline micronucleus is transcriptionally silent and contains ~100 megabase-sized chromosomes (Chen et al., 2014). In contrast, the somatic macronucleus is transcriptionally active, being the sole locus of Pol II-dependent RNA production in non-developing cells (Khurana et al., 2014). The Oxytricha macronuclear genome is extraordinarily fragmented, consisting of ~16,000 unique chromosomes with a mean length of ~3.2 kb, most encoding a single gene. Macronuclear chromatin yields a characteristic ~200 bp ladder upon digestion with micrococcal nuclease, indicative of regularly spaced nucleosomes (Gottschling and Cech, 1984; Lawn et al., 1978; Wada and Spear, 1980). Yet it remains unknown how and where nucleosomes are organized within these miniature chromosomes and if this in turn regulates (or is regulated by) 6mA deposition.

SUMMARY OF THE DISCLOSURE

[0006] The ciliate *Oxytricha* is a natural source of tools for RNA-guided genome reorganization and other nucleic acid modification. Long template RNAs instruct new linkages between pieces of DNA (Nowacki et al. 2008), and small RNAs instruct which DNA segments to keep (Fang et al. 2012) or eliminate. Foreseeable uses of these or other machinery derived from the *Oxytricha* genome include *in vitro* and/or *in vivo* modification of nucleic acids.

[0007] Intriguingly, in green algae, basal yeast, and ciliates, 6mA is enriched in ApT dinucleotide motifs within nucleosome linker regions near promoters (Fu et al., 2015; Hattman et al., 1978; Karrer and VanNuland, 1999; Mondo et al., 2017; Pratt and Hattman, 1981; Wang et al., 2017). In the present disclosure, four ciliate proteins-named MTA1, MTA9, p1, and p2 – have been identified as being necessary for 6mA methylation in a complex form termed MTA1c. MTA1 and MTA9 contain divergent MT-A70 domains, while p1 and p2 are homeobox-like proteins that likely function in DNA binding. The present disclosure delineates key biochemical properties of this methyltransferase and dissects the function of 6mA *in vitro* and *in vivo*.

[0008] The present disclosure provides a novel ciliate enzyme "MTA1" effective for N6-methyladenine (m6dA) methylation of DNA (see, *e.g.*, Appendix 4). MTA1 has been identified in a ciliate, *Tetrahymena thermophila*, and its functional role validated in m6dA methylation in *Oxytricha.* (See, Genbank ID: XP_001032074.3 [*Tetrahymena* MTA1] and EJY79437.1 [*Oxytricha* MTA1]). MTA1 is evolutionarily distinct from all known m6dA methyltransferases. Evolutionary analysis reveals that it is present in ciliates (including *Oxytricha* and *Tetrahymena*), algae, and basal fungi, but not multicellular eukaryotes. MTA1 exhibits a unique substrate specificity *in vivo*, being essential for the deposition of dimethylated AT (5'-A*T-3' / 3'-TA*-5'), as well as a wide range of other motifs *in vivo* (Figs. 1A – 1B). The inventors have been actively characterizing the biochemical properties and enzymology of *Tetrahymena* and *Oxytricha* MTA1, including its binding partners, *in vitro* substrate specificity (DNA vs. RNA and sequence motifs therein), methylation kinetics, and structural basis of these activities.

[0009] The present disclosure provides that MTA1c or any cmoponents thereof presents immediate commercial applications in: 1) generation of DNA substrates containing m6dA at locations distinct from known m6dA methyltransferases, circumventing the need for slow, expensive synthesis of methylated DNA; and 2) rational design of N6-adenine methylating enzymes with novel substrate specificities.

[0010] Accordingly, one embodiment of the present disclosure is a method of modifying a nucleic acid from a cell, the cell derived from a multicellular eukaryote. This method comprises the steps of: (a) obtaining the nucleic acid from the cell; and

(b) contacting the nucleic acid with MTA1c or any components thereof under conditions effective to methylate the nucleic acid.

[0011] The modified base, m6dA, has been discovered in a wide range of eukaryotes, including humans. m6dA levels are significantly reduced in gastric and liver cancer tissues, and disruption of m6dA promotes tumor formation (Xiao et al. 2018). As disclosed herein, MTA1 is a novel m6dA "writer", paving the way for cost-effective methods to understand mechanisms of m6dA function in biomedically relevant models.

[0012] Accordingly, another embodiment of the present disclosure is a method of treating or ameliorating the effects of a disease characterized by an abnormal level of m6dA in a subject. This method comprises administering to the subject an amount of MTA1c or any components thereof effective to modulate m6dA levels in the subject. In some embodiments, the modulation comprises restoring m6dA levels to normal or near-normal ranges in the subject.

[0013] Another embodiment of the present disclosure is a pharmaceutical composition comprising MTA1c or any components thereof that is effective to modulate m6dA levels in a subject in need thereof and a pharmaceutically acceptable carrier, diluent, adjuvant or vehicle.

[0014] Yet another embodiment of the present disclosure is a kit for treating or ameliorating the effects of a disease characterized by an abnormal level of m6dA in a subject, such as, *e.g.*, cancer, comprising an effective amount of MTA1c or any components thereof, packaged together with instructions for its use.

[0015] Another embodiment of the present disclosure is a cell line obtained from a multicellular eukaryote comprising a nucleic acid encoding MTA1c or any components thereof and/or an MTA1c protein compolex or any components thereof. As used herein, a "cell line" refers to all types of cell lines such as, *e.g.*, immortalized cell lines and primary cell lines. In certain embodiments, the nucleic acid encoding MTA1c or any components thereof is operably linked to a recombinant expression vector.

[0016] Another embodiment of the present disclosure is a recombinant expression vector comprising a polynucleotide encoding MTA1c or any components thereof.

[0017] Still another embodiment of the present disclosure is a transgenic organism whose genome comprises a transgene comprising a nucleotide sequence encoding MTA1c or any components thereof. Non-limiting examples of possible organism include an archaea, a bacterium, a eukaryotic single-cell organism, algae, a plant, an animal, an invertebrate, a fly, a worm, a cnidarian, a vertebrate, a fish, a frog, a bird, a mammal, an ungulate, a rodent, a rat, a mouse, and a non-human primate.

[0018] The present disclosure also provides a method of identifying protein binding sites on DNA. This method comprises the steps of: (a) providing DNA; (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA; (c) contacting the DNA with one or more proteins; (d) contacting the DNA with an enzyme effective to hydrolize the DNA in positions where no protein binding occurs; (e) removing the DNA bound protein; and (f) isolating and sequencing the DNA fragments. In certain embodiments, the one or more proteins in step (c) comprise histone octamers.

[0019] Another embodiment of the present disclosure is a method of mediating DNA N6-adenine methylation. This method comprises the steps of: (a) providing DNA; and (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA.

[0020] Another embodiment of the present disclosure is a method of modulating nucleosome organization and/or transcription in a cell, comprising providing to the cell an agent that is effective to modulate the expression of MTA1c or any components thereof.

[0021] The present disclosure also provides a method of generating a synthetic chromosome. This method comprises the steps of: (a) generating chromosome segments containing terminal restriction sites, wherein the chromosome segments comprise one or more m6dA bases; (b) digesting the chromosome segments with a restriction enzyme; and (c) purifying and ligating the digested chromosome segments to form a synthetic chromosome. In some embodiments, the method further comprises enriching the synthetic chromosome. A synthetic chromosome made by the method above is also provided.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0023] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present disclosure. The disclosure may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0024] Figs. 1A – 1E show epigenomic profiles of *Oxytricha* chromosomes.

[0025] Fig. 1A shows meta-chromosome plots of chromatin organization at *Oxytricha* macronuclear chromosome ends. Heterodimeric telomere end-binding protein complexes (orange ovals) protect each end in vivo. Horizontal red bar: promoter. The 5' chromosome end is proximal to TSSs. Nucleosome occupancy, normalized Mnaseseq coverage; 6mA, total 6mA number; Transcription start sites, total number of called TSSs.

[0026] Fig. 1B shows histograms of the total number of 6mA marks within each linker in *Oxytricha* chromosomes. Distinct linkers are depicted as horizontal blue lines.

[0027] Fig. 1C shows that poly(A)-enriched RNA-seq levels positively correlate with 6mA. Genes are sorted according to the total number of 6mA marks 0-800 bp downstream of the TSS. FPKM, fragments per kilobase of transcript per million mapped RNA-seq reads. Notch in the boxplot denotes median, ends of boxplot denote first and third quartiles, upper whisker denotes third quantile + 1.5 x interquartile range (IQR), and lower whisker denotes data quartile 1 - 1.5 x IQR.

[0028] Fig. 1D shows that composite analysis of 65,107 methylation sites reveals that 6mA (marked with 1 occurs within a 5'-ApT-3' dinucleotide motif.

[0029] Fig. 1E provides the distribution of various 6mA dinucleotide motifs across the genome. Asterisk, 6mA.

[0030] Figs. 2A – 2G show purification and characterization of the ciliate 6mA methyltransferase.

[0031] Fig. 2A provides phylogenetic analysis of MT-A70 proteins. Bold MTA1 and MTA9 genes are experimentally characterized in this study. Paralogs of MTA1

and MTA9 are labeled as "-B." Posterior probabilities >0.65 are shown. Gray triangle represents outgroup of bacterial sequences. The complete phylogenetic tree is shown in Fig. 9G. Gene names are in Table 5. Tth, *Tetrahymena thermophila*; Otri, *Oxytricha trifallax*.

[0032] Fig. 2B shows the phylogenetic distribution of the occurrence of ApT 6mA motifs and MT-A70 protein families. Filled square denotes its presence in a taxon. The basal yeast clade is comprised of *L. transversale, A.repens, H. vesiculosa, S. racemosum, L. pennispora, B. meristosporus, P. finnis*, and *A. robustus*.

[0033] Fig. 2C is an experimental scheme depicting the partial purification of DNA methyltransferase activity from *Tetrahymena* nuclear extracts.

[0034] Fig. 2D show gene expression and protein abundance of candidate genes in partially purified *Tetrahymena* nuclear extracts. UniProt IDs are listed in Table 5. RNA-seq data are from (Xiong et al.. 2012). FPKM, fragments per kilobase of transcript per million mapped RNA-seq reads. Low, Mid, and High DNA methylase activity correspond to fractions eluting from the Nuvia cPrime and Superdex 200 columns in Fig. 2C. Total spectrum counts, total number of LC-MS/MS fragmentation spectra that match peptides from a target protein.

[0035] Fig. 2E shows DNA methyltransferase assay using [3H]SAM. Vertical axis represents scintillation counts. Error bars represent SEM (n = 3).

[0036] Fig. 2F shows dot blot assay using cold SAM.

[0037] Fig. 2G shows DNA methyltransferase assay performed on different nucleic acid substrates in the presence of MTA1, MTA9, p1, and p2. Sense ssDNA are 5' → 3'; antisense are 3' → 5'. ApT dinucleotides are labeled in bold red. Horizontal blue lines in hemimethylated dsDNA substrates denote possible locations where 6mA may be installed by *EcoGII* (prior to this assay). Relative activity denotes scintillation counts normalized against the unmethylated 27 bp dsDNA substrate with two ApT motifs (top-most dsDNA substrate). An enlarged bar plot of relative activity on 27 bp unmethylated dsDNA substrates is included in Fig. 10K. Error bars represent SEM (n = 3).

[0038] Figs. 3A – 3E show genome-wide loss of 6mA in *mta1* mutants.

[0039] Fig. 3A shows schematic depicting the disruption of *Oxytricha* MTA1 open reading frame. Flanking dark blue bars: 5' and 3' UTR; yellow, open reading

frame; red, retention of 62 bp ectopic DNA segment; gray bar, intron; Internal light blue bar, annotated MT-A70 domain; ATG, start codon; TGA, stop codon. Agarose gel analysis shows PCR confirmation of ectopic DNA retention.

[0040] Fig. 3B shows dot blot analysis of RNase-treated genomic DNA.

[0041] Fig. 3C shows histogram of 6mA counts near 5' and 3' *Oxytricha* chromosome ends. Inset depicts histogram of fold change in total 6mA in each chromosome, between mutant and wild-type cell lines.

[0042] Fig. 3D shows that chromosomes are sorted into 10 groups according to total 6mA in wild-type cells (blue boxplots). For each group, the total 6mA per chromosome in mutants and the difference in total 6mA per chromosome are plotted below. Boxplot features are as described in Fig. 1C.

[0043] Fig. 3E shows motif distribution in wild-type and *mta1* mutants. Loss of ApT dimethylated motif is underlined.

[0044] Figs. 4A – 4E show effects of 6ma on nucleosome organization *in vitro* and *in vivo*.

[0045] Fig. 4A shows the experimental workflow for the generation of mini-genome DNA.

[0046] Fig. 4B shows agarose gel analysis of *Oxytricha* gDNA (Native) and mini-genome DNA before chromatin assembly.

[0047] Fig. 4C shows that methylated regions exhibit lower nucleosome occupancy *in vitro* but not *in vivo*. Overlapping 51 bp windows were analyzed across 98 chromosomes. For each window, the change in nucleosome occupancy in the absence versus presence of 6mA was calculated. Boxplot features are as described in Fig. 1C. p values were calculated using a two-sample unequal variance t test. N.S., non-significant, with $p > 0.05$.

[0048] Fig. 4D shows the reduction in nucleosome occupancy at methylated loci *in vitro* (black arrowheads). For *in vitro* MNase-seq, + 6mA refers to chromatin assembled on *Oxytricha* gDNA, while —6mA denotes chromatin assembled on mini-genome DNA. The vertical axis for SMRT-seq data denotes confidence score [-10 log(p value)] of detection of 6mA, while that for *in vitro* MNase-seq data denotes nucleosome occupancy.

[0049] Fig. 4E shows no change in nucleosome occupancy in linker regions despite loss of 6mA in *mta1* mutants. Vertical axes are the same as Fig. 4D.

[0050] Figs. 5A – 5C show modular synthesis of full-length *Oxytricha* chromosomes.

[0051] Fig. 5A shows features of the chromosome selected for synthesis. Gray boxes represent exons. All data tracks represent normalized coverage except for SMRT-seq, which represents the confidence score [-10 log(p value)] of detection of each methylated base.

[0052] Fig. 5B shows the schematic of chromosome construction. Different colors denote DNA building blocks ligated to form the full-length chromosome. Precise 6mA sites (bold red) represent cognate 6mA positions revealed by SMRT-seq in native genomic DNA. These are introduced via oligonucleotide synthesis. For chromosome 5, 6mA sites (non-bold red) represent possible locations ectopically installed by a bacterial 6mA methyltransferase, EcoGII. Intervening sequence within chromosomes 5 and 6 is represented as "...".

[0053] Fig. 5C shows native polyacrylamide gel analysis and anti-6mA dot blot analysis of building blocks and purified synthetic chromosomes.

[0054] Figs. 6A – 6E show quantitative modulation of nucleosome occupancy by 6mA.

[0055] Fig. 6A shows the experimental workflow. Chromatin is assembled using either salt dialysis or the NAP1 histone chaperone. Italicized blue steps are selectively included.

[0056] Fig. 6B shows the tiling qPCR analysis of synthetic chromosome with cognate 6mA sites. Horizontal gray box represents annotated gene, and vertical black lines depict native 6mA positions. Horizontal blue bars span ~100 bp regions amplified by qPCR. Red horizontal lines represent the region containing 6mA. Hemi methyl chromosomes contain 6mA on the antisense and sense strands, respectively, while the Full methyl chromosome has 6mA on both strands. Black arrowheads: decrease in nucleosome occupancy specifically at the 6mA cluster.

[0057] Fig. 6C shows the tiling qPCR analysis of ectopically methylated synthetic chromosome. Vertical black lines illustrate possible 6mA sites installed enzymatically. Red arrowheads: decrease in nucleosome occupancy in the ectopically methylated region. Black arrowheads: position of cognate 6mA sites (not in this construct).

[0058] Fig. 6D shows the tiling qPCR analysis of chromatin from Fig. 6B that is subsequently incubated with ACF and/or ATP. ACF equalizes nucleosome occupancy between the 6mA cluster and flanking regions in the presence of ATP (black line). Nucleosome occupancy at the methylated region is not restored to the same level as the unmethylated control (black arrowheads).

[0059] Fig. 6E shows that MNase-seq analysis of chromatin is assembled on native gDNA ("+" 6mA) and mini-genome DNA ("—" 6mA) using NAP1 ± ACF and ATP. p values were calculated using a two-sample unequal variance t test.

[0060] Figs. 7A – 7F show effects of 6mA on gene expression and cell viability *in vivo*.

[0061] Fig. 7A shows the following: Horizontal axis: the mean RNA-seq counts across all biological replicates from wild-type and *mta1* mutant data for each gene. Vertical axis: log2(fold change) in gene expression (mutant/wild type).

[0062] Fig. 7B shows that upregulated genes tend to be sparsely methylated compared to randomly subsampled genes (gray lines).

[0063] Fig. 7C shows RNA-seq analysis of MTA1 expression during the sexual cycle of *Oxytricha*. RNA-seq time course data are from Swart et al. (2013). The total duration of the sexual cycle is ~60 h.

[0064] Fig. 7D shows survival analysis of *Oxytricha* cells during the sexual cycle. The total cell number at each time point is normalized to 27 h data to obtain the percentage survival. Error bars represent SEM (n = 4).

[0065] Fig. 7E is a model illustrating the impact of 6mA methylation by MTA1c on nucleosome organization and gene expression.

[0066] Fig. 7F shows the comparison of DNA and RNA N6-adenine methyltransferases. Blue denotes catalytic subunit; yellow denotes subunit with predicted DNA or RNA binding domain.

[0067] Figs. 8A – 8B show MS analysis of 6mA in ciliate DNA.

[0068] Fig. 8A shows that *Oxytricha* and *Tetrahymena* genomic DNA were digested into nucleosides using degradase enzyme mix, followed by analysis using reverse-phase HPLC and mass spectrometry. Isotopically labeled dA and 6mA standards ($^{15}N_5$-dA and D_3-6mA) were mixed with each sample to allow quantitative measurement of endogenous dA and 6mA concentrations. MS/MS analysis of labeled dA and 6mA standards confirmed the mass of the nucleobase. Fluted peaks

with expected masses of dA and 6mA, and with highly similar retention times (RT) to internal standards are detected in *Oxytricha* and *Tetrahymena* nucleosides.

[0069] Fig. 8B shows the quantitation of dA and 6mA levels in *Oxytricha* and *Tetrahymena* gDNA using internal isotopically labeled nucleoside standards. The detected level of 6mA in *Tetrahymena* gDNA agrees with earlier reports (Gorovsky et al., 1973; Pratt and Hallman, 1981). The calculated abundance of 6mA relative to (dA + 6mA) in *Oxytricha* is ~0.71%, which is similar to the estimate from SMRT-seq base calls (0.78 -1.04%). Note that the calculation from SMRT-seq data is expected to be an overestimate because 6mA is scored at being present or absent at each site in the genome for this purpose. In actual fact, 6mA sites may be partially methylated (Fig. 11A). Neither 6mA nor dA was detected from LC-MS analysis of *Oxytricha* culture media, arguing against spurious signal arising from contamination or overall technical handling. The PacBio and LC-MS measurements of % 6mA in *Oxytricha* are both similar to thin layer chromatography analysis of nucleotides (0.6 - 0.7%) from a distinct but closely related species, *Oxytricha fallax* (Rae and Spear, 1978).

[0070] Figs. 9A – 9K show analysis of 6mA and methyltransferase components in *Tetrahymena*.

[0071] Fig. 9A shows *Tetrahymena* MNase-seq data from (Beh et al., 2015), while SMRT-seq data were generated in the present disclosure. Meta-chromosome plots overlaying *in vivo* MNase-seq (nucleosome occupancy) and SMRT-seq (6mA), relative to annotated transcription start sites. 6mA lies mainly within nucleosome linker regions, between the +1, +2, +3, and +4 nucleosomes.

[0072] Fig. 9B shows histograms of the total number of 6mA marks within each linker in *Tetrahymena* genes. Calculations are performed as described in Fig. 1B. Distinct linkers are highlighted with horizontal bold blue lines.

[0073] Fig. 9C shows the relationship between transcriptional activity and total number of 6mA marks in *Tetrahymena* genes. Analysis is performed as in Fig. 1C. RNA-seq data was obtained from (Xiong et al., 2012).

[0074] Fig. 9D shows that composite analysis of 441,618 methylation sites reveals that 6mA occurs within a 5'-ApT-3' dinucleotide motif in *Tetrahymena*, consistent with previous experiments (Bromberg et al., 1982; Wang et al., 2017) and similar to *Oxytricha*.

[0075] Fig. 9E shows distribution of various 6mA dinucleotide motifs across the genome.

[0076] Fig. 9F shows organization of transcription (mRNA-seq), nucleosome organization (MNase-seq), and 6mA (SMRT-seq) in a *Tetrahymena* gene.

[0077] Fig. 9G shows that all sequences used for phylogeny construction are listed in Table 1. Abbreviations: Cel: *Caenorhabditis elegans*; Ath: *Arabidopsis thaliana*; Sra: *Syncephalastrum racemosum*; Hve: *Hesseltinella vesiculosa*; Are: *Absidia repens*; Dre: *Danio redo*; Has: *Homo sapiens*; Ssc: Sus scrota; Mmu: *Mus musculus*; Xla: *Xenopus laevis*; Dme: *Drosophila melanogaster*; Cre: *Chlamydomonas reinhardtii*; Ltr: *Lobosporangium transversale*; Lpe: *Linderina pennispora*; Bme: *Basidiobolus meristosporus*; Pfi: *Piromyces finnis*; Aro: *Anaeromyces robustus*; Tth: *Tetrahymena thermophila*; Otri: *Oxytricha trifallax*. This Bayesian phylogenetic tree of MT-A70 proteins is the same as in Fig. 2A, except that all sequences are now included and labeled. TAMT-1 proteins are named according to (Luo et al., 2018).

[0078] Fig. 9H shows Bayesian phylogenetic tree of p1 proteins.

[0079] Fig. 9I shows Bayesian phylogenetic tree of p2 proteins. Dashed box depicts outgroup consisting of vertebrate SNAPC4 genes. These genes bear weak similarity to the homeobox-like domain of p2 proteins, but do not group phylogenetically with them and are therefore unlikely to be functionally homologs.

[0080] Fig. 9J shows phylogenetic distribution of ApT 6mA motif and various proteins, as depicted in Fig. 2B, but now also including TAMT-1, p1, and p2 proteins. Filled boxes denote the presence of a particular protein in a taxon. Open dashed boxes indicate the presence of SNAPC4 genes in vertebrates.

[0081] Fig. 9K shows the gene expression profiles of *Tetrahymena* MTA1, MTA9, p1 and p2. Microarray counts represent poly(A)' expression levels, and are obtained from TetraFGD (Miao et al., 2009; Xiong et al., 2011). MTA1, MTA9, p1 and p2 were found in our study to co-elute with 6mA methylase activity. On the other hand, TAMT-1 is a putative DNA methyltransferase described by (Luo et al., 2018). The horizontal axis categories beginning with "S" and "C" represent the number of hours since the onset of starvation and conjugation (mating), respectively. "Low," "Med," and "High" denote relative cell densities during log-phase growth. Blue and orange traces represent data from two biological replicates. Green and red shaded

regions show the peaks in poly(A)* RNA expression in vegetative growth and conjugation, respectively, for MTA1, MTA9, p1 and p2. Note that their expression pattern differs from TAMT-1.

[0082] Figs. 10A – 10N show further characterization of 6mA methyltransferase activity and MTA1c.

[0083] Fig. 10A shows that fractionation of nuclear extracts on a Q Sepharose column results in two distinct peaks of DNA methyltransferase activity, denoted as "Low Salt sample" and "High Salt sample" by black horizontal bars. FT denotes column flow-through. The DNA methyltransferase assay is performed as in Fig. 2E. The salt concentration at which individual fractions elute from the column is plotted against DNA methyltransferase activity of each fraction (counts per minute). Inset shows DNA methyltransferase activity of the input nuclear extract, flowthrough from the Q Sepharose column, and blank control (nuclear extract buffer). Orange and blue plots denote replicates derived from independent preparations of nuclear extract.

[0084] Fig. 10B is DNA methyltransferase assay showing that the activity from nuclear extracts is heat-sensitive and requires addition of DNA and SAM. Error bars represent s.e.m. (n = 3).

[0085] Fig. 10C is dot blot showing that nuclear extracts mediate 6mA methylation. Note that the low salt sample has substantial DNase activity, resulting in a lower amount of DNA available for dot blot analysis. DNA substrate, nuclear extract, and SAM cofactor were mixed as in panels A and B. The DNA was subsequently purified and used for dot blot analysis.

[0086] Fig. 10D shows domain organization of *Tetrahymena* MTA1, MTA9, p1, and p2. Protein domains are predicted using hmmscan on the EMBL-EBI webserver (Finn et al., 2015). "aa" denotes amino acids. Start and end coordinates of each domain are stated below each polypeptide.

[0087] Fig. 10E shows the sequence alignment of human (Hsa) METTL3 with *Tetrahymena* (Tth) and *Oxytricha* (Otri) MTA1 / MTA9, within the MT-A70 domain. Horizontal black bars underscore the DPPW catalytic motif, and the N549 / 0550 residues in human METTL3 that interact with the ribose moiety of the SAM cofactor. Note that the DPPW catalytic motif is conserved in MTA1 but not MTA9.

[0088] Fig. 10F shows dot blot analysis of hemimethylated dsDNA substrates. Sense or antisense oligonucleotides were first individually methylated using the

EcoGII bacterial 6mA methyltransferase. Each methylated ssDNA was subsequently purified and annealed with an unmethylated complementary strand to form hemimethylated constructs.

[0089] Fig. 10G shows SDS-PAGE analysis of recombinant proteins. Full length proteins were expressed and purified from *E. coli*. Bands of expected size are indicated with a black arrowhead.

[0090] Fig. 10H is methyttransferase assay using radiolabeled SAM on DNA and RNA substrates, coupled with gel analysis of nucleic acid integrity. ssRNA and dsRNA were produced by *in vitro* transcription from the 350bp dsDNA template using 17 RNA polymerase, and subsequently purified before use in this assay. Methyltransferase activity on equimolar amounts of each substrate was measured after incubation at 37°C for 6 hr, and depicted as either scintillation counts (Counts per minute), or normalized to the 350bp dsDNA sample (Relative activity). Only dsDNA, and not dsRNA or ssRNA, was methylated. Activity measurements are represented as scintillation counts (counts per minute). In addition, aliquots from each reaction containing DNA or RNA substrate and recombinant MTA1c (ie. MTA1, MTA7, p1 and p2 proteins) were withdrawn at 0, 1, 2, 3, or 6 hr during the 37°C incubation, purified using phenol:chloroform extraction and ethanol precipitation, and subsequently analyzed on a non-denaturing agarose gel. Both dsDNA and dsRNA substrates remained intact after 6 hr. The ssRNA migrates more diffusely on a nondenaturing agarose gel, with some decrease in size over time, suggesting partial degradation and/or RNA folding; however, there is no detectable methylation of ssRNA despite a significant presence on the agarose gel after 6 hr at 37°C. It is unlikely that this species is too short to be methylated, since MTA1c can methylate significantly shorter substrates such as 27bp dsDNA (Figs. 2G, 10I, 10J, and 10K). Error bars represent s.e.m. (n = 3).

[0091] Fig. 10I is DNA methyltransferase assay using radiolabeled SAM, on ssDNA oligonucleotides or annealed dsDNA substrates. All four recombinant MTA1c protein components — MTA1, MTA9, p1, and p2 — were included in each sample. Activity measurements are represented as scintillation counts (counts per minute). dsDNA substrates were prepared by annealing ssDNA oligonucleotides, as in Fig. 2G. Sense ssDNA nucleotide sequences are depicted in the 5' → 3' direction, while antisense ssDNA is depicted as 3' → 5'. Error bars represent s.e.m. (n = 3).

[0092] Fig. 10J is control [^{3}H]SAM assay using hemimethylated dsDNA. Reactions depicted in red represent hemimethylated dsDNA incubated with [3H]SAM in the absence of recombinant MTA1c (MTA1, MTA9, p1, and p2 proteins). These reactions showed no methyltransferase activity, verifying that there is no contaminating *EcoGII* methyltransferase in hemimethylated dsDNA preparations. Activity measurements are shown as scintillation counts, or as "Relative Activity" (normalized against the sample containing unmethylated DNA substrate, [3H]SAM, and MTA1c protein). Hemimethylated dsDNA substrates in this panel are the same as those used in Fig. 2G. The unmethylated dsDNA substrate used in this panel is the same as the top-most dsDNA substrate in Fig. 2G, with two uninterrupted ApT dinucleotides. Error bars represent s.e.m. (n = 3).

[0093] Fig. 10K is DNA methyltransferase assay using radiolabeled SAM, on dsDNA substrates with disrupted ApT dinucleotides. All four recombinant MTA1c protein components — MTA1, MTA9, p1, and p2 — were included in each sample. Activity measurements are normalized against the parent dsDNA construct with two uninterrupted ApT dinucleotides (top-most construct in this panel). ApT dinucleotide positions are labeled in bold red. Note that the parent dsDNA construct is identical to that in Fig. 10L. Error bars represent s.e.m. (n = 3).

[0094] Fig. 10L is DNA methyitransferase assay using radiolabeled SAM, on dsDNA substrates with shifted ApT dinucleotides. All four recombinant MTA1c protein components — MTA1, MTA9, p1, and p2 — were included in each sample. Activity measurements are normalized against the parent dsDNA construct with two uninterrupted ApT dinucleotides (top-most construct in this panel). The parent construct is identical to that in Fig. 10K. ApT dinucleotides are labeled in bold red. The adjacent nucleotides are labeled in bold black to highlight the 4-mer sequence that contains each ApT dinucleotide. Error bars represent s.e.m. (n = 3).

[0095] Fig. 10M shows motif frequencies of all 4-mer sequences containing methylated ApT dinucleotides in the *Tetrahymena* and *Oxytricha* genomes. A' denotes 6mA. The 4-mers TA'TA and CKTT are colored in red and blue, respectively, to highlight their large difference in genomic frequencies.

[0096] Fig. 10N shows motif frequencies of 4-mer sequences — regardless of methylation state — in *Tetrahymena* and *Oxytricha*. These were calculated from genomic sequence between the 5' chromosome end and the +4 nucleosome peak

(*Oxytricha*), or between the TSS and the +4 nucleosome peak (*Tetrahymena*). Analysis was restricted to these regions in order to serve as "background" frequencies for comparison to A'T methylated 4-mers, which are also mainly found downstream of TSSs. The 4-mers TATA and GATT are colored in red and blue, respectively, to facilitate comparison with methylated TA'TA and CA*TT in panel M.

[0097] Figs. 11A – 11D show supplemental SMRT-seq data analyses.

[0098] Fig. 11A shows the following: Top two panels depict PacBio coverage (horizontal axis) plotted against fractional methylation at each called 6mA site (vertical axis). Bottom left panel is a histogram of fractional methylation of all 6mA sites. Bottom right panel is a histogram of IPD ratios of all 6mA sites. Mutant datasets show significantly lower fractional methylation and IPD ratios at 6mA sites than wild-type data.

[0099] Fig. 11B shows that wild-type SMRT-seq data are randomly subsampled 15 times, such that the resulting coverage is lower than 'Mal mutant data. The difference in PacBio coverage between mutant and subsampled wild-type data is calculated for each chromosome, and is collectively represented as an olive boxplot (top panel). This set of calculations is repeated 15 times for each subsampled dataset, resulting in a series of 15 boxplots. The difference in PacBio coverage between mutant and fully sampled wild-type data is represented as a violet boxplot. Separately, the difference in total 6mA marks per chromosome is calculated for respective datasets, and boxplots are shown in the bottom panel. Mutant datasets consistently yield lower numbers of called 6mA marks than subsampled wild-type, despite the former having higher coverage than the latter.

[0100] Fig. 11C shows the scatterplot of total number of 6mA marks per chromosome in wild-type versus mutant data. PacBio cutoffs for calling 6mA marks are varied as shown. A greater number of 6mA marks per chromosome are consistently detected in wild-type than mutant data.

[0101] Fig. 11D shows the boxplot of PacBio chromosome coverage in individual wild-type and mutant biological replicates (left panel). Only chromosomes with 100-150x PacBio coverage are shown. The total number of 6mA marks in each of these chromosomes are plotted in the right panel. Wild-type replicates show consistently higher numbers of 6mA marks per chromosome than mutant replicates.

[0102] Figs. 12A – 12H show analysis of nucleosome organization and confirmation of ectopic DNA insertion in *mta1* mutants. Description of analysis in panels A-G: Nucleosomes are grouped according to their "starting" 6mA level, defined as the total number of 6mA marks ±200 bp from the nucleosome dyad in wild-type cells (WT). The dyad is assigned to be the peak position of MNase-seq reads. Similarly, linkers are grouped according to their "starting" methylation level, defined as the total number of 6mA marks between two flanking nucleosome dyads (or between the 5' chromosome end and the terminal nucleosome) in wild-type cells. Loci with high starting 6mA have methylation greater than or equal to the 90th percentile of starting 6mA levels, and show greater changes in methylation between mutant and wild-type cells (Fig. 3D). Those with low starting 6mA are in the lowest 10th percentile. if 6mA impacts nucleosome organization in vivo, then loci with high starting 6mA should show a greater change in nucleosome organization. Possible effects are illustrated in panels A — C. Vertical green lines depict 6mA marks, while blue and red peaks denote nucleosome occupancy. The plots shown in panels A — C illustrate the idealized result if 6mA disfavors nucleosomes in vivo. Actual effects are shown in panels D — G. "Wild type" is abbreviated as WT. Analyses are restricted to the 5' chromosome end.

[0103] Fig. 12A shows that 6mA loss may result in an increase in nucleosome fuzziness (highlighted with bold red double-sided arrow). The effect should be greater for nucleosomes with high starting 6mA due to greater change in 6mA between mutant and wild-type cells ("Change in nucleosome fuzziness" Box). Nucleosomes should, in turn, exhibit lower occupancy near the peak position, and higher occupancy in flanking regions ("Change in Nucleosome occupancy" Box; highlighted with red arrowheads and plotted ± 73bp from the dyad). Nucleosome fuzziness is calculated as the standard deviation of MNase-seq read locations ± 73bp from the dyad.

[0104] Fig. 12B shows that 6mA loss from nucleosome linker regions may result in a decrease in linker length (highlighted with bold red bracket). If so, the magnitude of decrease in linker length should be greater for linkers with high starting 6mA ("Change in linker length" Box).

[0105] Fig. 12C shows that 6mA loss may result in an increase in occupancy directly over the methylated linker region (highlighted with bold red bracket). If so,

the magnitude of increase in linker occupancy should be greater for regions with high starting 6mA ("Change in linker occupancy" Box). Linker occupancy denotes the average MNase-seq coverage ± 25bp from the midpoint between flanking nucleosome dyads or chromosome end. As an example, for the +1/+2 nucleosome linker, occupancy is calculated ± 25bp from the midpoint of the +1 and +2 nucleosome dyad positions. Since nucleosome linker length in *Oxytricha* is ~200bp (Fig. 12F, bottom panels), the genomic window used to calculate linker occupancy has minimal overlap with that for calculating nucleosome fuzziness and occupancy in panel A.

[0106] Fig. 12D shows the impact of 6mA loss on nucleosome fuzziness. For each nucleosome, the change in fuzziness between mutant and wild-type cells is calculated. Boxplots represent the distribution of changes in fuzziness scores. "MNase-seq" denotes sequencing of nucleosomal DNA obtained from Oxytricha chromatin *in vivo*, while "Control gDNA-seq" represents sequencing of MNase-digested, naked genomic DNA in vitro. Boxplot features are as described in Fig. 1C. Distributions are compared using a Wilcoxon rank-sum test. N.S denotes "non-significant," with $p > 0.01$.

[0107] Fig. 12E shows the impact of 6mA loss on nucleosome occupancy. For each nucleosome, the difference in nucleosome occupancy between mutant and wild-type cells is calculated at individual basepairs ± 73bp around the nucleosome dyad. Data are averaged and depicted as line plots. The change in occupancy at the dyad is compared between nucleosomes with high and low starting 6mA using a Wilcoxon rank-sum test.

[0108] Fig. 12F shows the impact of 6mA loss on linker length. Three types of linkers are analyzed: between the 5' chromosome end and +1 nucleosome dyad, between the +1 and +2 nucleosome dyads, and between the +2 and +3 nucleosome dyads. For each linker, the difference in its length between mutant and wild-type cells is calculated. The resulting distribution of linker length differences is plotted as a histogram (top-most row of this panel). Distributions of linker length differences are compared using two-sample unequal variance t test. N.S. indicates "not significant," with $p > 0.01$. Separately, the respective distributions of linker lengths in mutant and wild-type cells are plotted in the bottom two rows of this panel. The median linker length from each group is included as an inset.

[0109] Fig. 12G shows the impact of 6mA loss on linker occupancy. Linkers are binned as in panel F. For each linker, the difference in occupancy between mutant and wild-type cells is calculated. The resulting distribution of changes in linker occupancy is represented as a boxplot. Distributions are compared using two-sample unequal variance t test. N.S. indicates "not significant," with p> 0.01. Boxplot features are as described in Fig. 1C.

[0110] Fig. 12H shows poly(A)$^+$ RNaseq analysis of wild-type and *mta1* mutants. "ATG" denotes start codon of MTA1 gene. A 62bp ectopic DNA insertion results in a frameshift mutation in the MTA1 coding region. Three wild-type (WTI, WT2, wr3) and mutant (mtal', mta12, mta13) biological replicates are analyzed. Short horizontal bars represent RNaseq reads, which are ,-.75 nt in length and mapped to the reference sequence. For a read to be successfully mapped, it must have no more than 2 mismatches relative to the reference sequence. Unmapped reads are discarded. Blue and red bars denote RNaseq reads that map to native and ectopic regions, respectively. RNaseq reads overlapping the ectopic region are detected in mutant but not wild-type replicates. These reads span junctions between the ectopic and flanking coding regions, confirming the site of ectopic insertion.

[0111] Figs. 13A – 13I show gel analysis of histone octamers and assembled chromatin. Description for panels A-D: *Xenopus* unmodified core histones were recombinantly expressed. *Oxytricha* histones were acid-extracted from vegetative nuclei. *Oxytricha* and *Xenopus* histones were subsequently refolded into octamers and purified through size exclusion chromatography. Description for panels E-I: *Xenopus* or *Oxytricha* histone octamers were assembled on DNA and subsequently digested with MNase to obtain ~150bp mononucleosome-sized fragments (labeled with red arrowheads). The resulting products were visualized by agarose gel electrophoresis. Mononucleosomal DNA was gel-excised and analyzed using Illumina sequencing or tiling qPCR analysis in Figs. 4A – 4E, 6A – 6E, and 14A – 14F.

[0112] Fig. 13A shows reverse-phase HPLC purification of acid-extracted *Oxytricha* histones. Fractions 1-5 were individually collected and analyzed by Coomassie staining and western blotting.

[0113] Fig. 13B shows SDS-PAGE analysis of purified *Oxytricha* histone fractions.

[0114] Fig. 13C shows Western blot analysis of *Oxytricha* histone fractions 1-5. The fraction that is most enriched in each type of histone is colored in red. Arrowheads indicate likely histone bands.

[0115] Fig. 13D shows SDS-PAGE analysis of purified *Oxytricha* and *Xenopus* histone octamers.

[0116] Fig. 13E shows that chromatin was assembled on PCR-amplified *Oxytricha* mini-genome DNA, digested with MNase, and analyzed by agarose gel electrophoresis.

[0117] Fig. 13F shows that chromatin was assembled on native *Oxytricha* genomic DNA, digested with MNase, and analyzed by agarose gel electrophoresis.

[0118] Fig. 13G shows that chromatin was assembled with synthetic chromosome DNA, digested with MNase, and visualized by agarose gel electrophoresis. All assemblies with synthetic chromosomes were performed in the presence of an approximately 100-fold mass excess of buffer DNA relative to synthetic chromosome (see Example 1). This applies to panels G, H, and I. Representative assemblies with the unmethylated chromosome are shown. Methylated chromosome assemblies were separately performed in place of the unmethylated variant.

[0119] Fig. 13H shows that chromatin was assembled on unmethylated synthetic chromosomes by salt dialysis and subsequently incubated with ACF and/or ATP. The resulting mixture was digested with MNase and visualized by agarose gel electrophoresis. Regularly spaced nucleosomes (labeled with red dots) are observed only when chromatin was incubated with both ACF and ATP.

[0120] Fig. 13I shows chromatin assembled on unmethylated synthetic chromosomes using the NAP1 histone chaperone in the presence of ACF and/or ATP. The resulting mixture was digested with MNase and visualized by agarose gel electrophoresis. Nudeosomes are regularly spaced (labeled with red dots) in the presence of both ACF and ATP, although less apparent than in panel H.

[0121] Figs. 14A – 14F show control MNase-Seq and tiling qPCR analysis.

[0122] Fig. 14A is the same analysis as Fig. 4C, showing that 6mA quantitatively disfavors nucleosome occupancy *in vitro* but not *in vivo*. Here, the extent of MNase digestion was 40% of that in Fig. 4C. P-values were calculated

using a two-sample unequal variance t test. N.S denotes "non-significant," with $p > 0.05$.

[0123] Fig. 14B is the same analysis as Fig. 6E, showing that the ACF complex restores nucleosome occupancy over methylated DNA in an ATP-dependent manner in vitro. Here, the extent of MNase digestion was 25% of that in Fig. 6E. P-values were calculated using a two-sample unequal variance t test. N.S denotes "non-significant," with $p > 0.05$.

[0124] Fig. 14C is the same analysis as Fig. 12D, showing that nucleosomes with high starting 6mA show larger changes in fuzziness. Here, the extent of MNase digestion was 40% of that in Fig. 12D. Distributions are compared using a Wilcoxon rank-sum test. N.S denotes "non-significant," with $p > 0.01$.

[0125] Fig. 14D is the same analysis as Fig. 12E, showing that nudeosomes with high starting 6mA exhibit characteristic changes in nucleosome occupancy at and around the nucleosome dyad. Here, the extent of MNase digestion was 40% of that in Fig. 12E. The change in dyad occupancy is compared between nucleosomes with high and low starting 6mA using a Wilcoxon rank-sum test. N.S denotes "non-significant," with $p > 0.01$.

[0126] Fig. 14E shows tiling qPCR analysis of nucleosome occupancy in spike-in and homogeneous synthetic chromosome preparations. The blunt, unmethylated synthetic chromosome (construct #1 in Fig. 5B) was used for chromatin assembly with ("Spike-in") or without ("Homogeneous") a 100-fold excess of buffer DNA. In the latter case, an equivalent mass of synthetic chromosome was added in place of buffer DNA to maintain the same DNA concentration for chromatin assembly. The tiling qPCR assay was performed as in Fig. 6B. Shaded red bars depict the regions where 6mA modulates nucleosome occupancy in separate methylated chromosomes analyzed in Figs. 6B and 6C. Note that methylated chromosomes were not used to generate qPCR data for this figure. Black arrowheads indicate no decrease in nucleosome occupancy in these regions when buffer DNA is used. Thus, the decrease in nucleosome occupancy in methylated chromosomes reported in Figs. 6A – 6E cannot be attributed to spike-in versus homogeneous addition of DNA for chromatin assembly. Error bars in all panels represent s.e.m. (n = 3-4).

[0127] Fig. 14F shows that chromatin was assembled on synthetic chromosomes using the NAP1 histone chaperone in the presence of ACF and/or ATP, instead of set dialysis. qPCR analysis was performed as in Fig. 6B. Methylated chromosomes used in this experiment contain 6mA in native sites. The addition of ACF and ATP results in a partial restoration of nucleosome occupancy over the methylated region. These results are similar to Fig. 6D, where chromatin was assembled by sat dialysis instead of NAP1.

[0128] Fig. 15 shows that ciliate methyltransferase MTA1c mediates DNA N6-adenine methylation (6mA) *in vivo* and 6mA directly disfavors nucleosome occupancy *in vitro*.

DETAILED DESCRIPTION OF THE DISCLOSURE

[0129] DNA N6-adenine methylation (6mA) has recently been described in diverse eukaryotes, spanning unicellular organisms to metazoa. In the present disclosure, it's reported a DNA 6mA methyltransferase complex in ciliates, termed MTA1c. It consists of two MT-A70 proteins and two homeobox-like DNA-binding proteins and specifically methylates dsDNA. Disruption of the catalytic subunit, MTA1, in the ciliate *Oxytricha* leads to genome-wide loss of 6mA and abolishment of the consensus ApT dimethylated motif. Mutants fail to complete the sexual cycle, which normally coincides with peak MTA1 expression. The present disclosure investigates the impact of 6mA on nucleosome occupancy in vitro by reconstructing complete, full-length *Oxytricha* chromosomes harboring 6mA in native or ectopic positions. It's shown that 6mA directly disfavors nucleosomes *in vitro* in a local, quantitative manner, independent of DNA sequence. Furthermore, the chromatin remodeler ACF can overcome this effect. The present disclosure identifies a diverged DNA N6-adenine methyltransferase and defines the role of 6mA in chromatin organization.

[0130] One embodiment of the present disclosure is a method of modifying a nucleic acid from a cell, the cell derived from a multicellular eukaryote. This method comprises the steps of: (a) obtaining the nucleic acid from the cell; and (b) contacting the nucleic acid with MTA1c or any components thereof under conditions effective to methylate the nucleic acid.

[0131] In some embodiments, the nucleic acid is RNA or DNA. In some embodiments, the eukaryotic cell is mammalian. In some embodiments, the multicellular eukaryote is a human. In some embodiments, the modification is a DNA N6-adenine methylation including one of more of the following motifs: dimethylated AT (5'-A*T-3'/3'-TA*-5'), dimethylated TA (5'-TA*-3'/3'-A*T-5'), dimethylated AA (5'-A*A*-3'/3'-TT-5'), methylated AT (5'-A*T-3'/3'-TA-5'), methylated AA (5'-A*A-3'/3'-TT-5'), methylated AC (5'-A*C-3'/3'-TG-5'), methylated AG (5'-A*G-3'/3'-TC-5'), methylated TA (5'-TA*-3'/3'-AT-5'), methylated AA (5'-AA*-3'/3'-TT-5'), methylated CA (5'-CA*-3'/3'-GT-5'), and methylated GA (5'-GA* -3'/3'-CT-5'). In certain embodiments, the MTA1 or an ortholog thereof comprises a mutation effective to abrogate dimethylation of the nucleic acid. Preferably, the mutation comprises loss of a C-terminal methyltransferase domain. In some embodiments, the MTA1c or any components thereof is obtained from ciliates, algae, or basal fungi. Preferably, the MTA1c or any components thereof is obtained from *Oxytricha* or *Tetrahymena*.

[0132] As used herein, an "ortholog," or orthologous gene, is a gene with a sequence that has a portion with similarity to a portion of the sequence of a known gene, but found in a different species than the known gene. An ortholog and the known gene originated by vertical descent from a single gene of a common ancestor. As used herein an ortholog encodes a protein that has a portion of at least about 50%, such as at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80% or at least about 80% of the total length of the sequence of the encoded protein that is similar to a portion of a length of at least about 50%, such as at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80% or at least about 80% of a known protein. The respective portion of the ortholog and the respective portion of the known protein to which it is similar may be a continuous sequence or be fragmented a number, for example, into 1 to about 3, including 2, individual regions within the sequence of the respective protein. For example, the 1 to about 3 regions are arranged in the same order in the amino acid sequence of the ortholog and the amino acid sequence of the known protein. Such a portion of an ortholog has an amino acid sequence that has at least about 40%, at least about 45%, such as at least about 55%, at least about 60%, at least about 65%, at least about 70%, at

least about 75% or at least about 80% sequence identity to the amino acid sequence of the known protein encoded by a MTA1 gene.

[0133] As used herein, an asterisk "*" indicates the presence of a methylated base. For example, "A*" represents a methylated adenine.

[0134] The modified base, m6dA, has been discovered in a wide range of eukaryotes, including humans. m6dA levels are significantly reduced in gastric and liver cancer tissues, and disruption of m6dA promotes tumor formation (Xiao et al. 2018). As disclosed herein, MTA1 is a novel m6dA "writer", paving the way for cost-effective methods to understand mechanisms of m6dA function in biomedically relevant models.

[0135] Accordingly, another embodiment of the present disclosure is a method of treating or ameliorating the effects of a disease characterized by an abnormal level of m6dA in a subject. This method comprises administering to the subject an amount of MTA1c or any components thereof effective to modulate m6dA levels in the subject. In some embodiments, the modulation comprises restoring m6dA levels to normal or near-normal ranges in the subject.

[0136] In some embodiments, the subject is a mammal that can be selected from the group consisting of humans, veterinary animals, and agricultural animals. Preferably, the subject is a human.

[0137] In some embodiments, the disease is a cancer, *e.g.*, gastric cancer or liver cancer. In certain embodiments, the method further comprises administering to the subject one or more of anti-gastric cancer and anti-liver cancer drugs. Non-limiting examples of anti-liver cancer drugs include Nexavar™ (Sorafenib Tosylate) and Stivarga™ (Regorafenib). Non-limiting examples of anti-gastric cancer drugs include Cyramza™ (Ramucirumab), Doxorubicin Hydrochloride, 5-FU (Fluorouracil Injection), Fluorouracil Injection, Herceptin™ (Trastuzumab), Mitomycin C, Taxotere™ (Docetaxel), Trastuzumab, Afinitor™ (Everolimus), Somatuline Depot™ (Lanreotide Acetate), FU-LV, TPF, and XELIRI.

[0138] In some embodiments, the method furthering comprises co-administering to the subject an epigenetic agent that is selected from the group consisting of methylation inhibiting drugs, Bromodomain inhibitors, histone acetylase (HAT) inhibitors, protein methyltransferase inhibitors, histone methylation inhibitors,

histone deacetlyase (HDAC) inhibitors, histone acetylases, histone deacetlyases, and combinations thereof.

[0139] Another embodiment of the present disclosure is a pharmaceutical composition comprising MTA1c or any components thereof that is effective to modulate m6dA levels in a subject in need thereof and a pharmaceutically acceptable carrier, diluent, adjuvant or vehicle.

[0140] Yet another embodiment of the present disclosure is a kit for treating or ameliorating the effects of a disease characterized by an abnormal level of m6dA in a subject, such as, *e.g.*, cancer, comprising an effective amount of MTA1c or any components thereof, packaged together with instructions for its use.

[0141] Another embodiment of the present disclosure is a cell line obtained from a multicellular eukaryote comprising a nucleic acid encoding MTA1c or any components thereof and/or an MTA1c protein complex or any components thereof. As used herein, a "cell line" refers to all types of cell lines such as, *e.g.*, immortalized cell lines and primary cell lines. In certain embodiments, the nucleic acid encoding MTA1c or any components thereof is operably linked to a recombinant expression vector.

[0142] Another embodiment of the present disclosure is a recombinant expression vector comprising a polynucleotide encoding MTA1c or any components thereof.

[0143] Still another embodiment of the present disclosure is a transgenic organism whose genome comprises a transgene comprising a nucleotide sequence encoding MTA1c or any components thereof. Non-limiting examples of possible organism include an archaea, a bacterium, a eukaryotic single-cell organism, algae, a plant, an animal, an invertebrate, a fly, a worm, a cnidarian, a vertebrate, a fish, a frog, a bird, a mammal, an ungulate, a rodent, a rat, a mouse, and a non-human primate.

[0144] The present disclosure also provides a method of identifying protein binding sites on DNA. This method comprises the steps of: (a) providing DNA; (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA; (c) contacting the DNA with one or more proteins; (d) contacting the DNA with an enzyme effective to hydrolize the DNA in positions where no protein binding occurs; (e) removing the DNA bound protein; and (f) isolating and

sequencing the DNA fragments. In certain embodiments, the one or more proteins in step (c) comprise histone octamers.

[0145] Another embodiment of the present disclosure is a method of mediating DNA N6-adenine methylation. This method comprises the steps of: (a) providing DNA; and (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA.

[0146] Another embodiment of the present disclosure is a method of modulating nucleosome organization and/or transcription in a cell, comprising providing to the cell an agent that is effective to modulate the expression of MTA1c or any components thereof.

[0147] The present disclosure also provides a method of generating a synthetic chromosome. This method comprises the steps of: (a) generating chromosome segments containing terminal restriction sites, wherein the chromosome segments comprise one or more m6dA bases; (b) digesting the chromosome segments with a restriction enzyme; and (c) purifying and ligating the digested chromosome segments to form a synthetic chromosome. In some embodiments, the method further comprises enriching the synthetic chromosome. A synthetic chromosome made by the method above is also provided.

EXAMPLES

[0148] The following examples are provided to further illustrate certain aspects of the present disclosure. These examples are illustrative only and are not intended to limit the scope of the disclosure in any way.

Example 1

Materials and Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-H2A	Active Motif	Cat #: 39111
Anti-H2B	Abcam	Cat #: 1790
Anti-H3	Abcam	Cat #: 1791
Anti-H4	Active Motif	Cat #: 39269
Anti-N6-methyladenosine antibody	Cedarlane Labs / Synaptic Systems	Cat #: 202003(SY)
Goat Anti-Rabbit IgG (H + L)-HRP Conjugate	Bio-Rad	1706515
Bacterial and Virus Strains		
One Shot TOP10 chemically competent *E. coli*	Thermo Fisher	Cat #: C404006
BL21(DE3) pLysS	Thermo Fisher	Cat #: 70-236-4
SHuffle T7 Express Competent *E. coli*	NEB	Cat #: C3029J
Lemo21 (DE3) Competent *E. coli*	NEB	Cat #: C2528J
Chemicals, Peptides, and Recombinant Proteins		
Micrococcal nuclease	NEB	Cat #: M0247S
Q5 Site-Directed Mutagenesis Kit	NEB	Cat #: E0554S
ProBlock Gold bacterial protease inhibitor cocktail	GoldBio	Cat #: GB-330-5
Proteinase K	Roche	Cat #: 3115879001
Phenol:Chloroform:IAA, 25:24:1	Thermo Fisher	Cat #: AM9732
TRIzol reagent	Thermo Fisher	Cat #: 15596026
DNA Polymerase I, Large (Klenow) Fragment	NEB	Cat #: M0210S
Klenow Fragment (3' → 5' exo-)	NEB	Cat #: M0212S
BsaI	NEB	Cat #: R3535S
EcoGII	NEB	Cat #: M0603S
T4 DNA ligase	NEB	Cat #: M0202M
Phusion DNA polymerase	NEB	Cat #: M0530L
S-adenosyl-L-methionine	NEB	Cat #: B9003S
Mouse NAP1	This study	N/A
Drosophila ACF complex	Active Motif	Cat #: 31509
Xenopus histones	This study	N/A
Polyvinyl alcohol	Sigma Aldrich	Cat #: P8136
Polyethylene glycol 8000	Sigma Aldrich	Cat #: P2139
Adenosine 5'-triphosphate (ATP)	Sigma Aldrich	Cat #: A6559-25UMO
Creatine phosphate	Sigma Aldrich	Cat #: 10621714001
Creatine kinase	Sigma Aldrich	Cat #: 10127566001
Power SYBR Green PCR master mix	Thermo Fisher	Cat #: 4367659
Gum Arabic	Sigma Aldrich	Cat #: G9752-1KG
3H-labeled S-adenosyl-L-methionine ([3H]SAM)	PerkinElmer	Cat #: NET155V250UC
Ultima Gold	PerkinElmer	Cat #: 6013326
DNA degradase plus enzyme	Zymo Research	Cat #: E2020
$^{15}N_5$-dA nucleoside	Cambridge Isotope Laboratories	Cat #: NLM-3895-25
D_3-6mA	Synthesized in this study	N/A
Critical Commercial Assays		
QIAquick gel extraction kit	QIAGEN	Cat #: 28706
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat #: E7490S
ScriptSeq v2 RNA-Seq Library Prep Kit	Illumina	Cat #: SSV21124
Nucleospin Tissue Kit	Takara Bio USA	Cat #: 740952.250
MinElute Reaction Cleanup Kit	QIAGEN	Cat #: 28206
NEBNext Ultra II DNA Library Prep Kit	NEB	Cat #: E7645S
Hi-Scribe T7 High Yield RNA Synthesis Kit	NEB	Cat #: E2040S
Dynabeads Protein A	Thermo Fisher	Cat #: 10001D
TOPO TA cloning kit	Thermo Fisher	Cat #: K457501
Deposited Data		
Oxytricha trifallax SMRT-seq	This study	SRA: SRX2335608 and SRX2335607
Tetrahymena thermophila SMRT-seq	This study	GEO: GSE94421
Oxytricha trifallax, all Illumina data (RNA-seq, 6mA-IP-seq, MNase-seq, gDNA-seq)	This study	GEO: GSE94421
Experimental Models: Organisms/Strains		
Oxytricha trifallax cells, strain JRB310	Lab collection	N/A
Oxytricha trifallax cells, strain JRB510	Lab collection	N/A
Oxytricha trifallax cells, mtal mutant	Lab collection	N/A
Tetrahymena thermophila cells, strain SB210	*Tetrahymena* stock center	Cat #: SD00703
Oligonucleotides		
All are listed in Table S4	IDT	N/A
Recombinant DNA		
pET-His-NAP1 (expression vector for recombinant NAP1)	This study	N/A
pET-XenH2A (expression vector for recombinant *Xenopus* histone H2A)	This study	N/A
pET-XenH2B (expression vector for recombinant *Xenopus* histone H2B)	This study	N/A

pET-XenH3 (expression vector for recombinant *Xenopus* histone H3)	This study	N/A
pET-XenH4 (expression vector for recombinant *Xenopus* histone H4)	This study	N/A
pET-HisSUMO-MTA1 (expression vector for recombinant *Tetrahymena* MTA1)	This study	N/A
pET-HisSUMO-MTA7 (expression vector for recombinant *Tetrahymena* MTA7)	This study	N/A
pET-HisSUMO-p1 (expression vector for recombinant *Tetrahymena* p1)	This study	N/A
pET-HisSUMO-p2 (expression vector for recombinant *Tetrahymena* p2)	This study	N/A
pCR-TOPO-syntheticChromosome (cloned synthetic chromosomes to verify accuracy of ligation of component DNA building blocks)	This study	N/A
Software and Algorithms		
Galaxy	Galaxy Community Hub	https://usegalaxy.org/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
TopHat2	TopHat2 (Mortazavi et al., 2008)	https://ccb.jhu.edu/software/tophat/index.shtml
Python 2.7.10	Python Software Foundation	https://www.python.org/download/releases/2.7/
CAGEr	Haberle et al., 2015	https://bioconductor.org/packages/release/bioc/html/CAGEr.html
SMRT Analysis 2.3.0	Pacific Biosciences	https://www.pacb.com/documentation/smrt-analysis-software-installation-v2-3-0/
PSI-BLAST	NCBI / NIH	https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on
CD-HIT	Huang et al., 2010	http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi
MAFFT	Katoh et al., 2017; Kuraku et al., 2013	https://mafft.cbrc.jp/alignment/software/
MrBayes / CIPRES Science Gateway	Miller et al., 2010	https://www.phylo.org/
R (v3.2.5)	The R Foundation	https://www.r-project.org/
hmmscan	Finn et al., 2015	https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan
Other		
Agencourt Ampure XP beads	Beckman Coulter	Cat #: A63880
Acid-extracted Oxytricha histones	This study	N/A
Slide-A-Lyzer 3.5K MWCO cassette	Thermo Fisher	Cat #: PI66110
Amersham Hybond-XL membrane	GE Healthcare	Cat #: RPN3035
Amersham Hybond-N+ membrane	GE Healthcare	Cat #: RPN119B
Volvic water	Amazon	https://www.amazon.com/Volvic-500ml-6-Pack/dp/B013PCKEM8/ref=sr_1_3_a_it?ie=UTF8&qid=1538873999&sr=8-3&keyword=volvic&dpID=51BqEyu6yDL&preST=_SY300_QL70_&dpSrc=srch

Oxytricha trifallax

[0149] Vegetative *Oxytricha trifallax* strain J RB310 was cultured at a density of 1.5 x 10^7 cells/L to 2.5 x 10^7 cells/L in Pringsheim media (0.11mM Na_2HPO_4, 0.08mM $MgSO_4$, 0.85mM $Ca(NO_3)_2$, 0.35mM KCl, pH 7.0) and fed daily with *Chlamydomonas reinhardtii*. Cells were filtered through cheesecloth to remove debris and collected on a 10 pm Nitex mesh for subsequent experiments.

Tetrahymena thermophila

[0150] Stock cultures of vegetative *Tetrahymena thermophila* strain SB210 were maintained in Neff medium (0.25% w/v proteose peptone, 0.25% w/v yeast extract, 0.5% glucose, 33.3pM $FeCl_3$). These cultures were inoculated into SSP medium (2% w/v proteose peptone, 0.1% w/v yeast extract, 0.2% w/v glucose, 33pM $FeCl_3$) and grown to log-phase (~3.5 x 10^5 cells/mL) through constant shaking at 125 rpm / 30°C.

in vivo MNase-seq

[0151] 3 x 10^5 vegetative *Oxytricha* cells were fixed in 1% w/v formaldehyde for 10 min at room temperature with gentle shaking, and then quenched with 125mM glycine. Cells were lysed by dounce homogenization in lysis buffer (20mM Tris pH 6.8, 3% w/v sucrose, 0.2% v/v Triton X-100, 0.01% w/v spermidine trihydrochloride) and centrifuged in a 10%-40% discontinuous sucrose gradient (Lauth et al., 1976) to purify macronuclei. The resulting macronuclear preparation was pelleted by centrifugation at 4000 x g, washed in 50ml TMS buffer (10mM Tris pH 7.5, 10mM $MgCl_2$, 3mM $CaCl_2$, 0.25M sucrose), resuspended in a final volume of 300 µL, and equilibriated at 37°C for 5 min. Chromatin was then digested with MNase (New England Biolabs) at a final concentration of 15.7 Kunitz Units / µL at 37°C for 1 min 15 s, 3 min, 5 min, 7 min 30sec, 10 min 30 s, and 15 min respectively. Reactions were stopped by adding 1/2 volume of PK buffer (300mM NaCl, 30mM Tris pH 8, 75mM EDTA pH 8, 1.5% w/v SDS, 0.5mg/mL Proteinase K). Each sample was incubated at 65°C overnight to reverse crosslinks and deproteinate samples. Subsequently, nucleosomal DNA was purified through phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation. Each sample was loaded on a 2% agarose-TAE gel to check the extent of MNase digestion. The sample exhibiting ~80% mononucleosomal species was selected for MNase-seq analysis, in accordance with previous guidelines (Zhang and Pugh, 2011). Mononucleosome-sized DNA was gel-purified using a QIAquick gel extraction kit (QIAGEN). Illumina libraries were prepared using an NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) and subjected to paired-end sequencing on an Illumina HiSeq 2500 according to manufacturer's instructions. All vecietative Tetrahymena MNase-sea data were obtained from (Beh et al., 2015).

poly(A)$^+$ RNA-seq and TSS sequencing

[0152] *Oxytricha* cells were lysed in TRIzol reagent (Thermo Fisher Scientific) for total RNA isolation according to manufacturer's instructions. Poly(A)$^+$ RNA was then purified using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). *Oxytricha* poly(A)$^+$ RNA was prepared for RNA-seq using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina). *Tetrahymena* poly(A)$^+$ RNA-seq data was obtained from (Xiong et al., 2012). The 5' ends of capped RNAs were enriched from vegetative *Oxytricha* total RNA using the RAMPAGE protocol

(Batut et al., 2013), and used for library preparation, Illumina sequencing and subsequent transcription start site determination (ie. "TSS-seq"). These data were used to plot the distribution of *Oxytricha* TSS positions in Fig. 1A. TSS positions used for analysis outside of Fig. 1A were obtained from (Swart et al., 2013) and (Beh et al., 2015). For RNaseq analysis of genes grouped according to "starting" methylation level level: total 6mA was counted between 100 bp upstream to 250 bp downstream of the TSS. Genes with high starting methylation have total 6mA in the 90th percentile and higher. Genes with low starting methylation have total 6mA at or below the 10th percentile.

Immunoprecipitation and Illumina sequencing of methylated DNA (6mA IP-seq)

[0153] Genomic DNA was isolated from vegetative *Oxytricha* cells using the Nucleospin Tissue Kit (Takara Bio USA, Inc.). DNA was sheared into 150bp fragments using a Covaris LE220 ultra-sonicator (Covaris). Samples were gel-purified on a 2% agarose-TAE gel, blunted with DNA polymerase I (New England Biolabs), and purified using MinElute spin columns (QIAGEN). The fragmented DNA was dA-tailed using Klenow Fragment (3' - > 5' exo-) (New England Biolabs) and ligated to Illumina adaptors following manufacturer's instructions. Subsequently, 2.2 µg of adaptor-ligated DNA containing 6mA was immunoprecipitated using an anti-N6-methyladenosine antibody (Cedarlane Labs) conjugated to Dynabeads Protein A (Invitrogen). The anti-6mA antibody is commonly used for RNA applications, but has also been demonstrated to recognize 6mA in DNA (Fioravanti et al., 2013; Xiao and Moore, 2011). The immunoprecipitated and input libraries were treated with proteinase K, extracted with phenol:chloroform, and ethanol precipitated. Finally, they were PCR-amplified using Phusion Hot Start polymerase (New England Biolabs) and used for Illumina sequencing.

Sample preparation for SMRT-seq

[0154] Vegetative *Oxytricha* macronuclei were isolated as described in the subheading "in vivo MNase-seq" of this study. Vegetative *Tetrahymena* macronuclei were isolated by differential centrifugation (Beh et al., 2015). *Oxytricha* and *Tetrahymena* cells were not fixed prior to nuclear isolation. Genomic DNA was isolated from *Oxytricha* and *Tetrahymena* macronuclei using the Nucleospin Tissue Kit (Macherey-Nagel). Alternatively, whole *Oxytricha* cells instead of macronuclei were used. SMRT-seq according to manufacturer's instructions, using P5-C3 and

P6-C4 chemistry, as in (Chen et al., 2014). *Oxytricha* and *Tetrahymena* macronuclear DNA were used for SMRT-seq in Figs. 1A – 1E and 9A-9F, while *Oxytricha* whole cell DNA was used for all other Figures. Since almost all DNA in *Oxytricha* cells is derived from the macronucleus (Prescott, 1994), similar results are expected between the use of purified macronuclei or whole cells.

Illumina data processing

[0155] Reads from all biological replicates were merged before downstream processing. All Illumina sequencing data were quality trimmed (minimum quality score = 20) and length-filtered (minimum read length = 40nt) using Galaxy (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010). MNase-seq and 6mA IP-seq reads were mapped to complete chromosomes in the *Oxytricha trifallax* JRB310 (August 2013 build) or *Tetrahymena thermophila* SB210 macronuclear reference genomes (June 2014 build) using Bowtie2 (Langmead and Salzberg, 2012) with default settings, while poly(A). RNA-seq and TSS-seq reads were mapped using TopHat2 (Mortazavi et al., 2008) with August 2013 *Oxytricha* gene models or June 2014 *Tetrahymena* gene models, with default settings.

[0156] MNase-seq datasets were generated by paired-end sequencing. Within each MNase-seq dataset, the read pair length of highest frequency was identified. All read pairs with length ± 25bp from this maximum were used for downstream analysis. On the other hand, 6mA IP-seq datasets were generated by single-read sequencing. 6mA IP-seq single-end reads were extended to the mean fragment size, computed using cross-correlation analysis (Kharchenko et al., 2008). The per-basepair coverage of *Oxytricha* MNase-seq read pair centers and extended 6mA IP-seq reads were respectively computed across the genome. Subsequently, the per-basepair coverage values were normalized by the average coverage within each chromosome to account for differences in DNA copy number (and hence, read depth) between *Oxytricha* chromosomes (Swart et al., 2013). The per-basepair coverage values were then smoothed using a Gaussian filter of standard deviation = 15. This smoothed data is denoted as "normalized coverage" or "nucleosome occupancy." *Tetrahymena* MNase-seq data were processed similarly to *Oxytricha*, except that DNA copy number normalization was omitted as *Tetrahymena* chromosomes have uniform copy number (Eisen et al., 2006).

[0157] For the MNase-seq analysis in Figs. 4C, 6E, 14A, and 14B, nucleosome occupancy and 6mA IP-seq coverage were calculated within overlapping 51bp windows across the 98 assayed chromosomes. Windows were binned according to the number of 6mA residues within. The *in vitro* MNase-seq coverage from chromatinized native gDNA ("+" 6mA) was divided by the corresponding coverage from chromatinized mini-genome DNA ("-" 6mA) to obtain the fold change in nucleosome occupancy in each window. Alternatively, a subtraction was performed on these datasets to obtain the difference in nucleosome occupancy *in vitro*. Identical DNA sequences were compared for each calculation. These data are labeled as ("+" histones) in Figs. 4C and 14A. Naked native gDNA and mini-genome DNA were also MNase-digested, sequenced and analyzed in the same manner to control for Mnase sequence preferences ("-" histones). Nucleosome occupancy *in vivo* corresponds to normalized MNase-seq coverage from wild type and *mta1* mutant cells.

[0158] Nucleosome positions were iteratively called as local maxima in normalized MNase-seq coverage, as previously described (Beh et al., 2015). "Consensus" +1, +2, +3 nucleosome positions downstream of the TSS were inferred from aggregate MNase-seq profiles across the genome (Fig. 1A for *Oxytricha* and Fig. 9A for *Tetrahymena*). Each gene was classified as having a +1, +2, +3 and/or +4 nucleosome if there is a called nucleosome dyad within 75 bp of the consensus nucleosome position.

[0159] RNA-seq and TSS-seq read coverage were calculated without normalization by DNA copy number since there is no correlation between *Oxytricha* DNA and transcript levels (Swart et al., 2013).

[0160] *Oxytricha* TSSs were called from TSS-seq data using CAGEr (Haberle et al., 2015); with clusterCTSS parameters (threshold = 1.6, thresholdIsTpm = TRUE, nrPassThreshold = 1, method = "paraclu," removeSingletons = TRUE, keepSingletonsAbove = 5). Only TSSs with tags per million counts > 0.1 were used for downstream analysis. *Tetrahymena* TSSs were obtained from (Beh et al., 2015).

SMRT-seq data processing

[0161] We processed SMRT-seq data with SMRTPipe v1.87.139483 in the SMRT Analysis 2.3.0 environment using, in order, the P Fetch, P Filter (with minLength = 50, minSubreadLength = 50, readScore = 0.75, and artifact = -1000), P

FilterReports, P Mapping (with gff2Bed = True, pulsemetrics = DeletionQV, IPD, InsertionQV, PulseWidth, QualityValue, MergeQV, SubstitutionQV, DeletionTag, and load PulseOpts = byread), P_MappingReports, P_GenomicConsensus (with algorithm = quiver, outputConsensus = True, and enableMapQVFilter = True), P_ConsensusReports, and P Mod ificationDetection (with identifyModifcations = True, enableMapQVFilter = False, and mapQvThreshold = 10) modules. All other parameters were set to the default. The *Oxytricha* August 2013 reference genome build was used for mapping *Oxytricha* SMRT-seq reads, with Contig10040.0.1, Contig1527.0.1, Contig4330.0.1, and Contig54.0.1 removed, as they are perfect duplicates of other Contigs in the assembly. *Tetrahymena* SMRT-seq reads were mapped to the June 2014 reference genome build. Only chromosomes with high SMRT-seq coverage (> = 80x for Oxytricha; > = 100x for *Tetrahymena*) were used for all 6mA-related analyses.

Chromosome synthesis

[0162] Synthetic Contig1781.0 chromosomes were constructed from "building blocks" of native chromosome sequence (Figs. 5B and 5C). The dark blue building block in Fig. 5B was prepared by annealing synthetic oligonucleotides, while all other building blocks were generated by PCR-amplification from genomic DNA using Phusion DNA polymerase (New England Biolabs). All oligonucleotides used for annealing and PCR amplification are listed in Table 2. The PCR-amplified building blocks contain terminal restriction sites for *BsaI* (New England Biolabs), a type I IS restriction enzyme that cuts distal from these sites. BsaI cleaves within the native DNA sequence, generating custom 4nt 5' overhangs and releasing the non-native *BsaI* restriction site as small fragments that are subsequently purified away. The *BsaI*-generated overhangs are complementary only between adjacent building blocks, conferring specificity in ligation and minimizing undesired by-products. After *BsaI* digestion, PCR building blocks were purified by phenol:chloroform extraction and ethanol precipitation. Building blocks were then sequentially ligated to each other using T4 DNA ligase (New England Biolabs) and purified by phenol:chloroform extraction and ethanol precipitation. Size selection after each ligation step was performed using polyethylene glycol (PEG) precipitation or Ampure XP beads (Beckman Coulter) to enrich for the large ligated product over its smaller constituents. The size of individual building blocks and their corresponding order of

ligation were designed to maximize differences in size between ligated products and individual building blocks. This increases the efficiency in size selection of products over reactants. Chromosomes 1 and 6 in Fig. 5B was generated by full length PCR from genomic DNA. To prepare chromosomes 2-4 in Fig. 5B, the red, dark blue, and purple blocks were first ligated in a 3-piece reaction and purified from the individual components. This product was subsequently ligated with the turquoise building block to obtain the full length chromosome. To prepare chromosomes 5 in Fig. 5B, the red, orange, and emerald building blocks were ligated in a 3-piece reaction and subsequently purified. All chromosomes were subjected to Sanger sequencing to verify ligation junctions. 6mA was installed in synthetic chromosomes using annealed oligonucleotides, or by incubation of DNA building blocks with *EcoGII* methyltransferase (New England Biolabs).

Verification of synthetic chromosome sequences

[0163] All chromosomes were dA-tailed using Klenow Fragment (3' - > 5' exo-) (New England Biolabs), cloned using a TOPO TA cloning kit (Thermo Fisher) or StrataClone PCR Cloning Kit (Agilent Technologies), transformed into One Shot TOP10 chemically competent *E. coli*, and sequenced using flanking T7, T3, M13F, or M13R primers.

Preparation of *Oxytricha* histones

[0164] Vegetative *Oxytricha trifallax* strain JRB310 was cultured as described in the subheading: "Experimental model and subject details" of this study. Cells were starved for 14 hr and subsequently harvested for macronuclear isolation as described in the subheading: "*in vivo* MNase-seq" of this study. However, formaldehyde fixation was omitted. Purified nuclei were pelleted by centrifugation at 4000 x g, resuspended in 0.421mL 0.4N H_2SO_4 per 10^6 input cells, and nutated for 3 hr at 4°C to extract histones. Subsequently, the acid-extracted mixture was centrifuged at 21,000 x a for 15 min to remove debris. Proteins were precipitated from the cleared supernatant using trichloroacetic acid (TCA), washed with cold acetone, then dried and resuspended in 2.5% v/v acetic acid. Individual core histone fractions were purified from crude acid-extracts using semi-preparative RP-HPLC (Vydac C18, 12 micron, 10 mM x 250 mm) with 40%-65% HPLC solvent B over 50 min (Fig. 13A). The identity of each purified histone fraction was verified by western

analysis (Fig. 13C) using antibodies: anti-H2A (Active Motif #39111), anti-H2B (Abcam #ab1790), anti-H3 (Abcam #ab1791), anti-H4 (Active Motif #39269).

Preparation of recombinant *Xenopus* histones

[0165] All RP-HPLC analyses were performed using 0.1% TFA in water (HPLC solvent A), and 90% acetonitrile, 0.1% TFA in water (HPLC solvent B) as the mobile phases. Wild-type *Xenopus* H4, H3 C110A, H2B and H2A proteins were expressed in BL21(DE3) pLysS *E.coli* and purified from inclusion bodies through ion exchange chromatography (Debelouchina et al., 2017). Purified histones were characterized by ESI-MS using a MicrOTOF-Q II ESI-Qq-TOF mass spectrometer (Bruker Daltonics). H4: calculated 11,236 Da, observed 11,236.1 Da; H3 C110A: calculated 15,239 Da, observed 15,238.7 Da; H2A: calculated 13,950 Da, observed 13,949.8 Da; H2B: calculated 13,817 Da, observed 13,816.8 Da.

Preparation of histone octamers

[0166] *Oxytricha* and *Xenopus* histone octamers were respectively refolded from core histones using established protocols (Beh et al., 2015; Debelouchina et al., 2017). Briefly, lyophilized histone proteins (*Xenopus* modified or wild-type; *Oxytricha* acid-extracted) were combined in equimolar amounts in 6 M guanidine hydrochloride, 20 mM Tris pH 7.5 and the final concentration was adjusted to 1 mg/mL. The solution was dialyzed against 2M NaCl, 10mM Tris, 1mM EDTA, and the octamers were purified from tetramer and dimer species using size-exclusion chromatography on a Superdex 200 10/300 column (GE Healthcare Life Sciences). The purity of each fraction was analyzed by SDS-PAGE. Pure fractions were combined, concentrated and stored in 50% v/v glycerol at -20°C.

Preparation of mini-genome DNA

[0167] 98 full-length chromosomes were individually amplified from Oxytricha trifallax strain JRB310 genomic DNA using Phusion DNA polymerase (New England Biolabs). Primer pairs are listed in Table 2. Amplified chromosomes were separately purified using a MinElute PCR purification kit (QIAGEN), and then mixed in equimolar ratios to obtain "mini-genome" DNA. The sample was concentrated by ethanol precipitation and adjusted to a final concentration of ~1.6mg/mL.

Preparation of native genomic DNA for chromatin assemblystarry

[0168] Macronuclei were isolated from vegetative Oxytricha trifallax strain JRB310 as described in the subheading "in vivo MNase-seq" of this study. However,

cells were not fixed prior to nuclear isolation. Genomic DNA was purified using the Nucleospin Tissue kit (Macherey-Nagel). Approximately 200µg of genomic DNA was loaded on a 15%-40% linear sucrose gradient and centrifuged in a SW 40 Ti rotor (Beckman Coulter) at 160,070 x g for 22.5hr at 20°C. Sucrose solutions were in 1M NaCl, 20mM Tris pH 7.5, 5mM EDTA. Individual fractions from the sucrose gradient were analyzed on 0.9% agarose-TAE gels. Fractions containing high molecular weight DNA that migrated at the mobility limit were discarded as such DNA species were found to interfere with downstream chromatin assembly. All other fractions were pooled, ethanol precipitated, and adjusted to 0.5mg/mL DNA.

Chromatin assembly and preparation of mononucleosomal DNA

[0169] Chromatin assemblies were prepared by salt gradient dialysis as previously described (Beh et al., 2015; Luger et al., 1999), or using mouse NAP1 histone chaperone and *Drosophila* ACF chromatin remodeler as previously described (An and Roeder, 2004; Fyodorov and Kadonaga, 2003). Details of each chromatin assembly procedure are listed below. To reduce sample requirements while maintaining adequate DNA concentrations for chromatin assembly, synthetic chromosomes were first mixed with a hundred-fold excess of "buffer" DNA (PCR-amplified *Oxytricha* Contig17535.0). We verified that nucleosome occupancy in the methylated region (qPCR primer pairs 6 and 7) of the synthetic chromosome is unaffected by the presence of buffer DNA (Fig. 14E). Native and mini-genome DNA were not mixed with buffer DNA prior to chromatin assembly.

[0170] For chromatin assembly through salt dialysis: histone octamers and (synthetic chromosome + buffer) DNA were mixed in a 0.8:1 mass ratio, while histone octamers and (native or mini-genome) DNA were mixed in a 1.3:1 mass ratio, each in a 50 µL total volume. Samples were first dialyzed into start buffer (10mM Tris pH 7.5, 1.4M KCl, 0.1mM EDTA pH 7.5, 1mM DTT) for 1 hr at 4°C. Then, 350mL end buffer (10mM Tris pH 7.5, 10mM KCl, 0.1mM EDTA, 1mM DTT) was added at a rate of 1mUmin with stirring. The assembled chromatin was dialyzed ovemight at 4°C into 200mL end buffer, followed by a final round of dialysis in fresh 200mL end buffer for 1 hr at 4°C. The assembled chromatin was then adjusted to 50mM Tris pH 7.9, 5mM $CaCl_2$ and digested with MNase (New England Biolabs) to mainly mononucleosomal DNA as previously described (Beh et al., 2015).

[0171] For chromatin assembly using mouse NAP1 and *Drosophila* ACF: NAP1 was recombinantly expressed and purified as described in (An and Roeder, 2004). ACF was purchased from Active Motif. 0.49μM NAP1 and 58nM histone octamer were first mixed in a 302μl reaction volume containing 62mM KCl, 1.2% w/v polyvinyl alcohol (Sigma Aldrich), 1.2% w/v polyethylene glycol 8000 (Sigma Aldrich), 25mM HEPES-KOH pH 7.5, 0.1mM EDTA-KOH, 10% v/v glycerol, and 0.01% v/v NP-40. The NAP1-histone mix was incubated on ice for 30 min. Meanwhile, "AM" mix was prepared, consisting of 20mM ATP (Sigma Aldrich), 200mM creatine phosphate (Sigma Aldrich). 33.3mM $MgCl_2$, 33.3μg/μl creatine kinase (Sigma Aldrich) in a 56u1 reaction volume. After the 30 min incubation. 5.29 μl of 1.7 μM ACF complex (Active Motif) and the "AM" mix were sequentially added to the NAP1-histone mix. Then, 10.63μl of native or mini-genome DNA (2.66μg) was added, resulting in a 374μl reaction volume. The final mixture was incubated at 27°C for 2.5 hr to allow for chromatin assembly. Subsequently, $CaCl_2$ was added to a final concentration of 5mM, and the chromatin was digested with MNase (New England Biolabs) to mainly mononucleosomal DNA as previously described (Beh et al., 2015).

[0172] Mononucleosome-sized DNA from MNase-digested chromatin was gel-purified and used for tiling qPCR on a Viia 7 Real-Time PCR System with Power SYBR Green PCR master mix (Thermo Fisher), or *in vitro* MNase-seq on an Illumina HiSeq 2500, according to the manufacturer's instructions. qPCR primer sequences are listed in Table 2.

Tiling qPCR analysis of nucleosome occupancy

[0173] qPCR data were analyzed using the ΔΔCt method (Livak and Schmittgen, 2001). At each locus along the synthetic chromosome, ΔCt = (Ct at locus of interest) - (Ct at qPCR primer pair 22, far from the methylated region). See Fig. 6B for location of qPCR primer pair 22. Separate ΔCt values were calculated from mononucleosomal DNA and the corresponding naked, undigested synthetic chromosome. The ΔΔCt value was calculated from this pair of ΔCt values. This controls for potential variation in PCR amplification efficiency, especially over methylated regions. The fold change in mononucleosomal DNA relative to naked chromosomal DNA at a particular locus is calculated as $2^{-\Delta\Delta Ct}$, and denotes `nucleosome occupancy' for all presented qPCR data.

ACF spacing assay

[0174] ATP-dependent nucleosome spacing was performed in accordance with a previous study (Lieleg et al., 2015). Chromatin was assembled by salt gradient dialysis as described above, and then adjusted to 20mM HEPES-KOH pH 7.5, 80mM KCl, 0.5mM EGTA, 12% v/v glycerol, 10mM $(NH_4)_2SO_4$, 2.5mM DTT. Samples were then incubated for 2.5 hr at 27°C with 3mM ATP, 30mM creatine phosphate, 4mM $MgCl_2$, 5 ng/0 creatine kinase, and 11 ng/μL ACF complex (Active Motif). Remodeled chromatin was then adjusted to 5mM $CaCl_2$ and subjected to MNase digestion, mononucleosomal DNA purification, and qPCR analysis as described above.

Phylogenetic analysis

[0175] The MTA1 amino acid sequence (UniProt ID: J9IF92 9SPIT) was queried against the NCBI nr database using PSI-BLAST (Altschul et al., 1997; Schaffer et al., 2001) (maximum e-value = $1e^{-4}$; enable short queries and filtering of low complexity regions). Retrieved hits were collapsed using CD-HIT (Huang et al., 2010) with minimum sequence identity = 0.97 to remove redundant sequences. The resulting sequences were added to existing MT-A70 alignments from (Greer et al., 2015) using MAFFT (-add) (Katoh et al., 2017; Kuraku et al., 2013). Gaps and duplicate sequences were removed from the merged alignment. Only sequences corresponding to the taxa in Fig. 2B were retained. The alignment was then used for phylogenetic tree construction using MrBayes in the CIPRES Science Gateway (Miller et al., 2010) with 5×10^6 generations. Protein sequences used for MrBayes analysis are given in Table 1.

[0176] The above procedure was also used for constructing phylogenetic trees from p1 (UniProt ID: Q22VV9 TETTS) and p2 (UniProt ID: I7M8B9 TETTS). However, protein sequences were aligned using MAFFT without adding to an existing alignment.

Preparation of nuclear extracts with DNA methyltransferase activity

[0177] Vegetative *Tetrahymena* cells were grown in SSP medium to log-phase (~3.5×10^6 cells/mL) and collected by centrifugation at 2,300 x g for 5 min in an SLA-3000 rotor. The supernatant was discarded, and cells were resuspended in medium B (10mM Tris pH 6.75, 2mM $MgCl_2$, 0.1M sucrose, 0.05% w/v spermidine trihydrochloride, 4% w/v gum Arabic, 0.63% w/v 1-octanol, and 1mM PMSF). Gum

arabic (Sigma Aldrich) is prepared as a 20% w/v stock and centrifuged at 7,000 x g for 30 min to remove undissolved clumps. For each volume of cell culture, one-third volume of medium B was added to the *Tetrahymena* cell pellet. Cells were resuspended and homogenized in a chilled Waring Blender (Waring PBB212) at high speed for 40 s. The resulting lysate was subsequently centrifuged at 2,750 x g for 5 min in an SLA-3000 rotor to pellet macronuclei. The nuclear pellet was washed twice with medium B and then five times in MM medium (10mM Tris-HCl pH 7.8, 0.25M sucrose, 15mM $MgCl_2$, 0.1% w/v spermidine trihydrochloride, 1mM DTT, 1mM PMSF). Macronuclei were pelleted between wash steps by centrifuging at 2,500 x g for 5 min in an SLA-3000 rotor. Finally, the total number of washed macronuclei was counted with a hemocytometer using a Zeiss ID03 microscope. Nuclear proteins were extracted by vigorously resuspending the pellet in M Msalt buffer (10mM Tris-HCl pH 7.8, 0.25M sucrose, 15mM MgCl2, 350mM NaCl, 0.1% w/v spermidine trihydrochloride, 1mM DTT, 1mM PMSF). 1mL M Msalt buffer was added per 2.33 x 108 macronuclei. The viscous mixture was nutated for 45 min at 4°C, and then cleared at 175,000 x g for 30 min at 4°C in a SW 41 Ti rotor. Following this, the supernatant was dialyzed in a Slide-A-Lyzer 3.5K MWCO cassette (Thermo Fisher) overnight at 4°C against two changes of MMminus medium (10mM Tris-HCl pH 7.8, 15mM $MgCl_2$, 1mM DTT, 0.5mM PMSF). The dialysate was then centrifuged at 7,197 x g for 1 hr at 4"C to remove precipitates, and dialyzed overnight in a Slide-A-Lyzer 3.5K MWCO cassette (Thermo Fisher) at 4°C against two changes of MN3 buffer (30mM Tris-HCl pH 7.8, 1mM EDTA, 15mM NaCl, 20% v/v glycerol, 1mM DTT, 0.5mM PMSF). The final dialysate was cleared by centrifugation at 7,197 g for 1.5 hr at 4°C, flash frozen, and stored at -80°C. This nuclear extract was used for all subsequent biochemical fractionation and 6mA methylation assays.

Partial purification of MTA1c from nuclear extracts

[0178] Tetrahymena nuclear extracts were passed through a HiTrap O HP column (GE Healthcare) and eluted using a linear aradient of 15mM to 650mM NaCl in 30mM Tris-HCl pH 7.8, 1mM EDTA, 20% v/v glycerol, 1mM DTT, 0.5 mM PMSF, over 30 column volumes. Each fraction was assayed for DNA methyltransferase activity using radiolabeled SAM as described in the next section. The DNA methyltransferase activity eluted in two peaks, at ~60mM and ~365mM NaCl, termed the "low salt sample" and "high salt sample." Fractions corresponding to each peak

were pooled and passed through a HiTrap Heparin HP column (GE Healthcare). Bound proteins were eluted using a linear gradient of 60 mM to 1M NaCl (for the low salt sample) or 350mM to 1M NaCl (for the high salt sample) over 30 column volumes. Fractions with DNA methyltransferase activity were respectively pooled and dialyzed into 10mM sodium phosphate pH 6.8, 100mM NaCl, 10% v/v glycerol, 0.3mM $CaCl_2$, 0.5mM DTT (for the low salt sample); or 30mM Tris-HCl pH 7.8, 1mM EDTA, 200mM NaCl, 10% v/v glycerol, 1mM DTT, 0.2mM PMSF (for the high salt sample). The dialyzed low salt sample was passed through a Nuvia cPrime column (Bio-Rad) and eluted using a linear gradient of 100 mM to 1M NaCl in 50 mM sodium phosphate pH 6.8, 10% v/v glycerol, 0.5 mM DTT. Separately, the dialyzed high salt sample was fractionated using a Superdex 200 10/300 GL column (GE Healthcare) in 30mM Tris-HCl pH 7.8, 1mM EDTA, 200mM NaCl, 10% v/v glycerol, 1mM DTT. Fractions from the Nuvia cPrime and Superdex 200 columns were dialyzed into 30mM Tris-HCl pH 7.8, 1mM EDTA, 15mM NaCl, 20% v/v glycerol, 1mM DTT, 0.5mM PMSF and assayed for DNA methyltransferase activity. Those with qualitatively low, medium, and high activity were subjected to mass spectrometry to identify candidate methyltransferase proteins (Fig. 2D; Table 6). This experiment identified four proteins that co-purify with DNA methyltransferase activity - MTA1, MTA9, p1, and p2 - and are collectively termed as "MTA1c" in the present disclosure. All four proteins are necessary for 6mA methylation *in vitro*.

Recombinant expression of MTA1, MTA9, p1, and p2 proteins

[0179] Full length MTA1, MTA9, p1, and p2 open reading frames were codon-optimized for bacterial expression and cloned into a pET-His6-SUMO vector using ligation independent cloning. Protein sequences are listed in Table 3. The vector was a gift from Scott Gradia (Addgene plasmid #29659; http://addgene.org/29659; RRID: Addgene 29659). Mutations in the MTA1 open reading frame was introduced using the OS® Site-Directed Mutagenesis Kit (New England Biolabs). For recombinant expression, pET-His6-SUMO-MTA1 (wild-type and mutant) was transformed into SHuffle T7 competent E. co/i (New England Biolabs); pET-His6-SUMO-MTA9 was transformed into Lemo (DE3) competent E. coli (New England Biolabs); pET-His6-SUMO-p1 and pET-His6-SUMO-p2 were transformed into BL21(DE3) competent E. coli (New England Biolabs). IPTG induction was performed at 16'C overnight. Induced cells were resuspended in 25ml of lysis buffer B (50mM Tris pH 7.8, 300mM

NaCl, 5% v/v glycerol, 10mM imidazole, 5mM BME, 1mM PMSF, 0.5x ProBlock Gold Bacterial protease inhibitor cocktail [GoldBio]). The cells were sonicated at 35% amplitude for a total of 4 minutes, with a 10 s off, 10 s cycle using a Model 505 Sonic Dismembrator (Fisherbrand). Lysates were cleared by centrifugation at 30,000 g for 30 min at 4°C, mixed with pre-washed Ni-NTA agarose (Invitrogen), and nutated for 45min at 4°C. The resin was subsequently washed with lysis buffer and eluted in 50mM Tris pH 7.8, 300mM NaCl, 5%v/v glycerol, 400mM glycerol, 5mM BME, lx ProBlock Gold bacterial protease inhibitor cocktail [GoldBio]). Eluates were dialyzed into lysis buffer B and then digested with TEV protease (gift from S.H. Sternberg) at 4°C overnight. The resulting mixture was passed through a fresh batch of Ni-NTA agarose (Invitrogen) to remove cleaved affinity tags. The flow-through containing each recombinant protein was flash frozen and used for all downstream methyltransferase assays.

Methyltransferase assays

Generation of DNA and RNA substrates

[0180] A 954bp dsDNA PCR product was used in all assays involving *Tetrahymena* nuclear extract. This substrate was amplified by PCR from *Tetrahymena thermophila* strain SB210 macronuclear SB210 genomic DNA using PCR primers metGATC F2 and metGATC_R2 (Table 2). The resulting product was purified using Ampure XP beads (Beckman Coulter). This 954bp region of the genome contains a high level of 6mA in vivo. Thus, the underlying DNA sequence may be intrinsically amenable to methylation by *Tetrahymena* MTA1. Note that the amplified 954bp product is devoid of DNA methylation as unmodified dNTPs were used for PCR. Separately, a 350bp dsDNA PCR product was used in all assays involving recombinant MTA1, MTA9, p1 and p2. This sequence lacks 5'-NATC-3' motifs, and was used to reduce background DNA methylation from contaminating *Dam* methyltransferase in recombinant protein preparations. The 350bp dsDNA PCR product was amplified from *Tetrahymena thermophila* strain SB210 macronuclear SB210 genomic DNA using the PCR primers noGATC2 F and noGATC2_R (Table 2), and purified using Ampure XP beads (Beckman Coulter).

[0181] For short DNA substrates (< 50bp), oligonucleotides were purchased from Integrated DNA Technologies and either directly used as ssDNA, or annealed with its complementary sequence to obtain dsDNA. To prepare hemimethylated

27bp dsDNA in Fig. 2G, either strand was methylated using *EcoGII* methyltransferase (New England BioLabs) before annealing with the complementary sequence.

[0182] To generate ~350nt ssRNA and ~350bp dsRNA, the aforementioned 350bp dsDNA was first PCR-amplified using primers containing T7 overhangs (primer pairs T7noGATC2_F2 / noGATC2_R and T7noGATC2_F2 / T7noGATC2_R2 respectively; see Table 2 for primer sequences). Each PCR product was used as a template for *in vitro* transcription using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). The synthesized RNA was rigorously treated with DNase (ThermoFisher) purified using acid phenol:chloroform extraction, followed by two rounds of chloroform extraction. Each sample was subsequently ethanol precipitated and resuspended in water for use in methyltransferase assays.

Radioactive methyltransferase assay

[0183] For experiments involving nuclear extract, 2.18µg of 954bp dsDNA substrate was mixed with 4-8µl nuclear extract and 0.64µM 3H-labeled S-adenosyl-L-methionine ([^{3}H]SAM) in 33mM Tris-HCI pH 7.5. 6mM EDTA. 4.3mM BME. in a 15µl reaction volume. For experiments involving recombinant MTA1c protein components (ie. MTA1, MTA9, p1, and/or p2), ~3µM oligonucleotide ssDNA / annealed dsDNA is used. Alternatively, 1.3µg of 350bp dsDNA substrate (or an equimolar amount ~350nt ssRNA, or ~350bp dsDNA) was used in place of DNA oligonucleotide substrates. ssRNA was heated at 90°C for 2 min and snap cooled to minimize secondary structures before mixing with other components of the methyltransferase assay. All samples were incubated overnight at 37°C, and subsequently spotted onto 1cm x 1cm squares of Hybond-XL membrane (GE Healthcare). Membranes were then washed thrice with 0.2M ammonium bicarbonate, once with distilled water, twice with 100% ethanol, and finally air-dried for 1 hr. Each membrane was immersed in 5mL Ultima Gold (PerkinElmer) and used for scintillation counting on a TriCarb 2910 TR (Perkin Elmer).

Non-radioactive methyltransferase assay

[0184] For assays involving nuclear extract: 5.5pg of 954bp DNA substrate was mixed with 20 nuclear extract and 0.2mM S-adenosyl-Lmethionine (NEB) in 33mM Tris-HCI pH 7.5, 6mM EDTA, 4.3mM BME in a 15µl reaction volume. For

assays involving recombinant MTA1c protein components (ie. MTA1, MTA9, p1, and/or p2), 2.6µg of 350bp DNA substrate was mixed with 540nM MTA1, 90nM MTA9, 1.5µM p1, 1.0µM p2 proteins. The band of expected size in each recombinant protein preparation was compared against a series of BSA standards to calculate protein concentration. All methylation reactions were incubated at 37°C overnight, then purified using a MinElute purification kit (QIAGEN), denatured at 95°C for 10 min, and snap cooled in an ice water bath. Samples were spotted on a Hybond N+ membrane (GE Healthcare), air-dried for 5 min and UV-cross-linked with 120,000µJ / cm^2 exposure using an Ultra-Lum UVC-515 Ultraviolet Multilinker. The cross-linked membrane was blocked in 5% milk in TBST (containing 0.1% v/v Tween) and incubated with 1:1,000 anti-N6-methyladenosine antibody (Synaptic Systems) at 4°C overnight. The membrane was then washed three times with TBST, incubated with 1:3,000 Goat anti-rabbit HRP antibody (Bio-Rad) at room temperature for 1 hr, washed another three times with 1 x TBST, and developed using Amersham ECL Western Blotting Detection Kit (GE Healthcare). This dot blot assay was used to measure 6mA levels in Figs. 2F, 3B, 5C, and 10C.

Quantitative mass spectrometry analysis of dA and 6mA

[0185] 10.5µg *Oxytricha* or *Tetrahymena* macronuclear genomic DNA was first digested to nucleosides by mixing with 14µl DNA degradase plus enzyme (Zymo Research) in a 262.5µl reaction volume. Samples were incubated at 37°C overnight, then 70°C for 20 min to deactivate the enzyme.

[0186] The internal nucleoside standards $^{15}N_5$-dA and D_3-6mA were used to quantify endogenous dA and 6mA levels in ciliate DNA. $^{15}N_5$-dA was purchased from Cambridge Isotope Laboratories, while D_3-6mA was synthesized as described in the following section. Nucleoside samples were spiked with 1 ng/µl $^{15}N_5$-dA and 200 pg/µl D_3-6mA in an autosampler vial. Samples were loaded onto a 1mm x 100mm C18 column (Ace C18-AR, Mac-Mod) using a Shimadzu HPLC system and PAL auto-sampler (20µl / injection) at a flow rate of 70µl / min. The column was connected inline to an electrospray source couple to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher). Caffeine (2 pmol/µl in 50% Acetonitrile with 0.1 % FA) was injected as a lock mass through a tee at the column outlet using a syringe pump at 0.5µl/min (Harvard PHD 2000). Chromatographic separation was achieved with a linear gradient from 10% to 99% B (A: 0.1% Formic Acid, B: 0.1% Formic Acid in

Acetonitrile) in 5 min, followed by 5 min wash at 100% B and equilibration for 10 min with 1% B (total 20 min program). Electrospray ionization was achieved using a spray voltage of 4.50 kV aided by sheath gas (Nitrogen) flow rate of 18 (arbitrary units) and auxiliary gas (Nitrogen) flow rate of 2 (arbitrary units). Full scan MS data were acquired in the Orbitrap at a resolution of 60,000 in profile mode from the m/z range of 190-290. A parent mass list was utilized to acquire MS/MS spectra at a resolution of 7500 in the Orbitrap. LC-MS data were manually interpreted in Xcalibur's Qual browser (Thermo, Version 2.1) to visualize nucleoside mass spectra and to generate extracted ion chromatograms by using the theoretical [M+H] within a range of ± 2ppm. Peak areas were extracted in Skyline (Ver. 3.5.0.9319).

Synthesis of D_3-6mA nucleoside

[0187] 2'-Deoxyadenosine and CD3I were purchased from Sigma Aldrich. Flash chromatography was performed on a Biotage Isolera using silica columns (Biotage SNAP Ultra, HP-Sphere 25pm). Semi-preparative RP-HPLC was performed on a Hewlett-Packard 1200 series instrument equipped with a Waters XBridge BEH C18 column (5 µm, 10 x 250 mm) at a flow rate of 4mL/min, eluting using A (0.1% formic acid in H_2O) and B (0.1% formic acid in 9:1 MeCN/H_2O). ^{1}H NMR spectra were recorded on a Bruker UltraShield Plus 500 MHz instrument. Data for ^{1}H NMR are reported as follows: chemical shift (8 ppm), multiplicity (s = singlet, br = broad signal, d = doublet, dd = doublet of doublets) and coupling constant (Hz) where possible. ^{13}C NMR spectra were recorded on a Bruker UltraShield Plus 500 MHz.

[0188] D_3-6mA (2'Deoxy-6-[*D*3]-methyladenosine) were synthesized and purified according to (Schiffers et al., 2017). After an initial purification by flash column chromatography, the methylated compounds were further purified by semipreparative RP-HPLC (linear gradient of 0% to 20% B over 30 min) affording the desired compounds in 14% and 10% yields respectively after lyophilization.

2'Deoxy-6-[D3]-methyladenosine

[0189] ^{1}H NMR (500 MHz, D_2O) δ 7.98 (s, 1 H), 7.77 (s, 1H), 6.17 (m, 1 H), 4.54 (m, 1H), 4.10 (m, 1 H), 3.79 (dd, J = 12.7, 3.2 Hz, 1H), 3.71 (dd, J = 12.7, 4.3 Hz, 1 H), 2.60 (m, 1 H), 2.44 (ddd, J = 14.0, 6.3, 3.3 Hz, 1 H).

[0190] ^{13}C NMR (126 MHz, D_2O) δ 154.0, 151.5, 146.1, 138.9, 118.4, 87.3, 84.3, 71.1, 61.6, 39.2, 26.4 ppm. (Peak at 26.4 ppm appears as a broad signal. C-D coupling is not resolved).

[0191] HR-MS (ESI+): m/z calculated for $[C_{11}H_{13}D_3N_5O_3]^+$ ([M+Hr): 269.1436. found 269.1421.

Mass spectrometry analysis of proteins in *Tetrahymena* nuclear extracts

[0192] Samples where topped up to 200µl with 50mM ammonium bicarbonate pH 8. TCEP was added to 5mM final concentration and left to incubate at 60°C for 10 min. 15mM chloroacetamide was then added and left to incubate in the dark at room temperature for 30 min. 1µg of Trypsin Gold (Promega) was added to each sample and incubated end-over-end at 37°C for 16 hr. An additional 0.25µg of Trypsin Gold was added and incubated end-over-end at 37°C for 3 hr. Samples were acidified by adding TFA to 0.2% final concentration, and desalted using SDB stage-tips (Rappsilber et al., 2007). Samples were dried completely in a speedvac and resuspended in 20µl of 0.1% formic acid pH 3.5µl was injected per run using an Easy-nLC 1200 UPLC system. Samples were loaded directly onto a 45cm long 75pm inner diameter nano capillary column packed with 1.9µm C18-AQ (Dr. Maisch, Germany) mated to metal emitter in-line with an Orbitrap Fusion Lumos (Thermo Scientific, USA). The mass spectrometer was operated in data dependent mode with the 120,000 resolution MS1 scan (AGC 4e5, Max IT 50ms, 400-1500 m/z) in the Orbitrap followed by up to 20 MS/MS scans with CID fragmentation in the ion trap. Dynamic exclusion list was invoked to exclude previously sequenced peptides for 60 s if sequenced within the last 30 s, and maximum cycle time of 3 s was used. Peptides were isolated for fragmentation using the quadrupole (1.6 Da window). Ns was utilized. Ion-trap was operated in Rapid mode with AGC 2e3, maximum IT of 300 msec and minimum of 5000 ions.

[0193] Raw files were searched using Byonic (Bern et al., 2012) and Sequest HT algorithms (Eng et al., 1994) within the Proteome Discoverer 2.1 suite (Thermo Scientific, USA). 1Oppm MS1 and 0.4Da MS2 mass tolerances were specified. Caramidomethylation of cysteine was used as fixed modification, while oxidation of methionine, pyro-Glu from Gln and deamidation of asparagine were specified as dynamic modifications. Trypsin digestion with maximum of 2 missed cleavages were allowed. Files were searched against the *Tetrahymena themophila* macronuclear reference proteome (June 2014 build), supplemented with common contaminants (27,099 total entries).

[0194] Scaffold (version Scaffold 4.8.7, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 93.0% probability. Peptide Probabilities from Sequest and Byonic were assigned by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability to achieve an FDR less than 1.0% and contained at least 3 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Generation of mta1 mutant lines

[0195] A frameshift mutation in the MTA1 gene was created by inserting a small non-coding DNA segment immediately downstream of the MTA1 start codon (Figs. 3A and 12H). This non-coding DNA segment belongs to a class of genetic elements that are normally eliminated during the sexual cycle (Chen et al., 2014). When ssRNA homologous to such DNA segments is injected into *Oxytricha* cells undergoing sexual development, the DNA is erroneously retained (Khurana et al., 2018). This results in disruption of the MTA1 open reading frame. The ectopic DNA segment is propagated through subsequent cell divisions after completion of the sexual cycle. RNaseq analysis confirmed the presence of the ectopic insertion in mta1 mutant transcripts but not wild-type controls (Fig. 12H).

[0196] ssRNA was generated by in vitro transcription using a Hi-Scribe T7 High Yield RNA Synthesis Kit (New England Biolabs). The DNA template for *in vitro* transcription consists of the ectopic DNA segment flanked by 100-200bp cognate MTA1 sequence. Following DNase treatment, ssRNA was acid-phenol:chloroform extracted and ethanol precipitated. After precipitation, ssRNA was resuspended in nuclease-free water (Ambion) to a final concentration of 1 to 3 mg/mL for injection.

ssRNA microinjections

[0197] *Oxytricha* cells were mated by mixing 3mL of each mating type, JRB310 and JRB510, along with 6mL of fresh Pringsheim media. At 10 to 12 hr post mixing, pairs were isolated and placed in Volvic water with 0.2% bovine serum albumin (Jackson ImmunoResearch Laboratories) (Fang et al., 2012). ssRNA constructs were injected into the macronuclei of paired cells under a light microscope

as previously described with DNA constructs (Nowacki et al., 2008). After injection, cells were pooled in Volvic water. At 60 to 72 hr post mixing, the pooled cells were singled out to grow clonal injected cell lines. As clonal population size grew, lines were transferred to 10 cm Petri dishes and grown in Pringsheim media. Only water from the "Volvic" brand has been empirically tested in our laboratory to support *Oxytricha* growth. Similar products from other vendors have not been tested.

Survival analysis of Oxytricha mtal mutants

[0198] This experiment was performed in Fig. 7D. Wild-type or mutant *Oxytricha* cells were mixed at 0 hr to induce mating. Since not all cells enter the sexual cycle, mated cells are separated from unmated vegetative cells at 15 hr and transferred into a separate dish. The cells are allowed to rest for 12 hr to account for cell death during transfer. The number of surviving mated cells is counted from 27 hr onward. The total cell number at each time point is normalized to 27 hr data to obtain the percentage survival. An increase in survival at 108 hr is observed in wild-type samples because the cells have completed mating and reverted to the vegetative state, where they can proliferate and increase in number.

QUANTIFICATION AND STATISTICAL ANALYSIS

[0199] All statistical tests were performed in Python (v2.7.10) or R (v3.2.5), and described in the respective Figure and Table legends.

DATA AND SOFTWARE AVAILABILITY

[0200] *Oxytricha* SMRT-seq data are deposited in SRA under the accession numbers SRA: SRX2335608 and SRX2335607, and GEO: GSE94421. *Tetrahymena* SMRT-seq and all Oxytricha Illumina data are deposited in NCBI GEO under accession number GEO: GSE94421.

Table 1. Protein sequences for phylogenetic tree construction.

Protein sequences for phylogenetic analysis of MT-A70 proteins (including MTA1 and MTA9)

>NP_495127.1 DNA N6-methyl methyltransferase [Caenorhabditis elegans]
MDTEFAILDEEKYYDSVFKELNLKTRSELYEISSKFMPDSQFEAIKRRGISNRKRKIKETSENSNRMEQM
ALKIKNVGTELKIFKKKSILDNNLKSRKAAETALNVSIPSASASSEQIIEFQKSESLSNLMSNGMINNWV
RCSGDKPGIIENSDGTKFYIPPKSTFHVGDVKDIEQYSRAHDLLFDLIIADPPWFSKSVKRKRTYQMDEE
VLDCLDIPVILTHDALIAFWITNRIGIEEEMIERFDKWGMEVVATWKLLKITTQGDPVYDFDNQKHKVPF
ESLMLAKKKDSMRKFELPENFVFASVPMSVHSHKPPLLDLLRHFGIEFTEPLELFARSLLPSTHSVGYEP
FLLQSEHVFTRNISL
>NP_564080.1 Methyltransferase MT-A70 family protein [Arabidopsis thaliana]
MAKTDKLAQFLDSGIYESDEFNWFFLDTVRITNRSYTRFKVSPSAYYSRFFNSKQLNQHSSESNPKKRKR
KQKNSSFHLPSVGEQASNLRHQEARLFLSKAHESFLKEIELLSLTKGLSDDNDDDDSSLLNKCCDDEVSF
IELGGVWQAPFYEITLSFNLHCDNEGESCNEQRVFQVFNNLVVNEIGEEVEAEFSNRRYIMPRNSCFYMS
DLHHIRNLVPAKSEEGYNLIVIDPPWENASAHQKSKYPTLPNQYFLSLPIKQLAHAEGALVALWVTNREK
LLSFVEKELFPAWGIKYVATMYWLKVKPDGTLICDLDLVHHKPYEYLLLGYHFTELAGSEKRSDFKLLDK
NQIIMSIPGDFSRKPPIGDILLKHTPGSQPARCLELFAREMAAGWTSWGNEPLHFQDSRYFLKV
>QRY94237.1 MT-A70-domain-containing protein [Syncephalastrum racemosum]
MIVASSDTCDIVDCEAAFGIDGTVRLRPGDFSLGTPYFTSRLGQKRPRPDDDTLDNTPSDTIHAIVQQLP
VMAPDYWHDRPMEAVVMNAHVHFPSLVSLAEASLRFDPDNDEDEDNRQILRPDMALESLQVFYRHFEHPK
DSPILIRVQDAYYWIFPRTAFMMGSLENIHLPTLGKFDCIVMDPPWPNKSVRRSAHYETQEDIYDLFAIP
LPQLAGPNCLVAVWVTNKPKFIRFVQKLFAAWDVEPLTTWYWLKVTTHGEPVCPIDSPHRKPYEHLILGR
KRPVKININDPPALPRVLVSVPSKHHSRKPPLNDILMRYLPSDARRLELFARCLTPGWTSWGNECLKFQH
VDYFYDTNEAMEEGKQK
>ORX58127.1 MT-A70-domain-containing protein [Hesseltinella vesiculosa]
MANAARRFAQQDELPLDVSQDLQDLPLLDLFNRKVINDSDQCSSLHVASFGQYLVPRHTKFVMSDLDNID
LLRSENDVFDLIVMDPPWPNKSVHRSTDYETQDIYDLFHLPIKSLIKNQGLVAVWVTNKPKYRRFILDKL
FKAWQMTCVGEWLWLKVTSSGEPVFPLDSPHRKPYEQLILGRYQPDDTSPTLPNPPQQHVLISVPSIRHS
RKPPLGEVLADFLPKQPACLELFARCLTPGWTSWGNECLKFQHESYFISNDTPHSPSAS
>ORZ15132.1 MT-A70-like protein, partial [Absidia repens]
YDLVVMDPPWPNKSVHRSSHYETQDIYDLYQIPLTSLVHKNSLVAVWITNKPKYRRFVMDKLFKSWHVDC
VAEWTWLKVTNDGEPVFPLNSTHRKPYEQLIIGRYNGGSGGGNDNNDSIQEESEVKPIPYQHSIVSVPSK
RHSRKPPLQDLLQPYLPAKPRCLELFARCLTPGWSSWGNECLKFQNEYYYTRIENPLHIDRSDV
>XP_021879935.1 MT-A70-domain-containing protein [Lobosporangium transversale]
MLHESTVSVLDRLILISHISLQTYLLAKDREGFDIIVMDPPWQNASVDRMSHYRTMDLYELFKIPIPDLL
KANGSNVGGIVAVWITNKAKVKRVVVEKLFPAWGLDLVAHWFWLKVTTKGEPVLSLSNSHRRAYEGVLIG
RQRQGSKLSNKTMHETSASNPVNRLLVSIPAQHSRKPSLNALIEEEFFTSKLESRADRDRNAYVDSEALV
KKPLYRLELFARNLEEGVLSWGNEPLRYQYCGRGASNSQVVQDGYLIPCPIQSELVSQ
>XP_889178.3 methyltransferase-like protein 4 isoform X1 [Danio rerio]
MSVVCCNSWGWLLDSSSHIDKDFQRCVCYNEANGLEENTHFTCCFKRQYFNILMPHMQQSTAMSGFPLDS
GKHDSAEHEKIELQTRKKRKRKHHDLNTGEIEANIYHDKVRSVVLEGSRALLEAGRQCGYFTEALTESQT
ISTPSESTSAHECQLAAFCDLAKQLPLSEESPVHTLSRDGQNPALDLFSSITENPFDCACEITFMRERYL
LPPRCRFLLSDVTRMDPLVNSGDKFDLIVLDPPWENKSVKRSNRYSSLPSSQLKKLPVPALAAPGGLVVT
WVTNRAKHRRFVREELYPHWAVEVLAEWLWVKVTRSGEFVFPLDSQHKKPYEVLVLGRCRSTSDHTDRCS
AVNELPDQRLLVSVPSTLHSHKPSLAAVLKPYIRREPRCLELFARSLQSDWSCWGNEVLKFQHCSYFSRH
TDQEPTSDTLQRTHSHLQSTGLLETPETAR
>NP_073751.3 methyltransferase-like protein 4 isoform 1 [Homo sapiens]
MSVVHQLSAGWLLDHLSFINKINYQLHQHHEPCCRKKEFTTSVHFESLQMDSVSSSGVCAAFIASDSSTK
PENDDGGNYEMFTRKFVFRPELFDVTKPYITPAVHKECQQSNEKEDLMNGVKKEISISIIGKKRKRCVVF
NQGELDAMEYHTKIRELILDGSLQLIQEGLKSGFLYPLFEKQDKGSKPITLPLDACSLSELCEMAKHLPS
LNEMEHQTLQLVEEDTSVTEQDLFLRVVENNSSFTKVITLMGQKYLLPPKSSFLLSDISCMQPLLNYRKT
FDVIVIDPPWQNKSVKRSNRYSYLSPLQIQQIPIPKLAAPNCLLVTWVTNRQKHLRFIKEELYPSWSVEV
VAEWHWVKITNSGEFVFPLDSPHKKPYEGLILGRVQEKTALPLRNADVNVLPIPDHKLIVSVPCTLHSHK
PPLAEVLKDYIKPDGEYLELFARNLQPGWTSWGNEVLKFQHVDYFIAVESGS
>XP_020951799.1 methyltransferase-like protein 4 isoform X1 [Sus scrofa]
MSVVHQLSSGWLLDHLSFINKISYELHQHHEPCCSKNEPTSVHLDSLHKDSVFSFGASPAFIASSSKPEN
DDGGNREMSMQKYVFRSELFDVTKPYITSAIHKECQQSNEKEDLANDVKKEASISIKRKKRKRCVVFNQG
ELDAMEYHTKIRGLILDGSSQLIQEGLKSGFLHPLSEKCDKCSKPVTLPLDTCSLSELCEMAKHVPSLNE
MELQTLQLMEDDISVTEQDLFSRIVENNSSFTKMITLMGQKYLLPPKSSFLLSDISCIYPLLNCRKTYDV
IVIDPPWQNKSVKRSNRYSYLSPLQIKQIPIPKLAAPNCLVVTWVTNRQKHLRFVKEELYPSWSVEIVAE
WHWVKITNSGEFVFPIDSPHKKPYEVLVLGRVRERAALLLSRNAEVKELSIPDHKLIVSVPCILHSHKPP
LAEVLKDYIKPEGEYLELFARNLQPGWTSWGNEVLKFQHMDYFVALESRS
>XP_011245012.1 PREDICTED: methyltransferase-like protein 4 isoform X2 [Mus musculus]
MSVVHHLPPGWLLDHLSFINKVNYQLCQHQESFCSKNNPTSSVYMDSLQLDPGSPFGAPAMCFAPDFTTV
SGNDDEGSCEVITEKYVFRSELFNVTKPYIVPAVHKERQQSNKNENLVTDYKQEVSVSVGKKRKRCIAFN

QGELDAMEYHTKIRELILDGSSKLIQEGLRSGFLYPLVEKQDGSSGCITLPLDACNLSELCEMAKHLPSL
NEMELQTLQLMGDDVSVIELDLSSQIIENNSSFSKMITLMGQKYLLPPQSSFLLSDISCMQPLLNCGKTF
DAIVIDPPWENKSVKRSNRYSSLSPQQIKRMPIPKLAAADCLIVTWVTNRQKHLCFVKEELYPSWSVEVV
AEWYWVKITNSGEFVFPLDSPHKKPYECLVLGRVKEKTPLALRNPDVRIPPVPDQKLIVSVPCVLHSHKP
PLTGYLNSSFATLIPRVSNNMEYCRVVRTAFIA
>XP_018079135.1 PREDICTED: methyltransferase-like protein 4 [Xenopus laevis]
MSVVCETSAGWLVDELSLLRKWYQHSTSCQDAAHKKQLYDIKEDLFLILRPHIPVQSTPAPLPLCPETN
PGTINQRKKRKRSCAFNQGELDAMEYHKKIIDFIMEGTQPLLQEGFKRLFLRPVLVNDDDHSQTEPRLCN
NPCQLAELCNMAKCMPLLNPGEHAVQVLERGIYLPQETNVLSCITENKSECPEVIQFMGEKYIIPPKSTF
LMSDVSCMEPLLHYKRYNIIVMDPPWENKSVKRSKRYSSLSPNEIQQLPVPVLAAPDCLVITWVTNKQKH
LRFVKEDLYPHWSVKTLGEWHWVKITRSGEFVFPLDSTHKKPYEVLIIGRFKGAGNSTARKSEICLPPIP
ERKLIVSVPCKLHSHKPPLSEILKEYVKPDLECLELFARNLQPGWTSWGNEVLKFQHIDYFTPVDVED
>NP_650573.1 uncharacterized protein Dmel_CG14906 [Drosophila melanogaster]
MLKLQKKTEDSKFAVFLDHKTLINEAYDEFKLKSELFQFHAKKTDKGIEEDKTRKRKRKAGVEDASSLED
LHLVNEYLELLSKPVEPEDSSPMKRHWEDGYNVPQLHGANESGRMQRFLRVDGSRGVYLIPNQSRFFNHN
VDNLPALLHQLLPAYDLIVLDPPWRNKYIRRLKRAKPELGYSMLSNEQLSHIPLSKLTHPRSLVAIWCTN
STLHQLALEQQLLPSWNLRLLHKLRWYKLSTDHELIAPPQSDLTQKQPYEMLYVACRSDASENYGKDIQQ
TELIFSVPSIVHSHKPPLLSWLREHLLLDKDQLEPNCLELFARYLHPHFTSIGLEVLKLMDERLYEVRKV
EHCNQEEVN

>tr|A8J2E1|A8J2E1_CHLRE Predicted protein OS=Chlamydomonas reinhardtii OX=3055 GN=CHLREDRAFT_174824 PE=3 SV=1
MATLPGAAAAAPGANAEVGVPEPSLEPQDALQQRIALAEGLLALNEADAMQAWQQLPREA
LLEQVAKYRGAVRDMASALRSSTLPGGVPPHCVPIHANVTTFDWPSLYSHAQFDVIMMDP
PWQLATANPTRGVALGYSQLNDDHISRLPVPQLQRQGGYLFVWVINAKYKWTLDLFDRWG
YRLVDEVVWVKMTVNRRLAKSHGYYLQHAKEVCLVAKRGNPPVPPGCEGGVGSDIIFSER
RGQSQKPEEIYHLIEQLVPNGRYLEIFARKNNLRNYWVSIGNEVTGTGLPDEDMQALRDL
HHIPGAVYGKNAPHLVSKLFLYAPNSSREEG
>XP_021880122.1 MT-A70-domain-containing protein [Lobosporangium transversale]
MLDQINIDIEQLEASLDIDEGKAHSNNASGTGCLIGTGTSSGNASNGAGVADEDLEEEVDDLEEFEAPEW
CVPIKANVMTYDWDSLAAECQFDVILMDPPWQLATHAPTRGVAIAYQQLPDICIEELPVPKLSSNGFIFI
WVINNKYAKAFDLMRRWGYSYVDDITWVKQTVNRRMAKGHGYYLQHAKETCLVGKKGEDPPGCRHSIGSD
VIFSERRGQSQKPEELYEIIEELVPNGRYLEIFGRKNNLRDYWVTVGNEL
>ORX89627.1 MT-A70-domain-containing protein [Linderina pennispora]
MDVDSSSPAVVLQALRQREQKIRSRILVLEQEISDLEKRCGVEGSGDAANKVTEADLEEFKAPEWSVPIR
ANVMNFDWEKLAQACQFDVILMDPPWQLASQAPTRGVAIAYQQLPDVCIESLPIDLLQTSGFIFIWVINN
KYTKAFQLMKQWGYKYVDDIAWVKQTVNRRMAKGHGYYLQHAKETCLVGKKGPDPPNLRRSVASDVIFSE
RRGQSQKPEELYEIIEQLVPGGRYLEIFGRKNNLRDYWVTVGNEL
>ORX98979.1 allantoinase [Basidiobolus meristosporus CBS 931.73]
MSAIIFTGNRVLFDSTSKVEPATIHVDPWTGRIVKITNKRSTKADFPGIEDKDFVDAGDDLIMPGVIDAH
VHLNEPGRTDWEGFDTATRAAAAGGLTTVIDMPLNSIPPTTTLENLNTKKEAAKPQAWVDVGFYGGVIPG
NADQLRPMIAAGVCGFKCFLIESGVDEFPCVNEEEVRKAFAEFDGTDNVFMFHAEMECDDHSHETAAPQS
TDPSAYQTFLQSRPHALEVKAIEMIIRVCKDFPNVRAHIVHLSSAEALPMIRKAKAEGVKLTVETCYHYL
TLNAEDIINGATHFKCCPPIREGSNRELLWEALLDGTIDYVVSDHSPCTPELKRFDSGDFTAAWGGISSL
QFGLSLLWTEAKRRGCTLQDLTRWLSQNTARHAGILNRKGRLQIGSDADIVIWSPEETFVVDKKMIHFKN
KVTPYENMTLHGAVKKTFVRGRNVYDKSTAQLFSAKPLGNLLARFQVYSNPITAMPSYAQPPSSDNGDFE
EESEDYIESDEVDEDLRELLAKETSLRLRIDSLKEEILKLEREQRGETDGSKNEGEGGEEEIDLEEFEAP
EWCVPIKANVMTFEWKRLAEAAQFDVILMDPPWQLATHAPTRGVAIGYQQLPDVCIEELPIPLLQKNGFI
FIWVINNKYVKAFELMAKWGYRYVDDITWVKQTVNRRMAKGHGYYLQHAKETCLIGKKGEDPPNCRHSVC
SDVIFSERRGQSQKPEELYEMIEQLVPNGKYLEIFGRKNNLRDYWVTIGNEL
>OR200623.1 MT-A70-domain-containing protein [Syncephalastrum racemosum]
MSSREESPSSVSGFDLDTIDESTVTDTTLKNLLRREIELQLQIDALQTEILQIEESTAAGKNNKNDEELD
PQDLEEFEAPEWCVPIKANVMTFDWEALASEVQFDVIVADPPWQLATHAPTRGVAIGYQQLPDVCIEEIP
IQKLQKNGFIFIWVINNKYAKAFELMERWGYHYVDDITWVKQTVNRRMAKGHGYYLQHAKETCLVGKKGE
DPPNCRHSVGSDVIFSERRGQSQKPEELYELIEELVPNGKYLEIFGRKNNLRDYWVTVGNEL
>ORZ06213.1 MT-A70-domain-containing protein [Absidia repens]
MTSDTSAMTADVLNRKRKRSPAMNGDDLSNNSDEADNNTTTGTTTSVDSNENDYQEQDREPILRLPRLND
AKLLEEVVDDVDYEDQPERYDFDFKKLWLQERGLMERIDGLLKDIARLTDFKGHYRDMVIPSDDEDDLDD
EDSKAQYDAPEWCVPIKANVMTFDWESLGKEVQFDVIMADPPWQLATHAPTRGVAISYQQLPDVCIEDLP
LEKLQTNGFLFIWVINNKYAKAFEMMEKWGYKYVDDITWVKQTVNRRMAKGHGYYLQHAKETCLVGVKGT
LPPYCRRSVGSDVIYSERRGQSQKPEQIYELIEEMMPGGKYLEIFGRKNNLRDYWITVGNEL
>ORX43344.1 MT-A70-domain-containing protein [Hesseltinella vesiculosa]
MASESNISRESSPASISSTNSESGIENVQSLTDEDLKQLILKEMNLKEHIEQLQRKISKLTANDLSTNQD
SSDADDDLLNGDETMDDDSSSGSDSEVSGNEDIASVKSSPHAADKSESESESESDEGSSEDGNDEEDEFE

APKWCVPIKANVMTFDWEKLASETQFDVIVADPPWQLATHAPTRGVAIAYQQLPDVCIEDLPIEKLQTNG
FIFIWVINNKYAKAFELMEKWGYTYVDDITWVKQTVNRRMAKGHGYYLQHAKETCLVGKKGVDPPSCRHS
VGSDIIFSERRGQSQKPEELYELIEELVPNGKYLEIFGRKNNLRDYWVTVGNEL
>ORX52920.1 MT-A70 protein [Piromyces finnis]
MMIVANEIDYEEFTAPEWCIPIKANVIDFEWDKLASECQFDAILMDPPWQLATHAPTRGVAIAYQQLPDQ
FIEELPIEKLQKNGFIFIWVINNKYVKAFELMKKWGYTFVDDITWVKQTVNRRMAKGHGYYLQHAKETCL
VGKKGEDPVGCKHSISSDVIYSVRRGQSQKPEELYEMIEELIPNGKYLEIFGRKNNLRDYWVTIGNEL
>ORX86973.1 MT-A70-domain-containing protein [Anaeromyces robustus]
MDEKEVENSVLQSSNIEKSNATTSNMDVDETSNNETSTAIIKSEDGANSYDDFLKLDFTPEEEKDEVLKK
LIERETELKLKIEKEIEGIKNLELKGFSALTQKDEDVQDIDYEEFTAPEWCIPIKANVIDFEWDKLASEC
QFDAILMDPPWQLATHAPTRGVAIAYQQLPDQFIEELPIEKLQKNGFIFIWVINNKYVKAFELMKKWGYT
FVDDITWVKQTVNRRMAKGHGYYLQHAKETCLVGKKGDDPVGCRHKISSDVIYSVRRGQSQKPEELYEMI
EELIPNGKYLEIFGRKNNLRDYWVTIGNEL
>XP_001032074.3 MT-a70 family protein [Tetrahymena thermophila SB210]
MKKEQQFLIFKKSLIIAQKRKEINIKQLKQQFKNFLFVQIFSIIKLKLQDIIIKFKMSKAVNKKGLRPRK
SDSILDHIKNKLDQEFLEDNENGEQSDEDYDQKSLNKAKKPYKKRQTQNGSELVISQQKTKAKASANNKK
SAKNSQKLDEEEKIVEEEDLSPQKNGAVSEDDQQQEASTQEDDYLDRLPKSKKGLQGLLQDIEKRILHYK
QLFFKEQNEIANGKRSMVPDNSIPICSDVTKLNFQALIDAQMRHAGKMFDVIMMDPPWQLSSSQPSRGVA
IAYDSLSDEKIQNMPIQSLQQDGFIFVWAINAKYRVTIKMIENWGYKLVDEITWVKKTVNGKIAKGHGFY
LQHAKESCLIGVKGDVDNGRFKKNIASDVIFSERRGQSQKPEEIYQYINQLCPNGNYLEIFARRNNLHDN
WVSIGNEL
>EJY88228.1 MT-A70 family protein [Oxytricha trifallax]
MNQSSQDITTQKSSNGFNPQTQPETLIQVIRKESTFIFKYRKNPYYVPPPISSQTSPNLEVETSNDLNQM
SDYEGQIPNNYEINRNSTQFTNNDDQSDNDFYDNNSITTMQIDTSTAKILNNGPLEYNPDLPNKEQKLKD
SQVMQNQPPTATSTNSQQRTLQELINIMPSIEDISQQCKQQQQLKIQAKANSTQSASTANAANGGKGRKR
GRTVRFDQPLLGKVRQRNGDASDDEEPDEIEMLIRRLHTDILNDARNDPVEQAKKIRQARESQSDQTNST
TQLSVYERMILGSASQQSTDHQPGEFSNMFRTLEDEQIEINQNFLFDEYDSEDDSIADDKVEIASDDEQM
LLQEHKKRGKKYLQDEIVKEEDFDEDDDSDEDIHMDDLENESLSFDRNNRKSHKPVCKRTREENILDADL
GDEKDDEDTIFIDNLPSDEFSIRRQLQDVKSYIKQFEMLFFEEEDSDKEEQLKQITNVQKHEEALQNFKD
RSHLKNFWCIPLSSDVREIDWDVLIARQQEHTNGQLFDVITCDPPWQLSSANPTRGVAIAYETLNDGEIL
KIPWGRLQKDGFLFIWVINAKYRFALDMMGAHGYRVVDEIQWVKQTCNGKIAKGHGYYLQHAKEVCLVGC
KGDPAILAKKCRSNIESDVIFSERRDQSQKPEEIYELVEALVPNGYYMEIFGRRNNLHNGWVTVGNEL
>EJY79437.1 MT-A70 family protein [Oxytricha trifallax]
MHLPMQIITQNMFRQGNQHSCLNRTEILRTPRLTRSTKTELQEQTHFSKLPRRNYLKLQIDMREIQSLVD
KKVKESAAAQQQLSQSGIEDSAIKRSLRPRKVENYKNMLEGDEITLKTIQDEGIEVKRKKREASSQNRLE
DEDEDEDMLEVGQQIERASDDEDDDDFPISTRRSARKRTRRQDVDEDEEAIEVNQVESSDAEVEIPANDI
DTESYTEGTNKRKQKLKAKKQVLDKKKNKTEGDIDKEDAVEEEETVFIDNLPNDEFEIRRMLKEVKKHIK
SLEKQFFEEEDSEKEEELKQINNNSKHEEALQAFKETSHLKQFWCIPLSVNVTTLDFDLLAKSQMKQGGR
LFDVITIDPPWQLSSANPTRGVAIAYDTLNDKEILNMPFEKVQTDGFLFIWVINAKYRFALEMMEKFGYK
LVDEIAWVKQTVNGKIAKGHGYYLQHAKETCLVGVKGNVKGKARYNIESDVIFSQRRGQSQKPEEIYEIA
EALVPNGYYLEIFGRRNNLHNGWVTIGNEL
>NP_066012.1 N6-adenosine-methyltransferase non-catalytic subunit [Homo sapiens]
MDSRLQEIRERQKLRRQLLAQQLGAESADSIGAVLNSKDEQREIAETRETCRASYDTSAPNAKRKYLDEG
ETDEDKMEEYKDELEMQQDEENLPYEEEIYKDSSTFLKGTQSLNPHNDYCQHFVDTGHRPQNFIRDVGLA
DRFEEYPKLRELIRLKDELIAKSNTPPMYLQADIEAFDIRELTPKFDVILLEPPLEEYYRETGITANEKC
WTWDDIMKLEIDEIAAPRSFIFLWCGSGEGLDLGRVCLRKWGYRRCEDICWIKTNKNNPGKTKTLDPKAV
FQRTKEHCLMGIKGTVKRSTDGDFIHANVDIDLIITEEPEIGNIEKPVEIFHIIEHFCLGRRRLHLFGRD
STIRPGWLTVGPTLTNSNYNAETYASYFSAPNSYLTGCTEEIERLRPKSPPPKSKSDRGGGAPRGGGRGG
TSAGRGRERNRSNFRGERGGFRGGRGGAHRGGFPPR
>NP_964000.2 N6-adenosine-methyltransferase non-catalytic subunit [Mus musculus]
MDSRLQEIRERQKLRRQLLAQQLGAESADSIGAVLNSKDEQREIAETRETCRASYDTSAPNSKRKCLDEG
ETDEDKVEEYKDELEMQQEEENLPYEEEIYKDSSTFLKGTQSLNPHNDYCQHFVDTGHRPQNFIRDVGLA
DRFEEYPKLRELIRLKDELIAKSNTPPMYLQADIEAFDIRELTPKFDVILLEPPLEEYYRETGITANEKC
WTWDDIMKLEIDEIAAPRSFIFLWCGSGEGLDLGRVCLRKWGYRRCEDICWIKTNKNNPGKTKTLDPKAV
FQRTKEHCLMGIKGTVKRSTDGDFIHANVDIDLIITEEPEIGNIEKPVEIFHIIEHFCLGRRRLHLFGRD
STIRPGWLTVGPTLTNSNYNAETYASYFSAPNSYLTGCTEEIERLRPKSPPPKSKSDRGGGAPRGGGRGG
TSAGRGRERNRSNFRGERGGFRGGRGGTHRGGFTPR
>XP_003129279.3 N6-adenosine-methyltransferase subunit METTL14 [Sus scrofa]
MDSRLQEIRERQKLRRQLLAQQLGAESADSIGAVLNSKDEQREIAETRETCRASYDTSTPNAKRKYQDEG
ETDEDKIEEYKDELEMQQEEENLPYEEEIYKDSSTFLKGTQSLNPHNDYCQHFVDTGHRPQNFIRDVGLA
DRFEEYPKLRELIRLKDELIAKSNTPPMYLQADIEAFDIRELTPKFDVILLEPPLEEYYRETGITANEKC
WTWDDIMKLEIDEIAAPRSFIFLWCGSGEGLDLGRVCLRKWGYRRCEDICWIKTNKNNPGKTKTLDPKAV
FQRTKEHCLMGIKGTVKRSTDGDFIHANVDIDLIITEEPEIGNIEKPVEIFHIIEHFCLGRRRLHLFGRD

STIRPGWLTVGPTLTNSNYNAETYASYFSAPNSYLTGCTEEIERLRPKSPPPKSKSDRGGGAPRGGGRGG
TSAGRGRERNRSNFRGERGGFRGGRGGAHRGGFPPR
>XP_018099083.1 PREDICTED: N6-adenosine-methyltransferase subunit METTL14 isoform X2 [Xenopus laevis]
MNSRLQEIRARQTLRRKLLAQQLGAESADSIGAVLNSKDEQREIAETRETSRASYDTSAAVSKRKLPEEG
KADEEVVQECKDSVEPQKEEENLPYREEIYKDSSTFLKGTQSLNPHNDYCQHFVDTGHRPQNFIRDVGLA
DRFEEYPKLRELIRLKDELIAKSNTPPMYLQADLENFDLRELKSEFDVILLEPPLEEYFRETGIAANEKW
WTWEDIMKLDIEGIAGSRAFVFLWCGSGEGLDFGRMCLRKWGFRRSEDICWIKTNKDNPGKTKTLDPKAI
FQRTKEHCLMGIKGTVHRSTDGDFIHANVDIDLIITEEPEIGNIEKPVEIFHIIEHFCLGRRRLHLFGRD
STIRPDQSWEERLANSGGLREKEFLVGLLLGLLLPTATLIQRLMLLTLTLQIHLLLDAQRRSKDSVPKLH
LLSQIVALGHREEEDEVEHLQVAERGAGKGTEAVLGETEGISEDVEDHIGVSLLPVDFKCF
>NP_996954.1 N6-adenosine-methyltransferase non-catalytic subunit [Danio rerio]
MNSRLQEIRERQKLRRQLLAQQLGAESPDSIGAVLNSKDEQKEIEETRETCRASFDISVPGAKRKQLNEG
EDPEEDVEEQKEDVEPQHQEESGPYEEVYKDSSTFLKGTQSLNPHNDYCQHFVDTGHRPQNFIRDGGLAD
RFEEYPKQRELIRLKDELISATNTPPMYLQADPDTFDLRELKCKFDVILEPPLEEYYRESGIIANERFW
NWDDIMKLNIEEISSIRSFVFLWCGSGEGLDLGRMCLRKWGFRRCEDICWIKTNKNNPGKTKTLDPKAVF
QRTKEHCLMGIKGTVRRSTDGDFIHANVDIDLIITEEPEMGNIEKPVEIFHIIEHFCLGRRRLHLFGRDS
TIRPGWLTVGPTLTNSNFNIEVYSTHFSEPNSYLSGCTEEIERLRPKSPPPKSMAERGGGAPRGGRGGPA
AGRGDRGRERNRPNFRGDRGGFRGRGGPHRGFPPR
>NP_609205.1 methyltransferase like 14 [Drosophila melanogaster]
MSDVLKSSQERSRKRRLLLAQTLGLSSVDDLKKALGNAEDINSSRQLNSGGQREEEDGGASSSKKTPNEI
IYRDSSTFLKGTQSSNPHNDYCQHFVDTGQRPQNFIRDVGLADRFEEYPKLRELIKLKDKLIQDTASAPM
YLKADLKSLDVKTLGAKFDVILIEPPLEEYARAAPSVATVGGAPRVFWNWDDILNLDVGEIAAHRSFVFL
WCGSSEGLDMGRNCLKKWGFRRCEDICWIRTNINKPGHSKQLEPKAVFQRTKEHCLMGIKGTVRRSTDGD
FIHANVDIDLIISEEEEFGSFEKPIEIFHIIEHFCLGRRRLHLFGRDSSIRPGWLTVGPELTNSNFNSEL
YQTYFAEAPATGCTSRIELLRPKSPPPNSKVLRGRGRGFPRGRGRPR
>NP_567348.2 Methyltransferase MT-A70 family protein [Arabidopsis thaliana]
MKKKQEESSLEKLSTWYQDGEQDGGDRSEKRRMSLKASDFESSSRSGGSKSKEDNKSVVDVEHQDRDSKR
ERDGRERTHGSSSDSSKRKRWDEAGGLVNDGDHKSSKLSDSRHDSGGERVSVSNEHGESRRDLKSDRSLK
TSSRDEKSKSRGVKDDDRGSPLKKTSGKDGSEVVREVGRSNRSKTPDADYEKEKYSRKDERSRGRDDGWS
DRDRDQEGLKDNWKRRHSSSGDKDQKDGDLLYDRGREREFPRQGRERSEGERSHGRLGGRKDGNRGEAVK
ALSSGGVSNENYDVIEIQTKPHDYVRGESGPNFARMTESGQQPPKKPSNNEEEWAHNQEGRQRSETFGFG
SYGEDSRDEAGEASSDYSGAKARNQRGSTPGRTNFVQTPNRGYQTPQGTRGNRPLRGGKGRPAGGRENQQ
GAIPMPIMGSPFANLGMPPPSPIHSLTPGMSPIPGTSVTPVFMPPFAPTLIWPGARGVDGNMLPVPPVLS
PLPPGPSGPRFPSIGTPPNPNMFFTPPGSDRGGPPNFPGSNISGQMGRGMPSDKTSGGWVPPRGGGPPGK
APSRGEQNDYSQNFVDTGMRPQNFIRELELTNVEDYPKLRELIQKKDEIVSNSASAPMYLKGDLHEVELS
PELFGTKFDVILVDPPWEEYVHRAPGVSDSMEYWTFEDIINLKIEAIADTPSFLFLWVGDGVGLEQGRQC
LKKWGFRRCEDICWVKTNKSNAAPTLRHDSRTVFQRSKEHCLMGIKGTVRRSTDGHIIHANIDTDVIIAE
EPPYGSTQKPEDMYRIIEHFALGRRRLELFGEDHNIRAGWLTVGKGLSSSNFEPQAYVRNFADKEGKVWL
GGGGRNPPPDAPHLVVTTPDIESLRPKSPMKNQQQQSYPSSLASANSSNRRTTGNSPQANPNVVVLHQEA
SGSNFSVPTTPHWVPPTAPAAAGPPPMDSFRVPEGGNNTRPPDDKSFDMYGFN
>PNW88915.1 hypothetical protein CHLRE_01g050600v5 [Chlamydomonas reinhardtii]
MQDGQGPPGDGRGRGRGRSRGGRIMFAREGGRGPRPMHSDMGPPPPPMGMFPHDPSAMMGGPMPGMPPMD
FTPEMLLTMMGAGLGGPMGLAGPMGMMMPDFGAAAAGAPGGMMVPPGAMMPPPPQPPSGGPGGMGGGGMG
GMGGMMGHQQGMGGAGGPMGLPGGGMGMGMGGGGGGGGGGGYGGRGGHGEAGGGGGGGGRAGGAGGGGGA
GGAAEHLSNDYSQNFVDTGLRPQNFLRDTHLTDRYEEYPKLKELIVRKDRQVSAHATPPLFLRTDLRSTR
LSPELFGTKFDVILVDPPWEEYVRRAPGMVADPEVWSWQDIQALDIEAVADNPCFLFLWCGAEEGLEAGR
VCMQKWGFRRVEDICWIKTNKEGGKGPGGGRRPYLTAANQHPESMLVHTKEHCLMGIKGSVRRATDGHII
HTNVDTDVIVSEEPELGSTRKPEEMYHIIERFCNGRRRLELFGEDHNIRNGWVTVGRSLTSSNFSAKAYA
DHFRNRDGSVWVQNTYGPKPPPGSVILVPTTDEIEDLRPKSPTGPHGGSSFHHSR
>XP_001022374.1 MT-a70 family protein [Tetrahymena thermophila SB210]
MQPQQNQNQQQQQQQSQQQQQQNQQLPQLQQSMSSQQQQNQQQEKQIIIKRGTTSKRNDYCQNFVNTHE
RPQNFIMNIRPEERFIEYPKLQDLIKFKDDLIKKRNHPPVYLKADLKYYDLSKLGKFDVIMMDPPWKEYE
ERVQGLPIYSQYPEKFNSWDLNEIAALPIDEISDKPSFLFLWVGSDHLDQGRELFRKWGYKRCEDIVWVK
TNKDKTKEYIELPHSNLLVRVKEHCLVGLRGDVKRASDSHFIHANIDTDVIVAEEPPLGSTQKPAEIYDI
IERFCLGRKRLELFGEVHNVRQGWLTIGKLLDESNFNQDEYNSWFDGDKTYPQIQTYRGGRYVGTTPDIE
QLRPKSPTKNNQMNSNQNMSGSQVSEFDLGIQQKQQKLNQQF
>NP_009876.1 Kar4p [Saccharomyces cerevisiae S288C]
MAFQDPTYDQNKSRHINNSHLQGPNQETIEMKSKHVSFKPSRDFHTNDYSNNYIHGKSLPQQHVTNIENR
VDGYPKLQKLFQAKAKQINQFATTPFGCKIGIDSIVPTLNHWIQNENLTFDVVMIGCLTENQFIYPILTQ
LPLDRLISKPGFLFIWANSQKINELTKLLNNEIWAKKFRRSEELVFVPIDKKSPFYPGLDQDDETLMEKM
QWHCWMCITGTVRRSTDGHLIHCNVDTDLSIETKDTTNGAVPSHLYRIAENFSTATRRLHIIPARTGYET
PVKVRPGWVIVSPDVMLDNFSPKRYKEEIANLGSNIPLKNEIELLRPRSPVQKAQ
>XP_001691478.1 predicted protein [Chlamydomonas reinhardtii]

MRLGGGPGGSELDDLLGKRSVKEKVKVEKGSELLDILSKPTARESARVEQFRTAGGSAIREHCPHLTKDE
CRRVNGVPLACHRLHFLRVVQPHTDVALGNCSYLDTCRNMRTCKYVHYRPDPEPDVPGMGSEMARLRASV
PKKPVGDGQTSRGALDPQWINCDVRSFDMTVLGKFGVIMADPPWEIHQDLPYGTMKDDEMVNLNVGCLQD
NGVLFLWVTGRAMELARECMAKWGYKRVDELIWVKTNQLQRLIRTGRTGHWLNHSKEHCLVGIKGSPQLN
RYVDTDVVVAEVRETSRKPDEMYSLLERLSPGTRKLEIFARVHNCKPGWVGLGNQLKNVNLIEPEVRQRF
AARYGFEPDASKDCFVN
>NP_192814.1 mRNAadenosine methylase [Arabidopsis thaliana]
METESDDATITVVKDMRVRLENRIRTQHDAHLDLLSSLQSIVPDIVPSLDLSLKLISSFTNRPFVATPPL
PEPKVEKKHHPIVKLGTQLQQLHGHDSKSMLVDSNQRDAEADGSSGSPMALVRAMVAECLLQRVPFSPTD
SSTVLRKLENDQNARPAEKAALRDLGGECGPILAVETALKSMAEENGSVELEEFEVSGKPRIMVLAIDRT
RLLKELPESFQGNNESNRVVETPNSIENATVSGGGFGVSGSGNFPRPEMWGGDPNMGFRPMMNAPRGMQM
MGMHHPMGIMGRPPFFPLPLPLPVPSNQKLRSEEEDLKDVEALLSKKSFKEKQQSRTGEELLDLIHRPTA
KEAATAAKFKSKGGSQVKYYCRYLTKEDCRLQSGSHIACNKRHFRRLIASHTDVSLGDCSFLDTCRHMKT
CKYVHYELDMADAMMAGPDKALKPLRADYCSEAELGEAQWINCDIRSFRMDILGTFGVVMADPPWDIHME
LPYGTMADDEMRTLNVPSLQTDGLIFLWVTGRAMELGRECLELWGYKRVEEIIWVKTNQLQRIIRTGRTG
HWLNHSKEHCLVGIKGNPEVNRNIDTDVIVAEVRETSRKPDEMYAMLERIMPRARKLELFARMHNAHAGW
LSLGNQLNGVRLINEGLRARFKASYPEIDVQPPSPPRASAMETDNEPMAIDSITA
>EAS00013.2 N6-adenosine-methyltransferase 70 kDa subunit [Tetrahymena thermophila SB210]
MGSSVKDQEISNKKHKARNSSSGANNNSNSSNYQSSKRDIHQDRSYSKDDSQSRQYNSNNGGGGSSSKNS
NRNSSQQGYNQNSSSNQGQNSEYGGSGSGKNSQANSQRNSSQQGLQQLNQQQQSQQQQQQQMLQNQMNSMG
MMNQFQNSFGLMGMQPSQPLQLLNPSMIIPSGKKQKYDFLEFPPSSQHEFRAILLDYFLSDLFDYPMHSA
ELFENFIEAFSDIKDSSSFIKKLELIPLLQELNDKKAIKLETCAVGTKLFDFIVDINKDKIKQLSREFSK
DRPKFMPILDKKPQPSSSKTNSSSTTAPPKQAISKREIEDLLKKETGLQKEVITQSKEKSNLLNKISAAE
ESALAIFRKQGSRRIDYCDCGTRDKCIQIRNSTVPCNKAHFRKIIRPHTDENLGNCSYLDTCRHMDYCKF
VHYELDVDINNMNNDNLLLDGIEKKLNPQWINCDLRQIDFNILGKFNCIMADPPWDIHMTLPYGTLKDRE
MKAMRVDLLQEEGVIFLWVTGRAMELGRECLTNWGYRRVEEIIWVKTNQLQRIIRTGRTGHWLNHSKEHC
LVGIKGNPKINRKIDCDVIVSEVRETSRKPDEIYNLIERMCPGGKKIELFGRPHNTMPGWLTLGNQLPGI
YLEDEEIIERYMDAYPDQDISRETMERNRIRMKNENDIDHIYNSHIQNIPPFKTKQLTKDLQLQQQSSSM
QTTQQQSSSQMMPQMQQQQSSQSINSNTDLQMHGNGLYEQE
>ORX92345.1 MT-A70-domain-containing protein [Basidiobolus meristosporus CBS 931.73]
MKLERALFKMADMWGYNTIGIKREYDNDKSAISVIYFDPRNLRNVQHIEKTLEDICDVDSIDPDIFLDKT
TSAQVPSTYIPNEEARFSEDAEIEKLLSKPSFLEMEAFSSLIGVTELIERKTFREQEAEEMFKAQGNGGF
REFCEYLIKEDCKKMNTSGQPCAMTASILLTNMKLHFRRIMRPQTDLELGDCSYLNTCHRMDTCKYVHYE
LDDFEHPSSANITKTTIPTSLIFRPPKKVLPAQWINCDVRKFDFSILGKFSVIMADPPWDIHMTLPYGTM
TDDEMKAMAIHKLQDEGLIFLWVTARAMELGRECLATWGYDRVDEVVWIKTNQLQRLIRTGRTGHWLNHS
KEHCLVGIKGDPSRFNIGLACDVLVAEVRETSRKPDQIYGMIDRLSPGTRKIEIFGRQHNTRPGWFTLGN
QLKDVRIYEPEVLEAYNQRYPECPAQLSAIPES
>AJR96662.1 Ime4p [Saccharomyces cerevisiae YJM1248]
MINDKLVHFLIQNYDDILRAPLSGQLKDVYSLYISGGYDDEMQKLRNDKDEVLQFEQFWNDLQDIIFATP
QSIQFDQNLLVADRPEKIVYLDVFSLKILYNKFHAFYYTLKSSSSSCEEKVSSLTTKPEADSEKDQLLGR
LLGVLNWDVNVSNQGLPREQLSNRLQNLLREKPSSFQLAKERAKYTTEVIEYIPICSDYSHASLLSTAVY
IVNNKIVSLQWSKISACQENHPGLIECIQSKIHFIPNIKPQTDISLGDCSYLDTCHKLNMCRYIHYLQYI
PSCLQERADRETAIENKRIRSNVSIPFYTLGNCSAHCIKKALPAQWIRCDVRKFDFRVLGKFSVVIADPA
WNIHMNLPYGTCNDIELLGLPLHELQDEGIIFLWVTGRAIELGKESLNNWGYNVINEVSWIKTNQLGRTI
VTGRTGHWLNHSKEHLLVGLKGNPKWINKHIDVDLIVSMTRETSRKPDELYGIAERLAGTHARKLEIFGR
DHNTRPGWFTIGNQLTGNCIYEMDVERKYQEFMKSKTGTSHTGTKKIDKKQPSKLQQQHQQQYWNNMDMG
SGKYYAEAKQNPMNQKHTPFESKQQQKQQFQTLNNLYFAQ
>NP_651204.1 methyltransferase like 3 [Drosophila melanogaster]
MADAWDIKSLKTKRNTLREKLEKRKKERIEILSDIQEDLTNPKKELVEADLEVQKEVLQALSSCSLALPI
VSTQVVEKIAGSSLEMVNFILGKLANQGAIVIRNVTIGTEAGCEIISVQPKELKEILEDTNDTCQQKEEE
AKRKLEVDDVDQPQEKTIKLESTVARKESTSLDAPDDIMMLLSMPSTREKQSKQVGEEILELLTKPTAKE
RSVAEKFKSHGGAQVMEFCSHGTKVECLKAQQATAEMAAKKKQERRDEKELRPDVDAGENVTGKVPKTES
AAEDGEIIAEVINNCEAESQESTDGSDTCSSETTDKCTKLHFKKIIQAHTDESLGDCSFLNTCFHMATCK
YVHYEVDTLPHINTNKPTDVKTKLSLKRSVDSSCTLYPPQWIQCDLRFLDMTVLGKFAVVMADPPWDIHM
ELPYGTMSDDEMRALGVPALQDDGLIFLWVTGRAMELGRDCLKLWGYERVDELIWVKTNQLQRIIRTGRT
GHWLNHGKEHCLVGMKGNPTNLNRGLDCDVIVAEVRATSHKPDEIYGIERLSPGTRKIELFGRPHNIQP
NWITLGNQLDGIRLVDPELITQFQKRYPDGNCMSPASANAASINGIQK
>NP_001084701.1 methyltransferase like 3 L homeolog [Xenopus laevis]
MSDTWSSIQAHKKQLDNLRERLQRRRKDATSQLALDLQSSEGGIAPTFRSDSPVPSASSQPLKGPSGSAE
VTPDPELEKKLLHHLSDLSLVLPADSVSIQLAITTPDFPVTRQGVESLLQKFAAQELIEVKGWGQEDDDR
PTVVTFADYSKLSAMMGAVAERKGTTIPTGAKKRRLQEADPSASSLSSSLSASASREKKTSEPQKKARKH
ASHLDLEIESLLSQQSTKEQQSKKVSQERLELLSTSTAREQSIVEKFRSRGRAQVQEFCDFGTKEECMKA
AGADTPCRKLHFRRIINMHTDESLGDCSFLNTCFHMDTCKYVHYEIDAWVEPGGTAMGTEAIASLDTPLA

KAVGDSSVGRLFPAQWIRCDIRYLDVSILGKFSVVMADPPWDIHMELPYGTLTDDEMRKLQIPVLQDDGF
LFLWVTGRAMELGRECLKLWGYERVDEIIWVKTNQLQRIIRTGRTGHWLNHGKEHCLVGVKGSPQGFNRG
LDCDVIVAEVRSTSHKPDEIYGMIERLSPGTRKIELFGRPHNIQPNWITLGNQLDGIHLLDPDVVAQFKQ
KYPDGVIGMPKNM
>sp|F1R777.1|MTA70_DANRE RecName: Full=N6-adenosine-methyltransferase subunit METTL3; AltName: Full=N6-adenosine-methyltransferase 70 kDa subunit; Short=MT-A70
MSDTWSHIQAHKKQLDSLRERLQRRRKDPTQLGTEVGSVESGSARSDSPGPAIQSPPQVEVEHPPDPELE
KRLLGYLSELSLSLPTDSLTITNQLNTSESPVSHSCIQSLLLKFSAQELIEVRQPSITSSSSSTLVTSVD
HTKLWAMIGSAGQSQRTAVKRKADDITHQKRALGSSPSIQAPPSPPRKSSVSLATASISQLTASSGGGGG
GADKKGRSNKVQASHLDMEIESLLSQQSTKEQQSKKVSQEILELLNTSSAKEQSIVEKFRSRGRAQVQEF
CDYGTKEECVQSGDTPQPCTKLHFRRIINKHTDESLGDCSFLNTCFHMDTCKYVHYEIDSPPEAEGDALG
PQAGAAELGLHSTVGDSNVGKLFPSQWICCDIRYLDVSILGKFAVVMADPPWDIHMELPYGTLTDDEMRK
LNIPILQDDGFLFLWVTGRAMELGRECLSLWGYDRVDEIIWVKTNQLQRIIRTGRTGHWLNHGKEHCLVG
VKGNPQGFNRGLDCDVIVAEVRSTSHKPDEIYGMIERLSPGTRKIELFGRPHNVQPNWITLGNQLDGIHL
LDPEVVARFKKRYPDGVISKPKNM
>NP_062826.2 N6-adenosine-methyltransferase catalytic subunit [Homo sapiens]
MSDTWSSIQAHKKQLDSLRERLQRRRKQDSGHLDLRNPEAALSPTFRSDSPVPTAPTSGGPKPSTASAVP
ELATDPELEKKLLHHLSDLALTLPTDAVSICLAISTPDAPATQDGVESLLQKFAAQELIEVKRGLLQDDA
HPTLVTYADHSKLSAMMGAVAEKKGPGEVAGTVTGQKRRAEQDSTTVAAFASSLVSGLNSSASEPAKEPA
KKSRKHAASDVDLEIESLLNQQSTKEQQSKKVSQEILELLNTTTAKEQSIVEKFRSRGRAQVQEFCDYGT
KEECMKASDADRPCRKLHFRRIINKHTDESLGDCSFLNTCFHMDTCKYVHYEIDACMDSEAPGSKDHTPS
QELALTQSVGGDSSADRLFPPQWICCDIRYLDVSILGKFAVVMADPPWDIHMELPYGTLTDDEMRRLNIP
VLQDDGFLFLWVTGRAMELGRECLNLWGYERVDEIIWVKTNQLQRIIRTGRTGHWLNHGKEHCLVGVKGN
PQGFNQGLDCDVIVAEVRSTSHKPDEIYGMIERLSPGTRKIELFGRPHNVQPNWITLGNQLDGIHLLDPD
VVARFKQRYPDGIISKPKNL
>sp|Q8C3P7.2|MTA70_MOUSE RecName: Full=N6-adenosine-methyltransferase subunit METTL3; AltName: Full=Methyltransferase-like protein 3; AltName: Full=N6-adenosine-methyltransferase 70 kDa subunit; Short=MT-A70
MSDTWSSIQAHKKQLDSLRERLQRRRKQDSGHLDLRNPEAALSPTFRSDSPVPTAPTSSGPKPSTTSVAP
ELATDPELEKKLLHHLSDLALTLPTDAVSIRLAISTPDAPATQDGVESLLQKFAAQELIEVKRGLLQDDA
HPTLVTYADHSKLSAMMGAVADKKGLGEVAGTIAGQKRRAEQDLTTVTTFASSLASGLASSASEPAKEPA
KKSRKHAASDVDLEIESLLNQQSTKEQQSKKVSQEILELLNTTTAKEQSIVEKFRSRGRAQVQEFCDYGT
KEECMKASDADRPCRKLHFRRIINKHTDESLGDCSFLNTCFHMDTCKYVHYEIDACVDSESPGSKEHMPS
QELALTQSVGGDSSADRLFPPQWICCDIRYLDVSILGKFAVVMADPPWDIHMELPYGTLTDDEMRRLNIP
VLQDDGFLFLWVTGRAMELGRECLNLWGYERVDEIIWVKTNQLQRIIRTGRTGHWLNHGKEHCLVGVKGN
PQGFNQGLDCDVIVAEVRSTSHKPDEIYGMIERLSPGTRKIELFGRPHNVQPNWITLGNQLDGIHLLDPD
VVARFKQRYPDGIISKPKNL
>XP_003128628.1 N6-adenosine-methyltransferase 70 kDa subunit [Sus scrofa]
MSDTWSSIQAHKKQLDSLRERLRRRRKQDSGHLDLRNPEAALSPTFRSDSPVPTVPTSGGPKPSTASAVP
ELATDPELEKKLLHHLSDLALTLPTDAVSIRLAISTPDAPATQDGVESLLQKFAAQELIEVKRGLLQDDA
HPTLVTYADHSKLSAMMGAVAEKKGPGEVAGTITGQKRRAEQDSTTVAAFASSLTSGLASSASEVAKEPT
KKSRKHAASDVDLEIESLLNQQSTKEQQSKKVSQEILELLNTTTAKEQSIVEKFRSRGRAQVQEFCDYGT
KEECMKASDADRPCRKLHFRRIINKHTDESLGDCSFLNTCFHMDTCKYVHYEIDACMDSEAPGSKDHTPS
QELALTQSVGGDSNADRLFPPQWICCDIRYLDVSILGKFAVVMADPPWDIHMELPYGTLTDDEMRRLNIP
VLQDDGFLFLWVTGRAMELGRECLNLWGYERVDEIIWVKTNQLQRIIRTGRTGHWLNHGKEHCLVGVKGN
PQGFNQGLDCDVIVAEVRSTSHKPDEIYGMIERLSPGTRKIELFGRPHNVQPNWITLGNQLDGIHLLDPD
VVARFKQRYPDGIISKPKNL
>WP_009339935.1 MULTISPECIES: S-adenosylmethionine-binding protein [Afipia]
MTLPAKDLLSFAGQRRFSTILADPPWQFTNKTGKVAPEHKRLSRYGTMKLDEIMMLPVADIAAPTSHLYL
WCPNALLPEGLAVMKAWGFNYKSNIVWHKVRKDGGSDGRGVGFYFRNVTEVILFGVRGKNARTLAPGRRQ
VNLLATRKREHSRKPDEQYEIIESCSPGPFLELFARGTRKNWATWGNQADDDYKPTWKTYAHHSRAGLVA
AE
>WP_013486562.1 S-adenosylmethionine-binding protein [Ethanoligenens harbinense]
MSTAKETANNLLQFCGEKKYATVYADPPWRFQNRTGKVAPENKKLNRYPTMDLEDIKALPVGKIAAEKSH
LYLWVPNALLPDGLEVMKAWGFEYKGNIIWEKVRKDGEPDGRGVGFYFRNVTEILLFGIRGGNNRTLAPA
RSQVNLIRTQKREHSRKPDEIITIIESCSPGPYLELFARGDRENWDMWGNQATAEYEPTWNTYKNHTTKE
TTSGVSGSQSET
>WP_016343787.1 adenine-specific DNA methyltransferase [Mycobacteroides abscessus]
MAAPLREVNEPPPLPVTDGGFSTILADPPWRFTNRTGKVAPEHRRLDRYSTLSLDEICALGVSDVTADNA
HLYLWVPNALLPDGLRVMEEWGFRYVSNIVWSKVRRDGLPDGRGVGFYFRNTTELLLFGVRGSMRTLQPA
RSQVNQIVTRKREHSRKPDEQYELIEACSPGPYLEMFGRYRRPNWAVWGDEANEDVEPRGQTHKGYGGGE
ITRLPALEPHSRIPQWLAKPIAAAIKSAYDDGMSIDAIAAETGYSISRVRHLLDQAGAKKRGRGRPAKA
>WP_023133224.1 MULTISPECIES: MT-A70 protein [Rothia]
MLDPMNTNEEFAPLPTVEGGFQTVLADPPWRFTNRTGKVAPEHHRLGRYGTMSLDEIKALRVGDVTADNA
HLYLWVPNALLPEGLEVMQAWGFRYVSNIIWAKRRKDGGPDGRGVGFYFRNVTEPILFGVKGSMRTLAPG

RSTVNMIETRKREHSRKPDEQYDLIEACSPGPYLELFARYARPGWSVWGNEASNEIEPRGKAQKGYGGGE
IDRLPILEPNERMSEWLSGRVGELLAEEYTKGASVQELANQSGYSIARVRTLLTHSGVPLRGRGRPKKGQ
VAS
>ETW92643.1 S-adenosylmethionine-binding protein [Candidatus Entotheonella factor]
MSNSPHSAADDLLACGFPPHSFSTVLADPPWRFTNRTGKMAPEHRRLSRYPTLTLEEIADLPLAQLVQPD
SHLYLWVPNALLAEGLDVMRRWGFTYKTNLVWYKIRRDGGPDRRGVGFYFRNVTELVLFGVRGRMRTLAP
GRRQENLLASQKQEHSRKPDTFYDLIERCSPGPYLELFARHPRPGWHQFGNEPLVSSS
>AHJ63281.1 Adenine-specific methyltransferase [Granulibacter bethesdensis]
MTKQPDPIAEFRNQLNGGNFATVLADPPWRFQNRTGKMAPEHRRLSRYGTMELPEIMALPVSEVTAKTAH
LYLWVPNALLPEGLAVMQAWGFNYKSNLVWHKIRKDGGSDGRGVGFYFRNVTELVLFGVKGKNARTEAPG
RRQVNLLATQKREHSRKPDEFYDIVEACSPGPYLELFARGTRPGWCAWGNQAEEYDITWDTYSHHSQRQS
LWVAE
>WP_017364718.1 S-adenosylmethionine-binding protein [Methylococcus capsulatus]
MTENTLDPAADLLERLGDKRFRTILADPPWQFQNRTGKMAPEHKRLNRYGTMSLEAIAGLPVERLTADTA
HLYLWVPNALLLEGLKVMEAWGFTYKTNLVWHKIRKDGGPDGRGVGFYFRNVTELVLFGVRGKNARTLAA
GRRQVNFLATRKREHSRKPDEMYGIIEACSPGPYLELFARGARDRWSVWGNEADENYYPRWNTYANHSQA
EICPFE
>WP_027700599.1 S-adenosylmethionine-binding protein [Xylella fastidiosa]
MTKHKANTASDVGRDLLARHGGQRFHTILADPPWQFQNRTGKMAPEHKRLSRYGTMTLDDIMMLPVEQLV
TDTAHLYLWVPNALLPEGIKVLEAWGFSYKSNIVWHKVRKDGGPDGRGVGFYFRNVTELVLFGVRGKNAR
TLAPGRRQVNFLATQKREHSRKPDEFYDIVESCSPGPFLELFARGPRDGWKVWGNQADKYYPTWPTYSNH
SQAECELGRVEMIAQRLLSV
>WP_027488351.1 S-adenosylmethionine-binding protein [Rhizobium undicola]
MLNRNTDAPSPSDDFTNFISGRKFATIMADPPWQFMNRTGKVAPEHKRLNRYGTMELDAIKALPVATACA
PTAHLYLWVPNALLPEGLEVMKAWGFNYKANIVWHKLRKDGGSDGRGVGFYFRNVTELILFGTRGKNART
LPPGRSQVNYIGTRKREHSRKPDEQYPLIESCSPGPYLEMFGRGLRKGWTTWGNQADETYEPTWKTYGHN
SSTDRLEAAE
>ESK34829.1 hypothetical protein G986_02949 [Escherichia coli UMEA 3323-1]
MGWFMTKKYTLIYADPPWVYRDKAADGNRGAGFKYPVMSVLDICRLPVWDLADENCLLAMWWVPTQPLEA
LKVVEAWGFRLMTMKGFTWIKCGSRQPDKLVMGMGHMTRANSEDCLFAVKGKLPTRINAGIVQSFTAPRL
EHSRKPDIVREKLVQLLGDVSRIELFARQTSHGFDVWGNQCEDPAVQLHPGYALDKGGLTNAFSNAPLSP
TDIQGRERAA
>AIF94871.1 Adenine DNA methyltransferase, phage-associated [Escherichia coli O157:H7 str. SS17]
MTKKYTLIYADPPWTFRDKATDGQRGASFKYPVMSLLDICRLPVWELAADNCLLAMWWVPTQPLEALKVV
EAWGFRLVTMKGLTWNKCGKRQTDKLVMGMGSTTRANSEDCLFAVKGNLPERINAGIIQSFTAPRLDHSR
KPDMAREKLVQLLGDVPRIELFARHTSHGFDVWGNQCGTPSIEMVPGIVKFLEKTNERKNDVDKGITS
>WP_032715146.1 adenine methylase [Klebsiella aerogenes]
MTGKYTLIYADPPWSYRDKAADGDRGAGFKYPVMNVMDICRLPVWELSADDCLLAMWWVPTQPVEALKVV
EAWGFRLMTMKGFTWHKINKHKGNSAIGMGHMTRANSEDCLFAVRGKLPERMDASICQHVTAPRLENSRK
PDVIREKLVQLLGDVPRIELFARQSSHGFDVWGNQCIAPAVELLPGCAVPVVKTEAA
>AIA43360.1 DNA methyltransferase [Klebsiella pneumoniae subsp. pneumoniae KPNIH27]
MNYDLIYCDPPWEYGNRISNGAACNHYSTMSIDDLKFLPVRKLAADNAVLAMWYTGTHNREAVELAESWG
FRVRTMKGFTWVKLNGNAADRFNKALSTGELVDFNDLLEMLDRETRMNGGNHTRSNTEDVLIATRGTGLP
RASASVKQVVHTCLGEHSAKPWEVRNRLEQLYGDVKRIELFAREEWKGWDRWGNQCNNSIEIITGLIKEV
NHAA
>WP_009320301.1 DNA methyltransferase [Clostridioides difficile]
MPAVLFLLELHRRRKGGYKIENNQKYNIIYADPPWRYQQKRLSGAAEHHYPTMSVKDICGLKVEEIAAKD
CVLFLWATFPQLPEALRVIKAWGFQYKTVAFVWLKQNKSGKGWFFGLGFWTRGNAEICLLAIKGKPHRNS
NRVHQFLISPIRGHSQKPEEAREKIVELMGDLPRVELFAREKTEGWDAWGNEVESDIEISSDTEKEWR
>WP_012115592.1 MT-A70 family protein [Xanthobacter autotrophicus]
MNGLWQFGDLKMFGYDLIVADPPWDFELYSEAGEGKSAKAHYGTMKLDEIAALRVGDLARGDCLLLLWCC
EWMPPAARQRVLDAWGFTYKTTIIWRKVTRAGKVRMGPGYRARTMHEPVIVATVGNPKHTPFSSVFDGVA
REHSRKPEAFYRMVEAAAPKAARADLFSRQRRDGWDAFGNEVEKFDQPPAEAAE
>KFL31466.1 DNA methyltransferase [Devosia riboflavina]
MTAWPFGAMPMFSFDVVMADPPWSFDNWSEGGNAKNAKAQYDCMPTPDIKRLPVGHLAAGDCWLWLWATY
PMLPDAIEVMDAWGFRYVTAGPWVKRGTSGKLAMGTGYVLRSCSEIFLIGKNGEPKTHARDVRNVLEAPR
REHSRKPDEAYAMAEKLFGPGRRADLFSRETRPGWTSWGNESTKFDEVAA
>WP_016734132.1 DNA methyltransferase [Rhizobium phaseoli]
MRLFPDLWPFGDLQPHSFDFIMADPPWKMQEWSDNGDKSKSTQSKYRLMPLDEIKAMPVLDLAAPNCLLW
LWATNPMLPQALDVLHAWGFTFATAGSWMKTTRNGKQAFGTGYIFRTSNEPILIGKRGEPKTTRSVRSSF
PGLAREHSRKPEEGYREAERLMPRARRLELFSRTNRVGWTTWGDEVGKFGDVA
>KFB10357.1 Adenine-specific methyltransferase [Nitratireductor basaltis]
MHLFDWPFGDLNPHSFDLIMADPPWAFELRSDKGEGKSAQSHYKCQTLDEIKALPVLDLAAPDCLLWLWA

TNPMLPQAFEVMAAWGFTFKTAGAWGKTTVNGKLAFGTGYIFRSAHEPILIGTRGEPRTTKSVRSLIMGQ
VREHSRKPEEAYAAAEKLIPNARRLELFSRTDRAGWEVWGDEAGKFGEAA

Protein sequences for phylogenetic analysis of p1 proteins

>XP_001009903.1 [Tetrahymena thermophila SB210]
MSLKKGKFQHNQSKSLWNYTLSPGWREEEVKILKSALQLFGIGKWKKIMESGCLPGKSIGQIYMQTQRLLGQQSLGDFMGLQIDLEAVFNQNM
KKQDVLRKNNCIINTGDNPTKEERKRRIEQNRKIYGLSAKQIAEIKLPKVKKHAPQYMTLEDIENEKFTNLEILTHLYNLKAEIVRRLAEQGETIAQPS
IIKSLNNLNHNLEQNQNSNSSTETKVTLEQSGKKKYKVLAIEETELQNGPIATNSQKKSINGKRKNNRKINSDSEGNEEDISLEDIDSQESEINSEEI
VEDDEEDEQIEEPSKIKKRKKNPEQESEEDDIEEDQEEDELVVNEEEIFEDDDDDEDNQDSSEDDDDDED
>EJY79729.1 [Oxytricha trifallax]
MSSSISAAIIAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWKTIEQSQCLPTKTMSQMYLQTQRLVGQQSLAEFMGLHLDLEQ
IFIKNAERQGAGVFRKNGCIINTGDNMTKVQIAKLRKKNSKIFGLTQPFVQSLHLPKAKVKEWLKVLTLDQILSAKSNFSTAEKIHYLKILENALERKL
KKILRLQELVSIYRPCNIGIVVQKRLGSSIGDEYFEYVDCVKIEEKSVGNLDFALPNRNTDSTSLNEDFSFLDSTQKPQKLKAGSGRENKRKKMRD
GLKDERAQRQSLMEALDEQEFDETKFQDSDGEMPDLNM
>EJY78001.1 [Oxytricha trifallax]
MSVHHKMADSKSLHNYTLSPGWTREEVDILKIALMKFGIGKWKKIQKSGCLPSKTISQMNLQTQRLLGQQSLAEFMGLHVYLDRVFRDNSLKTG
PEIQRKNNFIINTGNNLTQPEKEKRLRLNKQKYGLDLAFIKTLRLPKPESATGGKREAILSMDQIFAQKSHFTVVEKLKHLEALKNALCSKLGKIERR
RRNKELSKIYRPLGQLIVVQKNADDQYEFVDIIDENE
>ORX69504.1 [Linderina pennispora]
MSSATPYAPRSMPTGQRNVVRSNDSASLWNCTLSPGWTQEEVQVLRKALMKFGVGNWMKIIESECLPGKTIAQMNLQTQRMLGQQSTAEFN
GLHLDAFVIGELNSKKQGPGIKRKNNCIVNTGGKLTRDEVVKRQQKHREQYEVKAEVWRAIVLPKPDNPLILLEKKREELKKVRLELEEIMKQIEE
>ORX78557.1 [Basidiobolus meristosporus CBS 931.73]
MTDVYKPRSMPVGARNVLRSNDSASLWNCTLSPGWTEPEVHILRKAVMKFGIGNWAKIIESQCLFGKTIAQMNLQLQRMLGQQSTAEFAGLHL
DPFVIGEINSKKQGPGIKRKNNCIVNTGGKLTREEIKRRLLEHKRTYEISEEEWRSIELPKPEDPGAVLIAKKDELKMLEDELLRVVQKIQKAREER
>EJY73777.1 [Oxytricha trifallax]
MSHATSHGNSTEKDKKNSGNMVAESKSLWNYALSPGWTPQEVDVLKIALMKFGIGKWTIIDKSGILPTKTIQQCYLQTQRILGQQSLAEFMGLHV
DIDKIALDNRRKNGIRKMGFLVNQGGKLTPEEKAHYQEINRQKYGLSPEEVETIKLPPPCSVEIYDINKIINPKSKLTTIEKINHCIKLQDALLEKLENI
KNKKIPTGAGFSSSRVYENMRGYDPQLLLNSHVTGQLDHSMQDLTIDERYSDLDEEEDPLAMASIIDSQATPQPQKIKSSVPNKASTTPSAKEMN
QIKDIIDSVIAENSAQQSKNLAQEKPKLKFSLVKATESNLLQSAAQNSDDVVMEEDSKLQHIETFSTVTQTATDQSNSQSKSQNNIASDSLKDSLE
QNDLSKSLTDSLEMQQYSAEKKLNQAPMSKNSDKPKKKRLNKRKLPSDDEFETL
>XP_021883515.1 [Lobosporangium transversale]
MSSGSTPRSMTAGARNILRSNDSASLWNYTVAPGWSMKEAEILRKALMKFGIGNWSKIIESNCLVGKTNAQMNLQTQRMLGQQSTAEFAGLHI
DPRVIGQKNSLIQSDHIRRKNGCIVNTGAKLSREEIRRRVAENKEQYELPEEEWSSIELPLPDDPHLLLEAKKSEKVRLELELKNVQRQIAMLRKV
GRKFETGSESPKTELDDDERDEFIEDQPLGKRARIEA
>EJY81929.1 [Oxytricha trifallax]
MSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLIGQQSLAEFMGLHLDIDRI
AADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLSAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKIKHLIKLEDALKSKLEMIRE
GKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSSHQNRNSDLDEENDESVMIIDESQQENLTPKGKAQAMLTHQKYNEVTQTMIKQ
GDDSRGQQHLPLDSTSASVSNPSSTSKSSTMKSNSMKQSETAIASMKPSSIGKKTKVDSSFVTKQSNQQSTAPIQKQAHQQNLDRNRSELGST
FAQQASVDTQNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVEDSE
>EJY76686.1 [Oxytricha trifallax]
MRVYLKFCNRKQIHYTHTMSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLI
GQQSLAEFMGLHLDIDRIAADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLTAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKI
KHLIKLEDALKSKLEMIREGKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSAHQNRNSDLDEENDQSVMIIDESQQQNLTPKGKAQ
TMLTNQTQTMKKQADDSRDEQHLPLISTSASVSNPSSTSKSSALKLNSMKQSDTAIASMKPSSSGKKTKVDSSFVSKQSNQQSTSYSETNVDT
QNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVEDSE
>EJY80746.1 [Oxytricha trifallax]
MRVYLKFCNRKQIHYTHTMSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLI
GQQSLAEFMGLHLDIDRIAADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLTAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKI
KHLIKLEDALKSKLEMIREGKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSAHQNRNSDLDEENDQSVMIIDESQQQNLTPKGKAQ
TMLTNQTQTMKKQADDSREEQHLPLNSTSASVSNPSSTSKSSALKLNSMKQSDTAIASMKPSSSGKKTKVDSSFVSKQSNQQSTGPIQKQAHQ
QNLDRNRSELGSTFAQQTNVDTQNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVKDSE
>ORX56566.1 [Piromyces finnis]
MSIPKPRSMPVGFRNILRPNDSTSLWNCTLSPGWTQEESDILRDALIFYGIGNWKDIIEHGCLPDKTNAQMNLQLQRMLGQQSTAEFQNLHIDPY
EIGKINSQKQGPNIRRKNGFIINTGGKLSREDIKRKIQENKENYELPEEVWSKIVLPNREVVTINEKRQKLNKLEEELDSVLKQIVNRRRELRGMTP
LKETEMKSIVNRSNQNDTKTEEKEIKEEESTTVNEEKIENTETSSISIISTNENEQSENISSSSPIVKSEQKKKRVVSRRKNKRRVNSDDEGFLPPG
KSRSKRTRRTPKKSSN
>ORX79686.1 [Anaeromyces robustus]
MSIPKPRSMPTGFRNILRPNDSTSLWNCTLSPGWTQEESDILRDALIYYGIGNWKDIIEHGCLPDKTNAQMNLQLQRMLGQQSTAEFQNLHIDPY
VIGKINSQKQGPNIRRKNGFIINTGGKLSREDIRRKIQENKENYELPKEEWSKIVLPNREVVIKNKVQEAINEKREKLNKLEDELDSVLKAIVNRRRRE
LRGMIPLKDSEMKSLVNRSAKNEGENKTETTNNEESNNTNNSDDIKDENNETSTSSHIFTNNDNELSENNSSSSSSNSISNKKKRFLRREVRRGK
RRYNYDDDDFMPSGNRSRKSRKI
>ORZ01404.1 [Syncephalastrum racemosum]
MSNNKENNVNKPRSMTAGARNVLRSNDSTSLWNCTLSPGWTQDESEVLRKALMKFGVGNWAKIIESGCLPGKTNAQMNLQLQRLLGQQSTAE
FAGLHIDPKVIGEKNSKIQGPHIKRKNNCIVNTGDKLSRDKLRARVMSNKEEYELPEEVWKNIELPKVKDPLMLLEGKKEEMRKLKTELEKVQAKI
QQLRQAQPARVQELQSQIEVARSPSPSAPDSPALSV
>XP_001698763.1 [Chlamydomonas reinhardtii]

MAFAAALAEKRGPRVGDAASLWNFTPAPGWSREEVQILRLCLMKHGVGQWMQILSTGLLPGKLIQQLNGQTQRLLGQQSLAAYTGLKVDVDRI
RVDNETRTDATRKAGLIINDGPNLTKEMKEKMRQDAVAKYGLTPEQVAEVDEQLAEIAAAFNPASTSAAAGAGSGAAAAGQAAAAGSGAGGSG
NLMAQPTEQLSAEQLGQLLLRLRNRLACLVDRARGRAGLPPRTAPRWATEAAAAAACLAAMAAAEASAPQAPAAAAGGQEGAAGPVMVSVPFS
REVLAEATACRVRSGTAAGARGNAPGAQGGVRKRTSKGGKAKGGDREWSPEGEENTAPQPRGGGKRKSGAVAGGEEADGVASGRAKRAS
RPKRGSSKHDPYVDDNDYGDEGIDPFDVGDDLDDMNPHGRYGNGGGRRADPSEAISALTAMGFTQSKARGALRECNFNVELAVEWLFANCL
>PNW76495.1 [Chlamydomonas reinhardtii]
MAFAAALAEKRGPRVGDAASLWNFTPAPGWSREEVQILRLCLMKHGVGQWMQILSTGLLPGKLIQQLNGQTQRLLGQQSLAAYTGLKVDVDRI
RVDNETRTDATRKAGLIINDGPNLTKEMKEKMRQDAVAKYGLTPEQVAEVDEQLAEIAAAFNPASTSAAAGAGSGAAAAGQAAAAGSGAGGSG
QAATAADAGGAAGRGTGSAGGAAAAAPPRNALAISTGVLAATLLDASLGNLMAQPTEQLSAEQLGQLLLRLRNRLACLVDRARGRAGLPPRTA
PRWATEAAAAACLAAMAAAEASAPQAPAAAAGGQEGAAGPVMVSVPFSREVLAEATACRVRSGTAAGARGNAPGAQGGVRKRTSKGGKAK
GGDREWSPEGEENTAPQPRGGGKRKSGAVAGGEEADGVASGRAKRASRPKRGSSKHDPYVDDNDYGDEGIDPFDVGDDLDDMNPHGRYG
NGGGRRADPSEAISALTAMGFTQSKARGALRECNFNVELAVEWLFANCL
>ORZ17038.1 [Absidia repens]
MSSPSSPSPIKPRSMLTGSRNVVRSNDSASLWNCTLSPGWNEEQSETLRHAVMKYGIGNWAKIIDSGYLPGKTNAQMNLQLQRLLGQQSTAEF
AGLHIDPKVIGEQNSRIQGPEIRRKNNTIVNTGDKLSREALRERILRNKEKYELPESVWQAIELEHVTDEDALLEEKKKTLREMKSQLKVVQRQIKN
LEFMHPLHAAKLKFELEKLAPSSSTSSSSSSPSPSSSSSPSSSSSKPSVSGTEEEMREAVDEERGSDEEIDELVEETDEEETSVSPKVGTRTKK
VRTN
>ORX56339.1 [Hesseltinella vesiculosa]
MIANSTATPKPRSMKAGARNVLRSNDSASLWNCTLSPGWTEQESEILRQLAIKFGIGNWAKIIESDCLPGKTNAQMNLQLQRLLGQQSTAEFAG
LHIDPKVIGEKNSKIQGPHIKRKNTTIVNTGGKLSREELRERQAKNKEMYEMPKSAWDSIDLDELRDMNSLKLKKKEDKDALKKQKLTQLKTKLTK
SQNNLKKVQAELKQIAMVDPERVAELKKELSRASSPLSNEVSVIEESPAKKQRTS
>ORX54764.1 [Piromyces finnis]
MVVEKDLAQENKIKEELNKKHEWVKEMRKKFCVRKEFENTKNLILEDGTLNQEYFRLSKGTVLKTNEVRKWTSIERNLLIKGIEKYGIGHFREISE
SLLPKWSGNDLRIKTIHLIGRQNLKLYKDWKGGEEDIKREYNRNKEIGLKCNAWKNNCLIDDGNGKVKEMIEATEPKH
>ORX84766.1 [Anaeromyces robustus]
MVVEKETNKENIKNIKEELDKKHAWVKEMRKKFCVRKEFENTKILILEDGTLNQDYFRLSKGTVLKTNEVRKWTSIERGLLIKGIEKYGIGHFREIS
ENLLPKWSGNDLRIKTIHLIGRQNLKLYKDWKGNEEDIKREYNRNKEIGLKCNAWKNNCLVDDGHGKVKAMIEATENN
>ORY98423.1 [Syncephalastrum racemosum]
MMTATDEDVDMKDVDIKLESNQETEQKILTPEEQKEKEKQDWIRQLRLKFCIRPEYEITKNMIFPDGTLNQDYFRPPKGAKVEEARKWTEVEKEL
LIQGIEKYGIGNFGEVSKALLPAWSTNDLRIKCIRLIGRQNLQLYRGWKGNADDIAREYNRNKELGLKYGTWKQGVLVYDDDGLVEKEILAQDAA
AKGEDVDMN
>XP_021686199.1 [Lobosporangium transversale]
MEINQEQLPSSSSILHPTSTSSSSSPSPSPSPASPKPERVFDARQRRINEIRLKFCIRDEFPITKNMIHPDGTLNQDYFRPPRGSKPVEVARKWTD
KERELLIKGIEKYGIGHFREISEEFLPLWSGNDLRIKTMRLVGRQNLQLYKDWKGNEQDLAREFELNKAIGLKYGAWKAGTLVADDDGLVAKAIEE
QWPGSNSGTGKTTAVIGISSEENSEVSTPLNDEDVDME
>ORY01319.1 [Basidiobolus meristosporus CBS 931.73]
MEVDQNDSSVAKETAEQPETPEISKELLERQEWIKNMRLQFCVRPEFEVTKNIIHEDGMLNQEYFLPPKGAKLEAEPERKWTETERNLLIQGIQQ
YGIGHFREISEALLPQWSGNDLRVKSMRLMGRQNLQLYKDWKGSIEDIEREYERNKAIGLKYNTWKNSTLVYDDAGLVLKAIEASEPKP
>ORZ26026.1 [Absidia repens]
MAIDSLQDTEDDRTNDQNDESRESSPTPLSPEEQAQKERHDWINQIRLKFCIRPEFEVTKNIIHPDGRLNQEYFHPPKGYKPEDARKWTETEKQL
LIKGIEEHGIGNFGLISKESLPKWSTNDLRVKCIRLIGRQNLQLYRGWKGNADDITREYERNKEIGLKYGTWKQGVLVYDDDGMVEKELLATAATP
ADSMSMEEDEDMATD
>ORX67568.1 [Linderina pennispora]
MDTASPDDGAIAQPMLGVEDADFWRQKQEWVKQMRLQFSRRPEFPETHNMIDDEGMLNQEYFQPPKDAVAPKERKWGDDEKRRLLEGIEKH
GIGHFREISEESLPEWSGNDLRMKAIRLMGRQNLQLYKGWKGDAAAIGLKHGTWKGGALVYDDDGVVLKAIQESNRANPP
>XP_001699352.1 [Chlamydomonas reinhardtii]
MAACSAACDSHVVPQPSPGSWGMPEDRDNYIVQMRRRYSPAGMLNADGSINQDFFKPRRVVLVADRAKWGDAEREGLYKGLEVHGVGKWR
EINRDYLKGQWDDQQVRIRAARLLGSQSLVRYMGWKGSKAKVDAEYAKNKAIGEATGCWKAGQLVEDDHGSVRKYFEAQQAGGEQ

Protein sequences for phylogenetic analysis of p2 proteins

>XP_001017830.3 [Tetrahymena thermophila SB210]
MNQMGVIAIKRKQSYQLNVKINYINTAHQIKKPCQYIQKCILFRLLYKFCKQLIPLNFNLFLIFYFYHLLFHLIFNYLLKFAKKINKLIRNQRKNREKKEA
FKHKKIQININHYNYLKQNIQQVGIIFQNKKSKLTLKLVQKKSLSEYYRKIKMKKNGKSQNQPLDFTQYAKNMRKDLSNQDICLEDGALNHSYFLTK
KGQYWTPLNQKALQRGIELFGVGNWKEINYDEFSGKANIVELELRTCMILGINDITEYYGKKISEEEQEEIKKSNIAKGKKENKLKDNIYQKLQQM
Q
>XP_001699352.1 [Chlamydomonas reinhardtii]
MAACSAACDSHVVPQPSPGSWGMPEDRDNYIVQMRRRYSPAGMLNADGSINQDFFKPRRVVLVADRAKWGDAEREGLYKGLEVHGVGKWR
EINRDYLKGQWDDQQVRIRAARLLGSQSLVRYMGWKGSKAKVDAEYAKNKAIGEATGCWKAGQLVEDDHGSVRKYFEAQQAGGEQ
>EJY77156.1 [Oxytricha trifallax]
MSTAKQQQAQQHLLPKRSNMRVGSVSNELDYAKRNYIIKMRQSFIEVNKNIYFEDGSLNFKYFNVKKGHYWSKEINEELIKGVIKYGATNYKDIKN
KMEIFKKEWSETEIRLRICRLLKCYNLKVYEGHKFNSREEILEQATLNKEEAIKQKKICGGILYNPPHEQDDGIMSSYFNLKNKNNTPVKASAQ
>ORZ26026.1[Absidia repens]
MAIDSLQDTEDDRTNDQNDESRESSPTPLSPEEQAQKERHDWINQIRLKFCIRPEFEVTKNIIHPDGRLNQEYFHPPKGYKPEDARKWTETEKQL
LIKGIEEHGIGNFGLISKESLPKWSTNDLRVKCIRLIGRQNLQLYRGWKGNADDITREYERNKEIGLKYGTWKQGVLVYDDDGMVEKELLATAATP
ADSMSMEEDEDMATD
>ORY96423.1 [Syncephalastrum racemosum]
MMTATDEDVDMKDVDIKLESNQETEQKILTPEEQKEKEKQDWIRQLRLKFCIRPEYEITKNMIFPDGTLNQDYFRPPKGAKVEEARKWTEVEKEL
LIQGIEKYGIGNFGEVSKALLPAWSTNDLRIKCIRLIGRQNLQLYRGWKGNADDIAREYNRNKELGLKYGTWKQGVLVYDDDGLVEKEILAQDAA
AKGEDVDMN
>XP_021886199.1 [Lobosporangium transversale]
MEINQEQLPSSSSILHPTSTSSSSSPSPSPSPASPKPERVFDARQRRINEIRLKFCIRDEFPITKNMIHPDGTLNQDYFRPPRGSKPVEVARKWTD
KERELLIKGIEKYGIGHFREISEEFLPLWSGNDLRIKTMRLVGRQNLQLYKDWKGNEQDLAREFELNKAIGLKYGAWKAGTLVADDDGLVAKAIEE
QWPGSNSGTGKTTAVIGISSEENSEVSTPLNDEDVDME
>ORY01319.1 [Basidiobolus meristosporus CBS 931.73]
MEVDQNDSSVAKETAEQPETPEISKELLERQEWIKNMRLQFCVRPEFEVTKNIIHEDGMLNQEYFLPPKGAKLEAEPERKWTETERNLLIQGIQQ
YGIGHFREISEALLPQWSGNDLRVKSMRLMGRQNLQLYKDWKGSIEDIEREYERNKAIGLKYNTWKNSTLVYDDAGLVLKAIEASEPKP
>ORX67588.1 [Linderina pennispora]
MDTASPDDGAIAQPMLGVEDADFWRQKQEWVKQMRLQFSRRPEFPETHNMIDDEGMLNQEYFQPPKDAVAPKERKWGDDEKRRLLEGIEKH
GIGHFREISEESLPEWSGNDLRMKAIRLMGRQNLQLYKGWKGDAAAIGLKHGTWKGGALVYDDDGVVLKAIQESNRANPP
>ORX84766.1 [Anaeromyces robustus]
MVVEKETNKENIKNIKEELDKKHAWVKEMRKKFCVRKEFENTKILILEDGTLNQDYFRLSKGTVLKTNEVRKWTSIERGLLIKGIEKYGIGHFREIS
ENLLPKWSGNDLRIKTIHLIGRQNLKLYKDWKGNEEDIKREYNRNKEIGLKCNAWKNNCLVDDGHGKVKAMIEATENN
>ORX54764.1 [Piromyces finnis]
MVVEKDLAQENKIKEELNKKHEWVKEMRKKFCVRKEFENTKNLILEDGTLNQEYFRLSKGTVLKTNEVRKWTSIERNLLIKGIEKYGIGHFREISE
SLLPKWSGNDLRIKTIHLIGRQNLKLYKDWKGGEEDIKREYNRNKEIGLKCNAWKNNCLIDDGNGKVKEMIEATEPKH
>ORX56334.1 [Hesseltinella vesiculosa]
MLAGDAELVEKPHNALNAEDTEMEDVDHSSHPDTTVDLSPEQLRLQEKQAWINQMRLKFCVREEFEITKNMIHPDGTLNQDYFKPPKKSKKKKS
KSKSKGTDETKDDTEAKGEDNKEDEDME
>PNW76495.1 [Chlamydomonas reinhardtii]
MAFAAALAEKRGPRVGDAASLWNFTPAPGWSREEVQILRLCLMKHGVGQWMQILSTGLLPGKLIQQLNGQTQRLLGQQSLAAYTGLKVDVDRI
RVDNETRTDATRKAGLIINDGPNLTKEMKEKMRQDAVAKYGLTPEQVAEVDEQLAEIAAAFNPASTSAAAGAGSGAAAAGQAAAAGSGAGGSG
QAATAADAGGAAGRGTGSAGGAAAAAPPRNALAISTGVLAATLLDASLGNLMAQPTEQLSAEQLGQLLLRLRNRLAGLVDRARGRAGLPPRTA
PRWATEAAAAACLAAMAAAEASAPQAPAAAAGGQEGAAGPVMVSVPFSREVLAEATACRVRSGTAAGARGNAPGAQGGVRKRTSKGGKAK
GGDREWSPEGEENTAPQPRGGGKRKSGAVAGGEEADGVASGRAKRASRPKRGSSKHDPYVDDNDYGDEGIDPFDVGDDLDDMNPHGRYG
NGGGRRADPSEAISALTAMGFTQSKARGALRECNFNVELAVEWLFANCL
>XP_001698763.1 [Chlamydomonas reinhardtii]
MAFAAALAEKRGPRVGDAASLWNFTPAPGWSREEVQILRLCLMKHGVGQWMQILSTGLLPGKLIQQLNGQTQRLLGQQSLAAYTGLKVDVDRI
RVDNETRTDATRKAGLIINDGPNLTKEMKEKMRQDAVAKYGLTPEQVAEVDEQLAEIAAAFNPASTSAAAGAGSGAAAAGQAAAAGSGAGGSG
NLMAQPTEQLSAEQLGQLLLRLRNRLACLVDRARGRAGLPPRTAPRWATEAAAAACLAAMAAAEASAPQAPAAAAGGQEGAAGPVMVSVPFS
REVLAEATACRVRSGTAAGARGNAPGAQGGVRKRTSKGGKAKGGDREWSPEGEENTAPQPRGGGKRKSGAVAGGEEADGVASGRAKRAS
RPKRGSSKHDPYVDDNDYGDEGIDPFDVGDDLDDMNPHGRYGNGGGRRADPSEAISALTAMGFTQSKARGALRECNFNVELAVEWLFANCL
>XP_011237366.1 [Mus musculus]

MPRRQAEAMDIDAEREKITQEIQELERILYPGSTSVHFEVSESSLSSDSEADSLPDEDLETAGAPILEEEGSSESSNDEEDPKDKALPEDPETCLQ
LNMVYQEVIREKLAEVSQLLAQNQEQQEEILFDLSGTKCPKVKDGRSLPSYMYIGHFLKPYFKDKVTGVGPPANEETREKATQGIKAFEQLLVTK
WKHWEKALLRKSVVSDRLQRLLQPKLLKLEYLHEKQSRVSSELERQALEKQIKEAEKEIQDINQLPEEALLGNRLDSHDWEKISNINFEGARSAE
EIRKFWQSSEHPSISKQEWSTEEVERLKAIAATHGHLEWHLVAEELGTSRSAFQCLQKFQQYNKTLKRKEWTEEEDHMLTQLVQEMRVGNHIP
YRKIVYFMEGRDSMQLIYRWTKSLDPSLKRGFWAPEEDAKLLQAVAKYGAQDWFKIREEVPGRSDAQCRDRYIRRLHFSLKKGRWNAKEEQQ
LIQLIEKYGVGHWARIASELPHRSGSQCLSKWKILARKKQHLQRKRGQRPRHSSQWSSSGSSSSSSEDYGSSSGSDGSSGSENSDVELEASLE
KSRALTPQQYRVPDIDLWVPTRLITSQSQREGTGCYPQHPAVSCCTQDASQNHHKEGSTTVSAAEKNQLQVPYETHSTVPRGDRFLHFSDTHS
ASLKDPACKPVLKVPLEKMPKLIRTRPPTQSHTLMKERPKQPLLPSSRSGSDPGNNTAGPHLRQLWHGTYQNKQRRKRQALHRRLLKHRLLLA
VIPWVGDINLACTQAPRRPATVQTKADSIRMQLECARLASTPVFTLLIQLLQIDTAGCMEVVRERKSQPPALLQPGTRNTQPHLLQASSNAKNNT
GCLPSMTGEQTAKRASHKGRPRLGSCRTEATPFQVPVAAPRGLRPKPKTVSELLREKRLRESHAKKATQALGLNSQLLVSSPVILQPPLLPVPH
GSPVVGPATSSVELSVPVAPVMVSSSPSGSWPVGGISATDKQPPNLQTISLNPPHKGTQVAAPAAFRSLALAPGQVPTGGHLSTLGQTSTTSQ
KQSLPKVLPILRAAPSLTQLSVQPPVSGQPLATKSSLPVNWVLTTQKLLSVQVPAVVGLPQSVMTPETIGLQAKQLPSPAKTPAFLEQPPASTDT
EPKGPQGQEIPPTPGPEKAALDLSLLSQESEAAIVTWLKGCQGAFVPPLGSRMPYHPPSLCSLRALSSLLLQKQDLEQKASSLAASQAAGAQPD
PKAGALQASLELVQRQFRDNPAYLLLKTRFLAIFSLPAFLATLPPNSIPTTLSPDVAVVSESDSEDLGDLELKDRARQLDCMACRVQASPAAPDPV
QSHLVSPGQRAPSPGEVSAPSPLDASDGLDDLNVLRTRRARHSRR

>XP_006497968.1 [Mus musculus]

MPRRQAEAMDIDAEREKITQEIQELERILYPGSTSVHFEVSESSLSSDSEADSLPDEDLETAGAPILEEEGSSESSNDEEDPKDKALPEDPETCLQ
LNMVYQEVIREKLAEVSQLLAQNQEQQEEILFDLSGTKCPKVKDGRSLPSYMYIGHFLKPYFKDKVTGVGPPANEETREKATQGIKAFEQLLVTK
WKHWEKALLRKSVVSDRLQRLLQPKLLKLEYLHEKQSRVSSELERQALEKQIKEAEKEIQDINQLPEEALLGNRLDSHDWEKISNINFEGARSAE
EIRKFWQSSEHPSISKQEWSTEEVERLKAIAATHGHLEWHLVAEELGTSRSAFQCLQKFQQYNKTLKRKEWTEEEDHMLTQLVQEMRVGNHIP
YRKIVYFMEGRDSMQLIYRWTKSLDPSLKRGFWAPEEDAKLLQAVAKYGAQDWFKIREEVPGRSDAQCRDRYIRRLHFSLKKGRWNAKEEQQ
LIQLIEKYGVGHWARIASELPHRSGSQCLSKWKILARKKQHLQRKRGQRPRHSSQWSSSGSSSSSSEDYGSSSGSDGSSGSENSDVELEASLE
KSRALTPQQYRVPDIDLWVPTRLITSQSQREGTGCYPQHPAVSCCTQDASQNHHKEGSTTVSAAEKNQLQVPYETHSTVPRGDRFLHFSDTHS
ASLKDPACKSHTLMKERPKQPLLPSSRSGSDPGNNTAGPHLRQLWHGTYQNKQRRKRQALHRRLLKHRLLLAVIPWVGDINLACTQAPRRPAT
VQTKADSIRMQLECARLASTPVFTLLIQLLQIDTAGCMEVVRERKSQPPALLQPGTRNTQPHLLQASSNAKNNTGCLPSMTGEQTAKRASHKGR
PRLGSCRTEATPFQVPVAAPRGLRPKPKTVSELLREKRLRESHAKKATQALGLNSQLLVSSPVILQPPLLPVPHGSPVVGPATSSVELSVPVAPV
MVSSSPSGSWPVGGISATDKQPPNLQTISLNPPHKGTQVAAPAAFRSLALAPGQVPTGGHLSTLGQTSTTSQKQSLPKVLPILRAAPSLTQLSV
QPPVSGQPLATKSSLPVNWVLTTQKLLSVQVPAVVGLPQSVMTPETIGLQAKQLPSPAKTPAFLEQPPASTDTEPKGPQGQEIPPTPGPEKAAL
DLSLLSQESEAAIVTWLKGCQGAFVPPLGSRMPYHPPSLCSLRALSSLLLQKQDLEQKASSLAASQAAGAQPDPKAGALQASLELVQRQFRDN
PAYLLLKTRFLAIFSLPAFLATLPPNSIPTTLSPDVAVVSESDSEDLGDLELKDRARQLDCMACRVQASPAAPDPVQSHLVSPGQRAPSPGEVSA
PSPLDASDGLDDLNVLRTRRARHSRR

>EJY86254.1 [Oxytricha trifallax]

MSVHHKMADSKSLHNYTLSPGWTREEVDILKIALMKFGIGKWKKIQKSGCLPSKTISQMNLQTQRLLGQQSLAEFMGLHVYLDRVFRDNSLKTG
PEIQRKNNFIINTGNNLTQPEKEKRLRLNKQKYGLDLAFIKTLRLPKPESATGGKREAILSMDQIFAQKSHFTVVEKLKHLEALKNALCSKLGKIERR
RRNKELSKIYRPLCQLIVVQKNADDQYEFVDIIDENE

>ORX69504.1 [Linderina pennispora]

MSSATPYAPRSMPTGQRNVVRSNDSASLWNCTLSPGWTQEEVQVLRKALMKFGVGNWMKIIESECLPGKTIAQMNLQTQRMLGQQSTAEFN
GLHLDAFVIGELNSKKQGPGIKRKNNCIVNTGGKLTRDEVVKRQQKHREQYEVKAEVWRAIVLPKPDNPLILLEKKREELKKVRLELEEIMKQIEE
TEKLVDVPEHAPGTKRARE

>NP_003077.2 [Homo sapiens]

MDVDAEREKITQEIKELERILDPGSSGSHVEISESSLESDSEADSLPSEDLDPADPPISEEERWGEASNDEDDPKDKTLPEDPETCLQLNMVYQE
VIQEKLAEANLLLAQNREQQEELMRDLAGSKGTKVKDGKSLPPSTYMGHFMKPYFKDKVTGVGPPANEDTREKAAQGIKAFEELLVTKWKNWE
KALLRKSVVSDRLQRLLQPKLLKLEYLHQKQSKVSSELERQALEKQGREAEKEIQDINQLPEEALLGNRLDSHDWEKISNINFEGSRSAEEIRKF
WQNSEHPSINKQEWSREEEERLQAIAAAHGHLEWQKIAEELGTSRSAFQCLQKFQQHNKALKRKEWTEEEDRMLTQLVQEMRVGSHIPYRRIV
YYMEGRDSMQLIYRWTKSLDPGLKKGYWAPEEDAKLLQAVAKYGEQDWFKIREEVPGRSDAQCRDRYLRRLHFSLKKGRWNLKEEEQLIELIE
KYGVGHWAKIASELPHRSGSQCLSKWKIMMGKKQGLRRRRRRARHSVRWSSTSSSGSSSGSSGGSSSSSSSSSEEDEPEQAQAGEGDRAL
LSPQYMVPDMDLWVPARQSTSQPWRGGAGAWLGGPAASLSPPKGSSASQGGSKEASTTAAAPGEETSPVQVPARAHGPVPRSAQASHSAD
TRPAGAEKQALEGGRRLLTVPVETVLRVLRANTAARSCTQKEQLRQPPLPTSSPGVSSGDSVARSHVQWLRHRATQSGQRRWRHALHRRLLN
RRLLLAVTPWVGDVVVPCTQASQRPAVVQTQADGLREQLQQARLASTPVFTLFTQLFHDTAGCLEVVRERKALPPRLPQAGARDPPVHLLQAS
SSAQSTPGHLFPNVPAQEASKSASHKGSRRLASSRVERTLPQASLLASTGPRPKPKTVSELLQEKRLQEARAREATRGPVVLPSQLLVSSSVIL
QPPLPHTPHGRPAPGPTVLNVPLSGPGAPAAAKPGTSGSWQEAGTSAKDKRLSTMQALPLAPVFSEAEGTAPAASQAPALGPGQISVSCPES
GLGQSQAPAASRKQGLPEAPPFLPAAPSPTPLPVQPLSLTHIGGPHVATSVPLPVTWVLTAQGLLPVPVPAVVSLPRPAGTPGPAGLLATLLPPL
TETRAAQGPRAPALSSSWQPPANMNREPEPSCRTDTPAPPTHALSQSPAEADGSVAFVPSEAQVAREIPEPRTSSHADPPEAEPPWSGRLPA
FGGVIPATEPRGTPGSPSGTQEPRGPLGLEKLPLRQPGPEKGALDLGLLSQEGEAATQGWLGGQRGVRVPLLGSRL
PYQPPALCSLRALSGLLLHKKALEHKATSLVVGGEAERPAGALQASLGLVRGQLQDNPAYLLLRARFLAAFTLPALLATLAPQGVRTTLSVPSRV
GSESEDEDLLSELELADRDGQPGCTTATCPIQGAPDSGKCSASSCLDTSNDPDDLDVLRTRHARHTRKRRRLV

>XP_016870547.1 [Homo sapiens]

MDVDAEREKITQEIKELERILDPGSSGSHVEISESSLESDSEADSLPSEDLDPADPPISEEERWGEASNDEDDPKDKTLPEDPETCLQLNMVYQE
VIQEKLAEANLLLAQNREQQEELMRDLAGSKGTKVKDGKSLPPSTYMGHFMKPYFKDKVTGVGPPANEDTREKAAQGIKAFEELLVTKWKNWE
KALLRKSVVSDRLQRLLQPKLLKLEYLHQKQSKVSSELERQALEKQGREAEKEIQDINQLPEEALLGNRLDSHDWEKISNINFEGSRSAEEIRKF
WQNSEHPSINKQEWSREEEERLQAIAAAHGHLEWQKIAEELGTSRSAFQCLQKFQQHNKALKRKEWTEEEDRMLTQLVQEMRVGSHIPYRRIV
YYMEGRDSMQLIYRWTKSLDPGLKKGYWAPEEDAKLLQAVAKYGEQDWFKIREEVPGRSDAQCRDRYLRRLHFSLKKGRWNLKEEEQLIELIE
KYGVGHWAKIASELPHRSGSQCLSKWKIMMGKKQGLRRRRRRARHSVRWSSTSSSGSSSGSSGGSSSSSSSSSSEEDEPEQAQAGEGDRAL
LSPQYMVPDMDLWVPARQSTSQPWRGGAGAWLGGPAASLSPPKGSSASQGGSKEASTTAAAPGEETSPVQVPARAHGPVPRSAQASHSAD
TRPAGAEKQALEGGRRLLTVPVETVLRVLRANTAARSCTQWLRHRATQSGQRRWRHALHRRLLNRRLLLAVTPWVGDVVVPCTQASQRPAVV
QTQADGLREQLQQARLASTPVFTLFTQLFHIDTAGCLEVVRERKALPPRLPQAGARDPPVHLLQASSSAQSTPGHLFPNVPAQEASKSASHKGS
RRLASSRVERTLPQASLLASTGPRPKPKTVSELLQEKRLQEARAREATRGPVVLPSQLLVSSSVILQPPLPHTPHGRPAPGPTVLNVPLSGPGAP
AAAKPGTSGSWQEAGTSAKDKRLSTMQALPLAPVFSEAEGTAPAASQAPALGPGQISVSCPESGLGQSQAPAASRKQGLPEAPPFLPAAPSPT
PLPVQPLSLTHIGGPHVATSVPLPVTWVLTAQGLLPVPVPAVVSLPRPAGTPGPAGLLATLLPPLTETRAAQGPRAPALSSSWQPPANMNREPE
PSCRTDTPAPPTHALSQSPAEADGSVAFVPGEAQVAREIPEPRTSSHADPPEAEPPWSGRLPAFGGVIPATEPRGTPGSPSGTQEPRGPLGLE
KLPLRQPGPEKGALDLEKPPLPQPGPEKGALDLGLLSQEGEAATGQWLGGQRGVRVPLLGSRLPYQPPALCSLRALSGLLLHKKALEHKATSLV
VGGEAERPAGALQASLGLVRGQLQDNPAYLLLRARFLAAFTLPALLATLAPQGVRTTLSVPSRVGSESEDEDLLSELELADRDGQPGCTTATCPI
QGAPDSGKCSASSCLDTSNDPDDLDVLRTRHARHTRKRRRLV
>XP_020936800.1 [Sus scrofa]

MDVDAEREKISKEIKELERILDPGSSGINDDVSESSLDSDSEAESLPDDDADATGPLLSEDERWGDASNDEDDAKERALPEDPETCLQLNMVYQ
EVVREKLAEVSLLLAQNREQQEEVSWALAGSGGRRVKDGRSPPARLYVGHFMKPYFKDKVTGAGPPANEDTREKAAQGVKAFEELLVTKWKS
WEKALLRKAVVSDRLQRLLQPKLLKLEYLQQKQSRATSDAERQALEKQVREAEKEVQDISQLPEEALLGHRLDSHDWEKIANVNFEGGRSAEET
RKFWQNHEHPSINKQEWSAQEVDRLKAIAAKHGHLRWQEIAEELGTRRSAFQCLQKYQQHNAALKRREWTQEEDRMLTQLVQAMGVGSHIPY
RRIAYYMEGRDSTQLIYRWTKSLDPALKKGLWAPEEDAKLLQAVAKYGEQDWFKIREEVPGRSDAQCRDRYLRRLRLSLKKGRWSAQEEERLL
ELIGKHGVGHWAKIASELPHRTDSQCLSKWKIMARKQQSRGRRRRRPLRRVCWSSSSEDSEDSGDSGGSSSSSSSSEDVEPEGAPEARADG
PAPPSAQHPVPDMDLWVPTRQSARVPWGVSPGAWPGHRSASPRPPEGSDVAPGEEAGRAQAPSETPSASLRGGGCPRSADARPSGSEGL
ADEGPRRPLTVPLETVLRVLRTNTAALCRALKEKLRRPRLLGSPLGPSPSDGSVARPRVQPRWRRRHALQRRLLERQLLMAVSPWVGDVTLPC
APWRPAVLHRRADGIGKQLQGARLASTPVFTLLIQLFRIDTAGCMEVVRERRAQPPALPSGGRVPSSARNSPGHLFQNGSARGAAKKSASHSG
GGGPQSAPAPSGPRPKPKTVSELLREKRLREARARKAAQGPAVLPPQGLLSSPAILQPLPPQQLPVSGAVLSGPGGPAVASPGAPGPWASAK
EGPPSLHALALAPASMAAGVTPAAPRAPALGPSQVPASCHLSSLGQSQAPATSRKQGLPEAPPFLPAAPSPIQLPVQPRSLTPALAAHTGASHV
VASTPLPVTWVLTAQGLLPVPAVVGLPRPAGPPGPEGLSGTPPPSLTETRAGRGPKQPPAHVSVGPDPPAKTPPTAQSPAEGDGDVAHGPGG
PSCPGEAQVAGEASVPRTLSPAKPLADHPEAEPCGSSQLPLPGGLSPGGAPTRHQGLERPPPPWPGPEKGAPDLRLLSQESEAAVRGWLTG
QRGVCVPPLASRLPYQPPTLCSLRALSGLLLHKKALEHRAASLVPSGAAGAQQAPLGQVRERLQSSPAYLLLKARFLAAFALPALLATLPPHGVP
TTLSAAAGVDSESDDDSLDELELADNGGPLSGWPSGRQAGPAAPTPTQGAPGEGSAAPGLDSDDLDILRTRHAWHARKRRRLV
>XP_021883515.1 [Lobosporangium transversale]

MSSGSTPRSMTAGARNILRSNDSASLWNYTVAPGWSMKEAEILRKALMKFGIGNWSKIIESNGLVGKTNAQMNLQTQRMLGQQSTAEFAGLHI
DPRVIGQKNSLIQGDHIRRKNGCIVNTGAKLSREEIRRRVAENKEQYELPEEEWSSIELPLPDDPHLLLEAKKSEKVRLELELKNVQRQIAMLRKV
GRKFETGSESPKTELDDDERDEFIEDQPLGKRARIEA
>EJY76686.1 [Oxytricha trifallax]

MRVYLKFCNRKQIHYTHTMSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLI
GQQSLAEFMGLHLDIDRIAADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLTAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKI
KHLIKLEDALKSKLEMIREGKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSAHQNRNSDLDEENDQSVMIIDESQQQNLTPKGKAQ
TMLTNQTQTMKKQADDSRDEQHLPLISTSASVSNPSSTSKSSALKLNSMKQSDTAIASMKPSSSGKKTKVDSSFVSKQSNQQSTSYSETNVDT
QNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVEDSE
>EJY73777.1 [Oxytricha trifallax]

MSHATSHGNSTEKDKKNSGNMVAESKSLWNYALSPGWTPQEVDVLKIALMKFGIGKWTIIDKSGILPTKTIQQCYLQTQRILGQQSLAEFMGLHV
DIDKIALDNRRKNGIRKMGFLVNQGGKLTPEEKAHYQEINRQKYGLSPEEVETIKLPPPCSVEIYDINKIINPKSKLTTIEKINHCIKLQDALLEKLENI
KNKKIPTGAGFSSSRVYENMRGYDPQLLLNSHVTGQLDHSMQDLTIDERYSDLDEEEDPLAMASIIDSQATPQPQKIKSSVPNKASTTPSAKEMN
QIKDIIDSVIAENSAQQSKNLAQEKPKLKFSLVKATESNLLQSAAQNSDDVVMEEDSKLQHIETFSTVTQTATDQSNSQSKSQNNIASDSLKDSLE
QNDLSKSLTDSLEMQQYSAEKKLNQAPMSKNSDKPKKKRLNKRKLPSDDEFETL
>EJY79729.1 [Oxytricha trifallax]

MSSSISAAIIAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWKTIEQSQCLPTKTMSQMYLQTQRLVGQQSLAEFMGLHLDLEQ
IFIKNAERQGAGVFRKNGCRNTGDNMTKVQIAKLRKKNSKIFGLTQPFVQSLHLPKAKVKEWLKVLTLDQILSAKSNFSTAEKIHYLKILENALERKL
KKILRLQELVSIYRPCNIGIVVQKRLGSSIGDEYFEYVDCVKIEEKSVGNLDFALPNRNTDSTSLNEDFSFLDSTQKPQKLKAGSGRENKRKKMRD
GLKDERAQRQSLMEALDEQEFDETKFQDSDGEMPDLNM
>EJY81929.1 [Oxytricha trifallax]

MSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLIGQQSLAEFMGLHLDIDRI
AADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLSAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKIKHLIKLEDALKSKLEMIRE
GKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSSHQNRNSDLDEENDESVMIIDESQQENLTPKGKAQAMLTHQKYNEVTQTMIKQ
GDDSRQQQHLPLDSTSASVSNPSSTSKSSTMKSNSMKQSETAIASMKPSSIGKKTKVDSSFVTKQSNQQSTAPIQKQAHQQNLDRNRSELGST
FAQQASVDTQNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVEDSE
>EJY80746.1 [Oxytricha trifallax]

MRVYLKFCNRKQIHYTHTMSSSISAAIMAGNQNKKIAESKSLWNYALSPGWTQQEVEILKIALMKFGVGRWSAINKSGVLPTKQIQQCYLQTQRLI
GQQSLAEFMGLHLDIDRIAADNKQKRGIRKQGFLVNQGCKLTPEEKDELRKINQEKYGLTAEHVEAIKLPAPCHLVEIFQIDKIMHPRSTLSTMDKI
KHLIKLEDALKSKLEMIREGKRQQKFEQLQQKLKTTEASGRGSVTRVQRQMSDLHLGSAHQNRNSDLDEENDQSVMIIDESQQQNLTPKGKAQ
TMLTNQTQTMKKQADDSREEQHLPLNSTSASVSNPSSTSKSSALKLNSMKQSDTAIASMKPSSSGKKTKVDSSFVSKQSNQQSTGPIGKQAHQ
QNLDRNRSELGSTFAQQTNVDTQNSNNQGTSTASGNFISQSDDEEALMPKLKRRRVKDSE

>ORX78557.1 [Basidiobolus meristosporus CBS 931.73]

MTDVYKPRSMPVGARNVLRSNDSASLWNCTLSPGWTEPEVHILRKAVMKFGIGNWAKIIESQCLFGKTIAQMNLQLQRMLGQQSTAEFAGLHL
DPFVIGEINSKKQGPGIKRKNNCIVNTGGKLTREEIKRRLLEHKRTYEISEEEWRSIELPKPEDPGAVLIAKKDELKMLEDELLRVVQKIQKAREER
RSKSVDSSSVDGSVDDEARETKRRRK

>ORX79686.1 [Anaeromyces robustus]

MSIPKPRSMPTGFRNILRPNDSTSLWNCTLSPGWTQEESDILRDALIYYGIGNWKDIIEHGCLPDKTNAQMNLQLQRMLGQQSTAEFQNLHIDPY
VIGKINSQKQGPNIRRKNGFIINTGGKLSREDIRRKIQENKENYELPKEEWSKIVLPNREVVIKNKVQEAINEKREKLNKLEDELDSVLKAIVNRRRE
LRGMIPLKDSEMKSLVNRSAKNEGENKTETTNNEESNNTNNSDDIKDENNETSTSSHIFTNNDNELSENNSSSSSSNSISNKKKRFLRREVRRGK
RRYNYDDDDFMPSGNRSRKSRKI

>ORX58566.1 [Piromyces finnis]

MSIPKPRSMPVGFRNILRPNDSTSLWNCTLSPGWTQEESDILRDALIFYGIGNWKDIIEHGCLPDKTNAQMNLQLQRMLGQQSTAEFQNLHIDPY
EIGKINSQKQGPNIRRKNGFIINTGGKLSREDIKRKIQENKENYELPEEVWSKIVLPNREVVTINEKRQKLNKLEEELDSVLKQIVNRRRELRGMTP
LKETEMKSIVNRSNQNDTKTEEKEIKEEESTTVNEEKIENTETSSISIISTNENEQSENISSSSPIVKSEQKKKRVVSRRKNKRRVNSDDEDFLPPG
KSRSKRTRRTPKKSSN

>XP_001009903.1 [Tetrahymena thermophila SB210]

MSLKKGKFQHNQSKSLWNYTLSPGWREEEVKILKSALQLFGIGKWKKIMESGCLPGKSIGQIYMQTQRLLGQQSLGDFMGLQIDLEAVFNQNM
KKQDVLRKNNCIINTGDNPTKEERKRRIEQNRKIYGLSAKQIAEIKLPKVKKHAPQYMTLEDIENEKFTNLEILTHLYNLKAEIVRRLAEQGETIAQPS
IIKSLNNLNHNLEQNQNSNSSTETKVTLEQSGKKKYKVLAIEETELQNGPIATNSQKKSINGKRKNNRKINSDSEGNEEDISLEDIDSQESEINSEEI
VEDDEEDEQIEEPSKIKKRKKNPEQESEEDDIEEDQEEDELVVNEEEIFEDDDDDEDNQDSSEDDDDDED

>XP_020936799.1 [Sus scrofa]

MDVDAEREKISKEIKELERILDPGSSGINDDVSESSLDSDSEAESLPDDDADATGPLLSEDERWGDASNDEDDAKERALPEDPETCLQLNMVYQ
EVVREKLAEVSLLLAQNREQQEEVSWALAGSGGRRVKDGRSPPARLYVGHFMKPYFKDKVTGAGPPANEDTREKAAQGVKAFEELLVTKWKS
WEKALLRKAVVSDRLQRLLQPKLLKLEYLQQKQSRATSDAERQALEKQVREAEKEVQDISQLPEEALLGHRLDSHDWEKIANVNFEGGRSAEET
RKFWQNHEHPSINKQEWSAQEVDRLKAIAAKHGHLRWQEIAEELGTRRSAFQCLQKYQQHNAALKRREWTQEEDRMLTQLVQAMGVGSHIPY
RRIAYYMEGRDSTQLIYRWTKSLDPALKKGLWAPEEDAKLLQAVAKYGEQDWFKIREEVPGVTFEARAFPASRQRTSLPCAPLWPPALWVSRL
GNRRGGROPRGFSRTPRSVCRRYLRRLRLSLKKGRWSAQEEERLLELIGKHGVGHWAKIASELPHRTDSQCLSKWKIMARKQQSRGRRRRRP
LRRVCWSSSSEDSEDSGDSGGSSSSSSSSEDVEPEGAPEARADGPAPPSAQHPVPDMDLWVPTRQSARVPWGVGPGAWPGHRSASPRPP
EGSDVAPGEEAGRAQAPSETPSASLRGGGCPRSADARPSGSEGLADEGPRRPLTVPLETVLRVLRTNTAALCRALKEKLRRPRLLGSPLGPSP
SDGSVARPRVQPRWRRRHALQRRLLERQLLMAVSPWVGDVTLPCAPWRPAVLHRRADGIGKQLQGARLASTPVFTLLIQLFRIDTAGCMEVVR
ERRAQPPALPSGGRVPSSARNSPGHLFQNGSARGAAKKSASHSGGGGPQSAPAPSGPRPKPKTVSELLREKRLREARARKAAQGPAVLPPQ
GLLSSPAILQPLPPQQLPVSGAVLSGPGGPAVASPGAPGPWASAKEGPPSLHALALAPASMAAGVTPAAPRAPALGPSQVPASCHLSGLGQSQ
APATSRKQGLPEAPPFLPAAPSPIQLPVQPRSLTPALAAHTQASHVVASTPLPVTWVLTAQGLLPVPAVVGLPRPAGPPDPEGLSGTPPPSLTET
RAGRGPKQPPAHVSVGPDPPAKTPPTAQSPAEGDGDVAHGPGGPSCPGEAQVAGEASVPRTLSPAKPLADHPEAEPCGSSQLPLPGGLSPG
GAPTRHQGLERPPPPWPGPEKGAPDLRLLSQESEAAVRGWLTGQRGVCVPPLASRLPYQPPTLCSLRALSGLLLHKKALEHRAASLVPSGAA
GAQGAPLGQVRERLQSSPAYLLLKARFLAAFALPALLATLPPHGVPTTLSAAAGVDSESDDDSLDELELADNGGPLGGWPSGRQAGPAAPTPT
QGAPGEGSAAPGLDSDDLDILRTRHAWHARKRRRLV

>XP_009300052.1 [Danio rerio]

MKCLSVNMTHLSRDSWLYTHDVQVTYNSFIKVSPCPKMASDDLRAQRDKIQREILALESTLGADSSIADQLSSDNSSDYESDDSGPTVKRVERD
DLETERLRIQREIEELENALGADAALENVLQDSDHDTDSSEDSADDLELPQNVETCLQMNLVYQEVLKEKLAELEQLLIENQGQQKEIEVQLSGP
GNSIFSVPGVPPQKQFLGYFLKPYFKDKLTGLGPPANEETKERMKHGSIPVDNLKIKRWEGWQKTLLTNAVARDTMKRMLQPKLSKMEYLSNKL
CRAEGEEKEQLKAQIELIEKQIAEIRTLKDDQLLGDLQDDHDWDKISNIDFEGLRQADDLKRFWQNFLHPSINKSVWKQDEIYKLQAVAEEFKMC
HWDKIAEALGTNRTAFMCFQTYQRYISKTFRRTHWTEEEDDLLRELVEKMRIGNFIPYIQMSHFMVGRDGSQLAYRWTSVLDPSLKKGPWSKE
EDQLLRNAVAKYGTREWGRIRTEVPGRTDSACRDRYLDCLRETVKKGTWSYAEMELLKEKVAKYGVGKWAKIASEIPNRVDAQCLHKWKLMT
RSKKPLKRPLSSITTSYPRNKRQKLLKTVKEEMFFNSSSDDESQINYMNSDESDDLAEDENLEIPQKEYVQTEMKEWIPRNAMVWTITPGSFRTL
WVRLPTNEEELRESTKESGLGSDSSENSACPNDEPIMERNTILDRFGDVERTYVGMNTVVLHRRTDDEKAMFKVCMSDVKQFIQMKATEFAVK
KKKKIKNKKRTLRDVFSLNTDLQKAVIPWIGNVIISTPANEAIFCEGDIVGIKAASIRLQKTSVFTFFIKAFHVDVNGCRTVIEIHKKLDIKMPLAINGNP
KPTPISTSPKTVAVLLQQSKAASEHKKPAEPSQQPSLPPSQKPSLPPAQQPTQPPSLPPSVPPSQQPTLPPPSQPSQPPPQPPSLPPSQPPAQ
QPPQQPSLPPPQPPSLPPPQPPSLPTSQQQSLPPSQQHSLPPFQNPSLPPSQQPSLPPSKQPPQPLPVRQITTPTLIYPNNLVITNPNMEGEVQ
HLVFKGLLLPQQPSKAVSHIPLPVMQPKTPAQPIVVSKSPSVQDSNSVKSSKRICKPTKKAQALMEQSKVKSRKKEPQKQNQGNKNVVFPTVTL
QTSPVIKILSPARLVQVTGLSPNFSSNQTINMPDKSLTIKSPQPCSSGNLHQSAPVVVHSSTNPTFVHSSVSNVSRDNLNVSSTINISPRVSRDAL
NPTSFLNSTTFPLPQNLSVQQSVQIVPQIPINVVHKATCTKAAKTSSDSSSDESVVKQHQLSPSTGRSIPPAVFNIQPNPSTPPTLSSGPVIFNPNN
KVVAPKLCGLNVSSSQLPTVSTGKTKYRPIRPLGPLPVVAPPSRKVTSMSRIRAQSEGEPLISLRDLPAAGVNFDSHLIFPEKSSEVDDWMDGKG
GIPLPHLDTSLPYLPPSAATIKTMTDLLRAKQPLLLAAKKVLPAQYQDECNEEVEVEAIRKVVAERFASNPAYLLCKARFLSCFTLPALLATINPCEE
RQLLSEDDEEDDHLATINPSEEHQSSTEDDEEDLQTNERSQPPTARTELNMNENEASAKQFSGIGPKRQRNQRIKRLIK

Table 2. Primer sequences.

Name	Sequence (5' to 3')	Description
1781.0_qF24	ACTAGTCTTAAATATGAGAAAGATGATTTGAATAAGAT	Contig1781.0 tiling qPCR primers
1781.0_qR24	ATCCTAGCAATATTATCTACTTATAATTCTATTGACTATTAG	Contig1781.0 tiling qPCR primers
1781.0_qF23	CTAATTAACTAATAGTCAATAGAATTATAAGTAGATAATATTGCT	Contig1781.0 tiling qPCR primers
1781.0_qR23	CATTAAATCATTAACAGAGTAATGTCGTCATATATTTGTC	Contig1781.0 tiling qPCR primers
1781.0_qF22	TTTAGTGAGCATAGACAAATATATGACGACATTACTC	Contig1781.0 tiling qPCR primers
1781.0_qR22	GCGGAGATGTCTTTTTGACCTTTTGATAG	Contig1781.0 tiling qPCR primers
1781.0_qF21	ATGTTAACATGCTTATTATTACTATCAAAAGGTCAA	Contig1781.0 tiling qPCR primers
1781.0_qR21	GGCTGCTACTGATATTTATGTTCTTTATGTTTA	Contig1781.0 tiling qPCR primers
1781.0_qF20	CAAAGAACACGAAGCTCATAAACATAAAGAACAT	Contig1781.0 tiling qPCR primers
1781.0_qR20	TGGAGCAAATGCTGCTAATAACGAG	Contig1781.0 tiling qPCR primers
1781.0_qF19	ACCTCCAGCAGCTCCGTTTCTATTATTTG	Contig1781.0 tiling qPCR primers
1781.0_qR19	GGCCTGGGTATTTTCCCTGCTTTA	Contig1781.0 tiling qPCR primers
1781.0_qF18	CTTCCCAGGTAAAATTTAAGGTAAATAAAGCAGG	Contig1781.0 tiling qPCR primers
1781.0_qR18	TCAAGGTGGAGGACTCTTCGGTAAC	Contig1781.0 tiling qPCR primers
1781.0_qF17	ATTACGAACCCACTACCTGAATTATTGTTACCG	Contig1781.0 tiling qPCR primers
1781.0_qR17	AAACGTCCTGCAGGACAACGC	Contig1781.0 tiling qPCR primers
1781.0_qF16	TTGATTGAAGTTTTAATTTGGTACTGGGC	Contig1781.0 tiling qPCR primers
1781.0_qR16	TTGGATGCTGATCTGTTTTGTTTAGAAAG	Contig1781.0 tiling qPCR primers
1781.0_qF15	TTGGGATTTCTTAACTGGATTTCTTTCTAAAC	Contig1781.0 tiling qPCR primers
1781.0_qR15	CTGCTTAAATTAAGTACTTCTATGTTTGAAATTAATGTTC	Contig1781.0 tiling qPCR primers
1781.0_qF14	CAATTAAAACACGTTGAACATTAATTTCAAACATAG	Contig1781.0 tiling qPCR primers
1781.0_qR14	TGAGGATCCAAGGTAAATTTCATACAATC	Contig1781.0 tiling qPCR primers
1781.0_qF13	GACTGCATGTATATGCTAATGATTGTATGAAATTTAC	Contig1781.0 tiling qPCR primers
1781.0_qR13	AGTGGCATTTCCAAGGAAACATTAATAC	Contig1781.0 tiling qPCR primers
1781.0_qF12	CAGTGTTTCCCTTTGTGTAAATGGG	Contig1781.0 tiling qPCR primers
1781.0_qR12	TCAGTGGATAAACTAGCCTAAGGAAACAC	Contig1781.0 tiling qPCR primers
1781.0_qF11	TTTTACAGACTGGACACAGTAGTGTTTCC	Contig1781.0 tiling qPCR primers
1781.0_qR11	CCAGTGGTATCAACATGCGGTCATC	Contig1781.0 tiling qPCR primers
1781.0_qF10	GATATATACACTCCCAGCAGTAAAGATGACC	Contig1781.0 tiling qPCR primers
1781.0_qR10	GAATAGGCTCACTCTAAATTCGAGTGC	Contig1781.0 tiling qPCR primers
1781.0_qF9	ATTCGCTAGGTCTAAGCAAATATTGCAC	Contig1781.0 tiling qPCR primers

1781.0_qR9	TAAATAGCCAAAACAACCAATAAAATTAACAATAACCTC	Contig1781.0 tiling qPCR primers
1781.0_qF8	CTTTTTGAGGGCGAGGTTATTGTTAATTTTATTG	Contig1781.0 tiling qPCR primers
1781.0_qR8	GATCCATTAATTACAGAAATAAATAATAGGCAGCATA	Contig1781.0 tiling qPCR primers
1781.0_qF7	ATATTGCCTGAATTATTATGCTGCCTATTATTTATT	Contig1781.0 tiling qPCR primers
1781.0_qR7	AAATGTGCACCGTCATCAAATACC	Contig1781.0 tiling qPCR primers
1781.0_qF6	GGATCACTATAATCATCTGGATGACTATTGG	Contig1781.0 tiling qPCR primers
1781.0_qR6	AAGTGTAATGTAGTTTCAATGGTAGTGATGTG	Contig1781.0 tiling qPCR primers
1781.0_qF5	TGACTTCTTCCAGTGGATTCACATC	Contig1781.0 tiling qPCR primers
1781.0_qR5	GCCAATTAATTCATTTGTTCGTAGAGATATGTAA	Contig1781.0 tiling qPCR primers
1781.0_qF4	CACTTTATAATAAATAAGAATTATTACATATCTCTACGAACAA	Contig1781.0 tiling qPCR primers
1781.0_qR4	CTCACCAGTAATTTGCAGACACC	Contig1781.0 tiling qPCR primers
1781.0_qF3	GGCTGACTGGGGTTGAGTTAATC	Contig1781.0 tiling qPCR primers
1781.0_qR3	AATATAAACAAAATGGAATATACAAAACTTGAATAAGAAATAG	Contig1781.0 tiling qPCR primers
1781.0_qF2	GAGACTGAGGATCTATTTCTTATTCAAGTTTTG	Contig1781.0 tiling qPCR primers
1781.0_qR2	ATTAATACATTATTAACTTAAATATAAATATTTAAAGAATTATGAACAATAAT	Contig1781.0 tiling qPCR primers
1781.0_qF1	CATTTTGTTTATATTATTGTTCATAATTCTTTAAATATTTATATTTAAGTTAAT	Contig1781.0 tiling qPCR primers
1781.0_qR1	ACAAGATAACATTGCTAATTTTCAATAAATTAAATTAATACATT	Contig1781.0 tiling qPCR primers

1781.0_F	CCCCAAAACCCCAAAACCCCACTAGTCTTAAATATGAGAAAGATGATTTGAATAAG	Primer pair for amplifying chromosome, to be added to mini-genome
1781.0_R	CCCCAAAACCCCAAAACCCCACAAGATAACATTGCTAATTTTCAATAAATTAAAT	
15118.0_F	CCCCAAAACCCCAAAACCCCGATTTATGAAAGTGCTGTATTATTAAGGAATG	Primer pair for amplifying chromosome, to be added to mini-genome
15118.0_R	CCCCAAAACCCCAAAACCCCATTATTCCTACTTTTAGCTATATTAGAAATTCG	
1339.1_F	CCCCAAAACCCCAAAACCCCATGATGATACATAGATTCATTAAAATAAAAAAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
1339.1_R	CCCCAAAACCCCAAAACCCCTTAGATGAATTAAATAAAGAATTCAAATAAATAG	
20718.0_F	CCCCAAAACCCCAAAACCCCATGAATCTGAAATCGGGCAGTTGAATACG	Primer pair for amplifying chromosome, to be added to mini-genome
20718.0_R	CCCCAAAACCCCAAAACCCCATTTATCATAATTATAGAGAAGATAGTGATGC	
20822.0_F	CCCCAAAACCCCAAAACCCCATGAGAGTTTGTGAAAAATTAAGTTTG	Primer pair for amplifying chromosome, to be added to mini-genome
20822.0_R	CCCCAAAACCCCAAAACCCCTATATTAAATATCAAGAAAAAGTAAAAAGACAG	
21162.0_F	CCCCAAAACCCCAAAACCCCAAGTCTCATTTTGGTTAGTGATGTTTGGATTG	Primer pair for amplifying chromosome, to be added to mini-genome
21162.0_R	CCCCAAAACCCCAAAACCCCGTATGATCGATGAATACAAAATCAAGTTGGAAG	
11991.0_F	CCCCAAAACCCCAAAACCCCACTTAAAAGGATTGCATGATTGTAAGGGAAATGTG	Primer pair for amplifying chromosome, to be added to mini-genome
11991.0_R	CCCCAAAACCCCAAAACCCCAATAATCGCACTTACATTATATCTGGAGAAATG	

5079.0_F	CCCCAAAACCCCAAAACCCCTTCTACTAAATTTCATTGATTTTTTTCAATTTC	Primer pair for amplifying chromosome, to be added to mini-genome
5079.0_R	CCCCAAAACCCCAAAACCCCATTTGATAGAATAGAAGAGAAATTATGGAATG	
13865.0_F	CCCCAAAACCCCAAAACCCCAAGTATAAATAAGGGAGTTGATATATAATATACTT	Primer pair for amplifying chromosome, to be added to mini-genome
13865.0_R	CCCCAAAACCCCAAAACCCCATGAGAATTCCTATTCAAAAATGAAAAAGTAGATTG	
22365.0_F	CCCCAAAACCCCAAAACCCCATAAGGTAGTATATTTTTATTAAGGATTGGAAATTA	Primer pair for amplifying chromosome, to be added to mini-genome
22365.0_R	CCCCAAAACCCCAAAACCCCATAAGACTAAATTTATTGAAATTATCTTGTTAATAG	
21620.0_F	CCCCAAAACCCCAAAACCCCTTGAGCCAATACTGAAAAGGATGATAGTGAATAGTG	Primer pair for amplifying chromosome, to be added to mini-genome
21620.0_R	CCCCAAAACCCCAAAACCCCTCATTTTTTAAATTGGATAGTAAGAAAAATTATAATAAAG	
15049.0_F	CCCCAAAACCCCAAAACCCCAAGGAATAAAATTCAATTCCAAAATGTAAGGTGAG	Primer pair for amplifying chromosome, to be added to mini-genome
15049.0_R	CCCCAAAACCCCAAAACCCCGTTAAAAGAACCAAGTGATATATTATAAGCCA	
16562.0_F	CCCCAAAACCCCAAAACCCCTTTATCAATTATAAATAAAAAGTTTTAAGTCTATTTTTAA	Primer pair for amplifying chromosome, to be added to mini-genome
16562.0_R	CCCCAAAACCCCAAAACCCCATAAGACAAATGCAACTTTATAAAGTAAATAAATTATC	
22360.0_F	CCCCAAAACCCCAAAACCCCAATGCAACATTTACTTTTAACATTAGAGATTATC	Primer pair for amplifying chromosome, to be added to mini-genome
22360.0_R	CCCCAAAACCCCAAAACCCCATAAGAGCAAAAGTTAATATAAAAATTCAAGGTG	
15836.0_F	CCCCAAAACCCCAAAACCCCGATTTGCACAGTTAATTTGAATTTGGTATTTG	Primer pair for amplifying chromosome, to be added to mini-genome
15836.0_R	CCCCAAAACCCCAAAACCCCTCATTTTTAGTATTTTAAATATCATTTAGTTTTAAGTAA	
2324.0_F	CCCCAAAACCCCAAAACCCCTTGATTGATTCCTGAATACAAATGAAATAATATAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
2324.0_R	CCCCAAAACCCCAAAACCCCAAGACCAAAATAAAGAGGAATAATGAGAAGTAC	
22404.0_F	CCCCAAAACCCCAAAACCCCATGTAGAATTAATATGAGAACATCATTTTTTAAGC	Primer pair for amplifying chromosome, to be added to mini-genome
22404.0_R	CCCCAAAACCCCAAAACCCCATAATGTAAGAAATCTGATACAATAGAGAGATAAAC	
15403.0_F	CCCCAAAACCCCAAAACCCCGAATGGAAAATTTGTATGAAGTTCAGAGAGAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
15403.0_R	CCCCAAAACCCCAAAACCCCATAAGATTATCAGTTATAAAAATTGATAGGGGATG	
17795.0_F	CCCCAAAACCCCAAAACCCCATCATACGATATCTTAAGTGTTGATCTGAATTAAAT	Primer pair for amplifying chromosome, to be added to mini-genome
17795.0_R	CCCCAAAACCCCAAAACCCCGTTAGGTTTAAGAGTAGAAATAAAAGGAGATAAG	
11141.0_F	CCCCAAAACCCCAAAACCCCTCTCACTATCTTTTGTAAAAAGTTGGTAGAT	Primer pair for amplifying chromosome, to be added to mini-genome
11141.0_R	CCCCAAAACCCCAAAACCCCGTTGGTTTAGAATAAAGAATTGTATTAACCAAATTTAT	
22342.0_F	CCCCAAAACCCCAAAACCCCGTGAATTAAAATATAAACGAATAAGATATAAAGATTG	Primer pair for amplifying chromosome, to be added to mini-genome
22342.0_R	CCCCAAAACCCCAAAACCCCTTAATTACTGAATTGTTTATTATAAGATTATAAG	
2240.0_F	CCCCAAAACCCCAAAACCCCGTAATGAATAAATTGTAAAGGTAAATTGCAA	Primer pair for amplifying chromosome, to be added to mini-genome
2240.0_R	CCCCAAAACCCCAAAACCCCAATGGCAAACATTTAAAATAAATATTAATATAAATTAC	
3531.0_F	CCCCAAAACCCCAAAACCCCTAAAAGGAAAACAAATAGAAGAAACTGAA	Primer pair for amplifying chromosome, to be added to mini-genome
3531.0_R	CCCCAAAACCCCAAAACCCCATTTGGATATTATGATTAGCAGTTTAGTG	

4701.0_F	CCCCAAAACCCCAAAACCCCTTTAAATAAAAATCGCATGAATTAAATGCAAG	Primer pair for amplifying chromosome, to be added to mini-genome
4701.0_R	CCCCAAAACCCCAAAACCCCTAGGTAAATGCAAATTGGAGAATTTCCAATAG	
20883.0_F	CCCCAAAACCCCAAAACCCCATATTAAGAATTGTGTAATTTTTGAGTAAATTG	Primer pair for amplifying chromosome, to be added to mini-genome
20883.0_R	CCCCAAAACCCCAAAACCCCATTTAGTAGAATCTTCAATAAATAAGCGTTATTG	
15191.0_F	CCCCAAAACCCCAAAACCCCTAGCATTAAATTTGTAAAAAGAATGAAATTTAATAT	Primer pair for amplifying chromosome, to be added to mini-genome
15191.0_R	CCCCAAAACCCCAAAACCCCAATATACATGATTTTAGATAAACAACAAATAAT	
19342.0_F	CCCCAAAACCCCAAAACCCCATCAAGAATGGATTAGAATTTTTAATGCTTTGC	Primer pair for amplifying chromosome, to be added to mini-genome
19342.0_R	CCCCAAAACCCCAAAACCCCGAGGAACTAGGGATTACTCATTTTACTTCAG	
15245.0_F	CCCCAAAACCCCAAAACCCCATGCATGTAATTTTCTGTCAAAATTGAGTAAATAG	Primer pair for amplifying chromosome, to be added to mini-genome
15245.0_R	CCCCAAAACCCCAAAACCCCGTAAGCTAAATAAGTAGACTAAATAGGTAG	
6109.0_F	CCCCAAAACCCCAAAACCCCAACCGCAAATAGAATATATAAAGGATAATTTA	Primer pair for amplifying chromosome, to be added to mini-genome
6109.0_R	CCCCAAAACCCCAAAACCCCGAAGTACTAAAAATAAAAAGTAAAGTATTAAAATAAAATC	
22610.0_F	CCCCAAAACCCCAAAACCCCGTAGACAGATTTTCCAGTTTATAGCTGTGTTTG	Primer pair for amplifying chromosome, to be added to mini-genome
22610.0_R	CCCCAAAACCCCAAAACCCCTTTATGAATTTTCTTAAATCTGTAAATAAATAAAATAAT	
11875.0_F	CCCCAAAACCCCAAAACCCCGTATGTTAATTTTATGCTTTAAATGATAGTTTA	Primer pair for amplifying chromosome, to be added to mini-genome
11875.0_R	CCCCAAAACCCCAAAACCCCTGGATTCCATTTTGAAGAATAATTTATTAAC	
15329.0_F	CCCCAAAACCCCAAAACCCCTTGTTTGGATTATATTCAAAATAGGAAATTTAG	Primer pair for amplifying chromosome, to be added to mini-genome
15329.0_R	CCCCAAAACCCCAAAACCCCATGAATTTCAATAACTTTTTATGAAAATGAATTTA	
20179.0_F	CCCCAAAACCCCAAAACCCCTAGGAAGAAAATCTTGTGTGCAATTTGAGATTAAC	Primer pair for amplifying chromosome, to be added to mini-genome
20179.0_R	CCCCAAAACCCCAAAACCCCTTGATAAAAACATAGATTAAATACTAGTGTATAAA	
9936.0_F	CCCCAAAACCCCAAAACCCCATATGGAATATTTAATTTGATTTAAATGAAACGAAATA	Primer pair for amplifying chromosome, to be added to mini-genome
9936.0_R	CCCCAAAACCCCAAAACCCCTTGTAACAGTAAATAGAATATTTTAATTACCAAAAC	
16267.0_F	CCCCAAAACCCCAAAACCCCTCATTTTAGAATTATCTGTACTTAATTATTTTG	Primer pair for amplifying chromosome, to be added to mini-genome
16267.0_R	CCCCAAAACCCCAAAACCCCATGAGCATGTTATTTTACTTCATTAGTCAATTTG	
4486.0_F	CCCCAAAACCCCAAAACCCCATGAAATGAATTCTAAGATTGAATTGCATG	Primer pair for amplifying chromosome, to be added to mini-genome
4486.0_R	CCCCAAAACCCCAAAACCCCAGAAGAGATCAATAAATTGAGAAGGAATTG	
8551.0_F	CCCCAAAACCCCAAAACCCCGTGTTACAATTTGCGTTTGAAATAGTTGGTTGATA	Primer pair for amplifying chromosome, to be added to mini-genome
8551.0_R	CCCCAAAACCCCAAAACCCCATATGGTAAAAATTGAAGAAAGAAATTCAAGAGAA	
11748.0_F	CCCCAAAACCCCAAAACCCCGTATTGATGATAAAATTGTATACAAGTTGATAG	Primer pair for amplifying chromosome, to be added to mini-genome
11748.0_R	CCCCAAAACCCCAAAACCCCTAGATGCTTAATTATTAAGAAGATTCTGGAATG	
22291.0_F	CCCCAAAACCCCAAAACCCCATAAACCAATGTAATTAATTTATTGGGTGTGTTG	Primer pair for amplifying chromosome, to be added to mini-genome
22291.0_R	CCCCAAAACCCCAAAACCCCTTAGATTAAATTTAGAGAGTTATAGAAATGTAGTAAAT	

17536.0_F	CCCCAAAACCCCAAAACCCCATCTCAATTTATAAAATCAGAATAAGAGATTGTC	Primer pair for amplifying chromosome, to be added to mini-genome
17536.0_R	CCCCAAAACCCCAAAACCCCAGAATAAAACAACTGAAGTAAATATGAGTTAC	
15372.0_F	CCCCAAAACCCCAAAACCCCTTTCAAATATAAAATAAACAGAAGAATGGCAAACG	Primer pair for amplifying chromosome, to be added to mini-genome
15372.0_R	CCCCAAAACCCCAAAACCCCAAATTCAATATTAAATGAAATAATTTTCAAAAGTG	
13537.0_F	CCCCAAAACCCCAAAACCCCATGAGATCAAATTTTTTTTATTAAAATTCTTC	Primer pair for amplifying chromosome, to be added to mini-genome
13537.0_R	CCCCAAAACCCCAAAACCCCTTGGATTCATATTTTTGTTTAAGGCTTAGATA	
22613.0_F	CCCCAAAACCCCAAAACCCCATTAGAAAAGAGGATTTCAATAAAAGCAAATAT	Primer pair for amplifying chromosome, to be added to mini-genome
22613.0_R	CCCCAAAACCCCAAAACCCCATCGATTTATTATTGTTGAATTTAAAAGTATTGAA	
12585.0_F	CCCCAAAACCCCAAAACCCCGAGAGGTTTGATAAGTAGAATTAGTAAAATCTAT AAAG	Primer pair for amplifying chromosome, to be added to mini-genome
12585.0_R	CCCCAAAACCCCAAAACCCCATTAGTACTATTTTCATAGATCTATGTATAAATTG AA	
5317.0_F	CCCCAAAACCCCAAAACCCCAATGGAAAGATAAACAGATTTTAATTTGGAAATAA AAT	Primer pair for amplifying chromosome, to be added to mini-genome
5317.0_R	CCCCAAAACCCCAAAACCCCTTTAAGCAGTATTTCTAAAATGTTGATGAAATAAA AAT	
17894.0_F	CCCCAAAACCCCAAAACCCCATAAGATAAAATTTAACGAAAAAAAAGTTAAGTC	Primer pair for amplifying chromosome, to be added to mini-genome
17894.0_R	CCCCAAAACCCCAAAACCCCATAAGATGAAATATAGAGATAATTGAGCCTA	
3513.0_F	CCCCAAAACCCCAAAACCCCAATTACATATTAATGTACTTATGATAGAATG	Primer pair for amplifying chromosome, to be added to mini-genome
3513.0_R	CCCCAAAACCCCAAAACCCCTAATGATCAAATAACCTGAGTTAAAGAAG	
16420.0_F	CCCCAAAACCCCAAAACCCCAAATTATGAAAATAGACACTAATTGGATGTTC	Primer pair for amplifying chromosome, to be added to mini-genome
16420.0_R	CCCCAAAACCCCAAAACCCCTGATTCGTCATATGAAATTGAAAAGGAGTAAAT	
1084.1_F	CCCCAAAACCCCAAAACCCCAGCGCCATGAATCTGATGCATTTATTTTTAAG	Primer pair for amplifying chromosome, to be added to mini-genome
1084.1_R	CCCCAAAACCCCAAAACCCCGTAGATCATTTATGTAAAAGATTTTGAGAGATG	
22551.0_F	CCCCAAAACCCCAAAACCCCATACAATTATTATAAATGAAAAAGCGCACTAATC	Primer pair for amplifying chromosome, to be added to mini-genome
22551.0_R	CCCCAAAACCCCAAAACCCCATAGTTACTATGAAAGGACTGGTACATAGAAATA ATAG	
8670.0_F	CCCCAAAACCCCAAAACCCCTTAAGTCAATATCTAAATCAAATATTAGTAGTATA AT	Primer pair for amplifying chromosome, to be added to mini-genome
8670.0_R	CCCCAAAACCCCAAAACCCCGTCATATGGTTTTATAAAATAAAATTGAGATTTTT TTG	
19107.0_F	CCCCAAAACCCCAAAACCCCATAAGGATAAATTCTATCATATAAGTGGAAGTGC	Primer pair for amplifying chromosome, to be added to mini-genome
19107.0_R	CCCCAAAACCCCAAAACCCCATTCTTGAATATTGATTATGCATATTGTGTAAAAT AG	
21021.0_F	CCCCAAAACCCCAAAACCCCAAGCGTTGAATTTTTTATAATATATGATAAAC	Primer pair for amplifying chromosome, to be added to mini-genome
21021.0_R	CCCCAAAACCCCAAAACCCCTTAATGCCAATAAACAGATGAAAGTAGAGTTATA G	
15004.0_F	CCCCAAAACCCCAAAACCCCATAGAGAGTGTTTTATTGAAGGACAGAGAATATT G	Primer pair for amplifying chromosome, to be added to mini-genome
15004.0_R	CCCCAAAACCCCAAAACCCCGAGCGTAAGAAATATTCTTAGATAAATGGAAACT G	
18789.0_F	CCCCAAAACCCCAAAACCCCATGGCAATATCTTTGCGTGTTTCTGGC	Primer pair for amplifying chromosome, to be added to mini-genome
18789.0_R	CCCCAAAACCCCAAAACCCCATAAGAATAAATTAAAGAAGATTTGAGAAAGATAT GC	

1335.1_F	CCCCAAAACCCCAAAACCCCAAATGCTAAAAATAATGAAAAATCTGAGGG	Primer pair for amplifying chromosome, to be added to mini-genome
1335.1_R	CCCCAAAACCCCAAAACCCCTAATGACAGGTTTAGTAATAATTTAGCTG	
17286.0_F	CCCCAAAACCCCAAAACCCCACGACTTAACATTGCTGTTAAATATTCAGAAAT	Primer pair for amplifying chromosome, to be added to mini-genome
17286.0_R	CCCCAAAACCCCAAAACCCCTAAAATTGGAAAGGGGCAAATTTGCTTATGA	
7278.0_F	CCCCAAAACCCCAAAACCCCATGAGTAATATATACAAATTTTAAATGTATTTTGATTTTA	Primer pair for amplifying chromosome, to be added to mini-genome
7278.0_R	CCCCAAAACCCCAAAACCCCATTGAGTGAGTATTTTTATATTTATTGCGAGTTA	
7752.0_F	CCCCAAAACCCCAAAACCCCACAATAGGCATATTTAATAATTAATTGTTAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
7752.0_R	CCCCAAAACCCCAAAACCCCACTCATTATATAAGGCTGAAAAAATCAGAGG	
244.1_F	CCCCAAAACCCCAAAACCCCTAAATGTAAGAGTAAACTATCATATGAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
244.1_R	CCCCAAAACCCCAAAACCCCATAATGCGAAATATTCATCAGAGTAAATAATG	
20383.0_F	CCCCAAAACCCCAAAACCCCATACGTCATGATTATAAGATTATTATAGAATGCTTAC	Primer pair for amplifying chromosome, to be added to mini-genome
20383.0_R	CCCCAAAACCCCAAAACCCCTCTTGTAAAATAATAAGTTTAAGAAATTGAATTTAG	
331.1_F	CCCCAAAACCCCAAAACCCCATAATATCAAATTAATGAATATTTATCAATTTTATTAAT	Primer pair for amplifying chromosome, to be added to mini-genome
331.1_R	CCCCAAAACCCCAAAACCCCCCCTAATGTCCATAATTTATGTATCAAATAAGG	
22208.0_F	CCCCAAAACCCCAAAACCCCATGATGGTGGAGGAGTGAAGATAAATTAGAATG	Primer pair for amplifying chromosome, to be added to mini-genome
22208.0_R	CCCCAAAACCCCAAAACCCCAAAGTGCAATAAAAAGAGTGAAAATAAATTTTTG	
21396.0_F	CCCCAAAACCCCAAAACCCCATATACCAATGTTAAAAATGAATATTGATATAGAATAG	Primer pair for amplifying chromosome, to be added to mini-genome
21396.0_R	CCCCAAAACCCCAAAACCCCATAATACAAAGTAAAATTGTTTTTTATAGTTCATAA	
11890.0_F	CCCCAAAACCCCAAAACCCCACATAGTGAATGAATTAATGAATAAGTTTGAG	Primer pair for amplifying chromosome, to be added to mini-genome
11890.0_R	CCCCAAAACCCCAAAACCCCGTGATAATAAATTCCTGAGTATATAGTTTAAGAAG	
13521.0_F	CCCCAAAACCCCAAAACCCCGTGATTGCATTTTTTTGCGAAATATTTGC	Primer pair for amplifying chromosome, to be added to mini-genome
13521.0_R	CCCCAAAACCCCAAAACCCCTGAGTTCTCATGTAATAAAAGAATCCATG	
3511.0_F	CCCCAAAACCCCAAAACCCCATGATGCTACAAAAACGCTATATAATCTATAAC	Primer pair for amplifying chromosome, to be added to mini-genome
3511.0_R	CCCCAAAACCCCAAAACCCCTTGAACTTTCAATAGATGTTTGATTAAATTC	
22209.0_F	CCCCAAAACCCCAAAACCCCAAAGATATGTGGCTGGATTTTAAAATATGGTTG	Primer pair for amplifying chromosome, to be added to mini-genome
22209.0_R	CCCCAAAACCCCAAAACCCCAAGACTAATGAATTTGAGAATTATAAAATAATGAATC	
18924.0_F	CCCCAAAACCCCAAAACCCCATCAACTTTAATTCATTGTAGGAATTAAAGATGTAATAC	Primer pair for amplifying chromosome, to be added to mini-genome
18924.0_R	CCCCAAAACCCCAAAACCCCGTGAGAACAAATAATAATAAAAATAAAGGAATTAA	
14977.0_F	CCCCAAAACCCCAAAACCCCAATTCTTTATCTGAATTAGATAAGAATTCATAAGC	Primer pair for amplifying chromosome, to be added to mini-genome
14977.0_R	CCCCAAAACCCCAAAACCCCGTGAGTATGCAATAGATTGTTAATTAAATTTG	
18694.0_F	CCCCAAAACCCCAAAACCCCAAGTTGCTAAAAATAGTTGATAGCAACAAGTTAT	Primer pair for amplifying chromosome, to be added to mini-genome
18694.0_R	CCCCAAAACCCCAAAACCCCGTGGATGTGTTTTTTTTCCAAATTAATGAACAAAAATTAAA	

13237.0_F	CCCCAAAACCCCAAAACCCCAACATTCTAAATTTCTTCTTTATAAGATTATTG	Primer pair for amplifying chromosome, to be added to mini-genome
13237.0_R	CCCCAAAACCCCAAAACCCCATCTAAACTAATCTGAAACCAAAGATAGTATG	
21338.0_F	CCCCAAAACCCCAAAACCCCGTTATCCATATATACGTAAGCATTTTGCGATTG	Primer pair for amplifying chromosome, to be added to mini-genome
21338.0_R	CCCCAAAACCCCAAAACCCCGAAACCTATGCATTATTTTTAAAGAAATATTAAATTAA	
215.1_F	CCCCAAAACCCCAAAACCCCTCGTACATTAATAGTTGAAATTGCTTTTATTAAATTG	Primer pair for amplifying chromosome, to be added to mini-genome
215.1_R	CCCCAAAACCCCAAAACCCCGTAGTCTAAAATAAATTTTATTTTGGGTTTTAA	
13236.0_F	CCCCAAAACCCCAAAACCCCGTTAAATGATAATCATAGCAAAATTGCGGTAT	Primer pair for amplifying chromosome, to be added to mini-genome
13236.0_R	CCCCAAAACCCCAAAACCCCAAGGATAAATATTGAAAGTAAATGTTCTAATTAATTTGC	
16827.0_F	CCCCAAAACCCCAAAACCCCAGAAATGAAAAGAATGATTTTTGAGGGGATTC	Primer pair for amplifying chromosome, to be added to mini-genome
16827.0_R	CCCCAAAACCCCAAAACCCCTAAAGGCAAAAGTCGATTTAAATGCTCAGTTTC	
15136.0_F	CCCCAAAACCCCAAAACCCCTTAAGGCTAAAATACTTGTTTTACTAGAGAAC	Primer pair for amplifying chromosome, to be added to mini-genome
15136.0_R	CCCCAAAACCCCAAAACCCCATAAATCAAATTAAATTGCATAACATGAAC	
115.1_F	CCCCAAAACCCCAAAACCCCAGAGGATGTAAATTACAATAAATCGTAAAAAC	Primer pair for amplifying chromosome, to be added to mini-genome
115.1_R	CCCCAAAACCCCAAAACCCCTTCTAAAAAATATAAAGATAAATTGACGTC	
21295.0_F	CCCCAAAACCCCAAAACCCCATCCAGTTGAAATCTAAAACAATTTTGTATATTTAAAG	Primer pair for amplifying chromosome, to be added to mini-genome
21295.0_R	CCCCAAAACCCCAAAACCCCTTAAGAGATTGCATTATAAATAAGATAGGATTC	
16269.0_F	CCCCAAAACCCCAAAACCCCATTGATTGATAAACTTGGAAGTTAAGAAAGATTTG	Primer pair for amplifying chromosome, to be added to mini-genome
16269.0_R	CCCCAAAACCCCAAAACCCCATGAATAACAGATGGAATGCTTCAAGATATG	
644.1_F	CCCCAAAACCCCAAAACCCCAAATGTTAGTATTTGAATTAAAGAGAGGTAAAAC	Primer pair for amplifying chromosome, to be added to mini-genome
644.1_R	CCCCAAAACCCCAAAACCCCTTATGAAAATGAAATGGTTTTGATTGGCTAATAA	
5586.0_F	CCCCAAAACCCCAAAACCCCATGAGTAAAATTTAGCTTAAGTAATGTAAGAATC	Primer pair for amplifying chromosome, to be added to mini-genome
5586.0_R	CCCCAAAACCCCAAAACCCCATATATCAAAATATCAACATTTTTTTGTGTGATTGTTAC	
13085.0_F	CCCCAAAACCCCAAAACCCCTTGATGAAATTTGAAAATGAATAGAGAGTAC	Primer pair for amplifying chromosome, to be added to mini-genome
13085.0_R	CCCCAAAACCCCAAAACCCCGTAATGCTACATTTGCAAAAAAGTACAAACAG	
13838.0_F	CCCCAAAACCCCAAAACCCCGTAAGGCCAGAATCAATGAATAAAAAGGTC	Primer pair for amplifying chromosome, to be added to mini-genome
13838.0_R	CCCCAAAACCCCAAAACCCCGAAAAGGGAGATTTACAAAAATTTGTAGATGTTATATTG	
1415.1_F	CCCCAAAACCCCAAAACCCCATTGATCATTAATAAAGAAGAATTGCTAATAT	Primer pair for amplifying chromosome, to be added to mini-genome
1415.1_R	CCCCAAAACCCCAAAACCCCAATGCGATGAAATGTTTTTTATTATGAAAAG	
19468.0_F	CCCCAAAACCCCAAAACCCCAAGGAAGTTCAATGCTATTTAGCAAATTAGG	Primer pair for amplifying chromosome, to be added to mini-genome
19468.0_R	CCCCAAAACCCCAAAACCCCTTGATTCAAAATATGCACAAGATTAAAAATTCAC	
20407.0_F	CCCCAAAACCCCAAAACCCCATAAGAAAGATAAGTTGCAATTAAATAATAAGG	Primer pair for amplifying chromosome, to be added to mini-genome
20407.0_R	CCCCAAAACCCCAAAACCCCATGAAGACAAGTCTGATGAAAATAGAATGG	

19922.0_F	CCCCAAAACCCCAAAACCCCATAGTCTTAAAATTTTATACTATCATGAAATAATATTAAG	Primer pair for amplifying chromosome, to be added to mini-genome
19922.0_R	CCCCAAAACCCCAAAACCCCGTAAGTCTAAAGTTTAACAGTTTTTAGTAAATATC	
20459.0_F	CCCCAAAACCCCAAAACCCCTTATGCTAGTTGAGTGATTGAAAATATATTTGTGC	Primer pair for amplifying chromosome, to be added to mini-genome
20459.0_R	CCCCAAAACCCCAAAACCCCTTGACGTAGAATAATGGGCTTATAGAAG	
20493.0_F	CCCCAAAACCCCAAAACCCCTTAATCAACTCACTTTACCCACTAATCAAACAC	Primer pair for amplifying chromosome, to be added to mini-genome
20493.0_R	CCCCAAAACCCCAAAACCCCATATTTAAGATATACAGAAATATAGAGAATACAAC	
9925.0_F	CCCCAAAACCCCAAAACCCCATTGGATCAATTTTGAAGAGAATTCATGGAAAAT	Primer pair for amplifying chromosome, to be added to mini-genome
9925.0_R	CCCCAAAACCCCAAAACCCCATCAGAAAAAATATTTGAAAATTCGATAAAGC	
22456.0_F	CCCCAAAACCCCAAAACCCCATTTCACTTTATTTATATATAGATTTGAAATTAAAGTT	Primer pair for amplifying chromosome, to be added to mini-genome
22456.0_R	CCCCAAAACCCCAAAACCCCAGTTGACATGTTATTTCCAAATTTTCATGGATA	
17712.0_F	CCCCAAAACCCCAAAACCCCATGATAACAGGAATATTTTATAAAATAGTTAAG	Primer pair for amplifying chromosome, to be added to mini-genome
17712.0_R	CCCCAAAACCCCAAAACCCCTCACTCTATGCAATAAATTTGTTGATATATT	
11116.0_F	CCCCAAAACCCCAAAACCCCTTAAAAAAAGAATAGTTGGAATAAAAATGAATTT	Primer pair for amplifying chromosome, to be added to mini-genome
11116.0_R	CCCCAAAACCCCAAAACCCCAATAGATAAAGATGCCTTTTTTAATAAGTATTTAAC	
19275.0_F	CCCCAAAACCCCAAAACCCCGAGAGGATAAATTTATATGAAAATAAAAATAAAGC	Primer pair for amplifying chromosome, to be added to mini-genome
19275.0_R	CCCCAAAACCCCAAAACCCCATAAATAAGAAATTTTAAGAATAACGGGCAAATTAG	
21217.0_F	CCCCAAAACCCCAAAACCCCTTGAATTTTAAATAAAGTTCTTTGTATGATTTAAATG	Primer pair for amplifying chromosome, to be added to mini-genome
21217.0_R	CCCCAAAACCCCAAAACCCCATAGATTACTTTTCAAAGAATTTCTTGACATTC	
10537.0_F	CCCCAAAACCCCAAAACCCCAAAGCAAAGAAATCTGATGTTTTATTAGAAAAAGTG	Primer pair for amplifying chromosome, to be added to mini-genome
10537.0_R	CCCCAAAACCCCAAAACCCCATGAGATGATAATATTGCCTTTTTGCATATAAT	
22670.0_F	CCCCAAAACCCCAAAACCCCATCCTTATACAAATTCAGAAAACTTAGCAAAT	Primer pair for amplifying chromosome, to be added to mini-genome
22670.0_R	CCCCAAAACCCCAAAACCCCGTGGAGAATTTTCTAAAGAATTTTCGGAAATTTG	

1781.0_F	CCCCAAAACCCCAAAACCCCACTAGTCTTAAATATGAGAAAGATGATTTGAATAAG	PCR primers for amplifying synthetic chromosome 1 and 8 in Fig 5B
1781.0_R	CCCCAAAACCCCAAAACCCCACAAGATAACATTGCTAATTTTCAATAAATTAAAT	

1781.0_Purple_F	GTCAGTGGTCTCAGTATGAAATTTACCTTGGATCCTCAGTGTTTCCCTTTGTG	PCR primers for amplifying purple DNA building block in synthetic chromosomes 2-4 in Fig 5B
1781.0_Purple_R	AACGCTCGGTCTCGCAGAAATAAATAATAGGCAGCATAATAATTCAGG	
1781.0_red_F	GTCAGTGGTCTCTCCAGTGGATTCACATCACTACCATTG	PCR primers for amplifying red DNA building block in synthetic chromosomes 2-4 in Fig 5B
1781.0_red_R	CCCCAAAACCCCAAAACCCCACAAGATAACATTGCTAATTTTCAATAAATTAAAT	
1781.0_turquoise_F	CCCCAAAACCCCAAAACCCCACTAGTCTTAAATATGAGAAAGATGATTTGAATAAG	PCR primers for amplifying turquoise DNA building block in synthetic chromosomes 2-4 in Fig 5B
1781.0_turquoise_R	ACGCTCGGTCTCGATACAATCATTAGCATATACATGCAGTCTGCTTAAATTAAG	

DarkBlue_6mA_top	TCTGTAATTAATGGATCACTATAATCATCTGGATGACTATTGGTATTTGATGACG GTGCACATTTGACTTCTT	Oligos for annealing to make blue DNA building block in synthetic chromosomes 2-4 in Fig 5B. Bold red nucleotides represent 6mA.
DarkBlue_6mA_bottom	ATTAATTACCTAGTGATATTAGTAGACCTACTGATAACCATAAACTACTGCCACG TGTAAACTGAAGAAGGTC	

1781.0_red2_F	AGCCTAGGTCTCGTTCTTTTTGAGGGCGAGGTTATTGTTAAT	PCR primers for amplifying red DNA building block in synthetic chromosome 5 in Fig 5B
1781.0_red2_R	CCCCAAAACCCCAAAACCCCACAAGATAACAT	
1781.0_orange_F	TAGTCAGGTCTCTAGAATAGGCTCACTCTAAATTCGAGTGCAAT	PCR primers for amplifying orange DNA building block in synthetic chromosome 5 in Fig 5B
1781.0_orange_R	TCTACTGGTCTCAGTATGAAATTTACCTTGGATCCTCAGTGT	
1781.0_emerald_F	ATCGTAGGTCTCAATACAATCATTAGCATATACATGCAGT	PCR primers for amplifying emerald DNA building block in synthetic chromosome 5 in Fig 5B
1781.0_emerald_R	CCCCAAAACCCCAAAACCCCACTAGTCTTAAAT	

17535.0_F	CCCCAAAACCCCAAAACCCCATCTCAATTTATAAAATCAGAATAAGAGATTGTC	PCR primers for amplifying "buffer" chromosome (Contig17535.0) for use in chromatin assembly
17535.0_R	CCCCAAAACCCCAAAACCCCAGAATAAAACAACTGAAGTAAATATGAGTTAC	

12701assay_F	AAGAAGAACTAGCCAGCTCTCACTCAGTTC	PCR primers for assaying the presence of ectopic DNA insertion in *mta1* mutants
12701assay_R	TGTCTATCTCATCAGGCTCATCAGCATAGG	

12701_firstround_T7_F	AAGAAGAACTAGCCAGCTCTCACTCAGTTC	PCR primers to generate DNA template for ssRNA *in vitro* transcription. This ssRNA is injected into *Oxytricha* cells to induce ectopic DNA retention in MTA1 gene. PCR product is amplified from *Oxytricha* gDNA of cell strain JRB310. The resulting PCR product is subjected to a second round of PCR amplification using primers "12701_secondround_T7_F" and "12701_secondround_T7_R". The final, second round PCR product is then used for ssRNA *in vitro* transcription.
12701_firstround_T7_R	CCTCTCTGCCCACTAAATTATTCTGACAGC	
12701_secondround_T7_F	CTACTTGATATAATACGACTCACTATAGGGAATTCCTAAGGGGAGTGAAGCCAA CAACAG	PCR primers for second round amplification of DNA template, to be used for ssRNA *in vitro* transcription. Forward primer contains T7 promoter sequence, which is required for subsequent *in vitro* transcription.
12701_secondround_T7_R	TGTCTATCTCATCAGGCTCATCAGCATAGG	

metGATC_F2	GTGCTATGCATTTTAAATTTATTCGCATTGAAGA	PCR primers for amplification of DNA substrate for use in 6mA methyltransferase assay involving Tetrahymena nuclear extracts, and for DpnI digestion assay on nuclear extract-methylated products
metGATC_R2	ATTCAGAATTTTAGTGTGTGGAGTATGATAGTA	

noGATC2_F	GGTCTATATTATTTTAGTATTCTTTCTATAAATG	PCR primers for amplifying 350bp dsDNA substrate for methyltransferase assays involving recombinant proteins (in Fig 2E, 2F, and S4H)
noGATC2_R	GTTACAAGAATATAAGAAAAGAAAGGGTGAATAGG	

T7noGATC2_F2	TAATACGACTCACTATAGGG GGTCTATATTATTTTAGTATTCTTTC	PCR primers for amplifying DNA ~350bp dsDNA template with T7 overhangs at one end, for

noGATC2_R	GTTACAAGAATATAAGAAAAGAAAGGGTGAATAGG	subsequent ssRNA production by in vitro transcription
T7noGATC2_F2	TAATACGACTCACTATAGGG GGTCTATATTATTTTAGTATTCTTTC	PCR primers for amplifying DNA ~350bp dsDNA template with T7 overhangs at the 5' and 3' ends, for subsequent dsRNA production by in vitro transcription
T7noGATC2_R2	TAATACGACTCACTATAGGG GTTACAAGAATATAAGAAAAG	
noGATC_F3	AACTTCTGTCATTACATTAAGCTTTAA	
noGATC_R3	TTAAAGCTTAATGTAATGACAGAAGTT	
noGATC_F12	AACTTCTGTCATTAACTTAAGCTTTAA	
noGATC_R12	TTAAAGCTTAAGTTAATGACAGAAGTT	
noGATC_F13	AACTTCTGTACTTACATTAAGCTTTAA	
noGATC_R13	TTAAAGCTTAATGTAAGTACAGAAGTT	
noGATC_F14	AACTTCTGTAGTTAACTTAAGCTTTAA	
noGATC_R14	TTAAAGCTTAAGTTAAGTACAGAAGTT	DNA oligonucleotides for use in DNA methyltransferase assays in Figure 2G, S4I, S4J, S4K, S4L
noGATC_F1	AACTTCTGTCATTACATTAAGCTTTAAAAAATTCAATTCCTTTTATT	
noGATC_R1	AATAAAAGGAATTGAATTTTTTAAAGCTTAATGTAATGACAGAAGTT	
noGATC_F2	TGTCATTACATTAAGCTTTAAAAAATTCAATTCCT	
noGATC_R2	AGGAATTGAATTTTTTAAAGCTTAATGTAATGACA	
noGATC_F3	AACTTCTGTCATTACATTAAGCTTTAA	
noGATC_R3	TTAAAGCTTAATGTAATGACAGAAGTT	
noGATC_F8	TATTAGAATTATGTTCTTCATGAAATT	
noGATC_R8	AATTTCATGAAGAACATAATTCTAATA	

Table 3. Recombinant protein sequences.

>MTA1 (manually curated from Tetrahymena DB gene ID: TTHERM_00704040)

MSKAVNKKGLRPRKSDSILDHIKNKLDQEFLEDNENGEQSDEDYDQKSLNKAKKPYKKRQTQNGSELVISQQK
TKAKASANNKKSAKNSQKLDEEEKIVEEEDLSPQKNGAVSEDDQQQEASTQEDDYLDRLPKSKKGLQGLLQDI
EKRILHYKQLFFKEQNEIANGKRSMVPDNSIPICSDVTKLNFQALIDAQMRHAGKMFDVIMMDPPWQLSSSQPS
RGVAIAYDSLSDEKIQNMPIQSLQQDGFIFVWAINAKYRVTIKMIENWGYKLVDEITWVKKTVNGKIAKGHGFYL
QHAKESCLIGVKGDVDNGRFKKNIASDVIFSERRGQSQKPEEIYQYINQLCPNGNYLEIFARRNNLHDNWVSIG
NEL

>MTA9 (manually curated from Tetrahymena DB gene ID: TTHERM_00301770)

MAPKKQEQEPIRLSTRTASKKVDYLQLSNGKLEDFFDDLEEDNKPARNRSRSKKRGRKPLKKADSRSKTPSRV
SNARGRSKSLGPRKTYPRKKNLSPDNQLSLLLKWRNDKIPLKSASETDNKCKVVNVKNIFKSDLSKYGANLQA
LFINALWKVKSRKEKEGLNINDLSNLKIPLSLMKNGILFIWSEKEILGQIVEIMEQKGFTYIENFSIMFLGLNKCLQSI
NHKDEDSQNSTASTNNTNNEAITSDLTLKDTSKFSDQIQDNHSEDSDQARKQQTPDDITQKKNKLLKKSSVPSI
QKLFEEDPVQTPSVNKPIEKSIEQVTQEKKFVMNNLDILKSTDINNLFLRNNYPYFKKTRHTLLMFRRIGDKNQKL
ELRHQRTSDVVFEVTDEQDPSKVDTMMKEYVYQMIETLLPKAQFIPGVDKHLKMMELFASTDNYRPGWISVIEK

>p1 (manually curated from Tetrahymena DB gene ID: TTHERM_00161750)

MSLKKGKFQHNQSKSLWNYTLSPGWREEEVKILKSALQLFGIGKWKKIMESGCLPGKSIGQIYMQTQRLLGQQ
SLGDFMGLQIDLEAVFNQNMKKQDVLRKNNCIINTGDNPTKEERKRRIEQNRKIYGLSAKQIAEIKLPKVKKHAP
QYMTLEDIENEKFTNLEILTHLYNLKAEIVRRLAEQGETIAQPSIIKSLNNLNHNLEQNQNSNSSTETKVTLEQSG
KKKYKVLAIEETELQNGPIATNSQKKSINGKRKNNRKINSDSEGNEEDISLEDIDSQESEINSEEIVEDDEEDEQIE
EPSKIKKRKKNPEQESEEDDIEEDQEEDELVVNEEEIFEDDDDDEDNQDSSEDDDDDED

>p2 (manually curated from Tetrahymena DB gene ID: TTHERM_00439330)

MKKNGKSQNQPLDFTQYAKNMRKDLSNQDICLEDGALNHSYFLTKKGQYWTPLNQKALQRGIELFGVGNWKE
INYDEFSGKANIVELELRTCMILGINDITEYYGKKISEEEQEEIKKSNIAKGKKENKLKDNIYQKLQQMQ

Sequences were manually curated by mapping RNaseq reads to reference gene annotations and verifying the accuracy of predicted exon boundaries.

Example 2

Epigenomic Profiles of Chromatin and Transcription in *Oxytricha*

[0201] We generated genome-wide in vivo maps of nucleosome positioning, transcription, and 6mA in the macronuclei of asexually growing (vegetative) *Oxytricha trifallax* cells using Mnase sequencing (MNase-seq), poly(Ar RNA sequencing (RNA-seq), transcriptional start site sequencing (TSS-seq), and single-molecule real-time sequencing (SMRT-seq) (Figs. 1A – 1E). The smallest *Oxytricha* chromosome is only 430 bp in length, with a single well-positioned nucleosome. Strikingly, 6mA is enriched in three consecutive nucleosome-depleted regions directly downstream of transcription start sites (TSSs; Fig. 1A). Each region contains varying levels of 6mA (Fig. 1B), with the +1/+2 nucleosome linker being most

densely methylated (Table 4). In general, highly transcribed chromosomes tend to bear more 6mA, suggesting a positive role of this DNA modification in gene regulation (Fig. 1C). The majority of methylation marks are located within an ApT motif (Figs. 1D and 1E). 6mA occurs on sense and antisense strands with approximately equal frequency, indicating that the underlying methylation machinery does not function strand-specifically. Quantitative LC-MS/MS analysis confirmed the presence of 6mA in *Oxytricha* (Figs. 8A and 8B; see Example 1).

Table 4. Descriptive statistics of 6mA distribution in the genome.

	Number of 6mA sites									
	Oxytricha					*Tetrahymena*				
	Minimum	Maximum	Median	Mean	Standard Deviation	Minimum	Maximum	Median	Mean	Standard Deviation
Methyl Cluster 1	0	14	2	2.03	2.27	0	27	10	9.88	6.10
Methyl Cluster 2	0	24	6	5.99	4.24	0	26	9	8.78	5.78
Methyl Cluster 3	0	16	2	2.49	2.91	0	25	5	5.75	5.53

Properties of 6mA distribution in nucleosome linkers. In *Oxytricha,* methyl cluster 1 = between 5' chromosome end and +1 nucleosome; methyl cluster 2 = between +1 and +2 nucleosome; methyl cluster 3 = between +2 and +3 nucleosome. In *Tetrahymena,* methyl cluster 1 = between +1 and +2 nucleosome; methyl cluster 2 = between +2 and +3 nucleosome; methyl cluster 3 = between +3 and +4 nucleosome. Consensus +1/+2/+3/+4 nucleosome positions: 193, 402, 618, 837 bp downstream of *Oxytricha* 5' chromosome ends; 112, 304, 497, 698 bp downstream of *Tetrahymena* TSSs.

Example 3

Purification and Reconstitution of the Ciliate 6mA Methyltransferase, MTA1c

[0202] To uncover the functions of 6mA *in vivo*, we set out to identify and disrupt putative 6mA methytransferases (MTases). The *Oxytricha* genome encodes a large number of candidate methyltransferases (Table 5), rendering it impractical to test gene function, one at a time or in combination. To identify the ciliate 6mA MTase, we undertook a biochemical approach by fractionating nuclear extracts and

identifying candidate proteins that co-purified with DNA methylase activity. The organism of choice for this experiment was *Tetrahymena thermophila*, a ciliate that divides significantly faster than *Oxytricha* (~2 h versus 18 h; Cassidy-Hanley, 2012; Laughlin et al., 1983). This faster growth time rendered it feasible to culture large amounts of *Tetrahymena* cells for nuclear extract preparation. *Tetrahymena* and *Oxytricha* exhibit similar genomic localization and 6mA abundance (Figs. 8A – 8B and 9A-9F). We thus reasoned that the enzymatic machinery responsible for 6mA deposition is conserved between *Tetrahymena* and *Oxytricha*, and that Tetrahymena could serve as a tractable biochemical system for identifying the ciliate 6mA MTase.

[0203] We prepared nuclear extracts from log-phase *Tetrahymena* cells, since 6mA could be readily detected at this developmental stage through quantitative MS and PacBio sequencing (Figs. 8A-8B and 9A-9F). Nuclear extracts were incubated with radiolabeled S-adenosyl-L-methionine (SAM) and PCR-amplified DNA substrate to assay for DNA methylase activity. Passage of the nuclear extract through an anion exchange column resulted in the elution of two distinct peaks of DNA methylase activity, both of which were heat sensitive (Figs. 2C, 10A, and 10B). Western blot analysis confirmed that both peaks of activity mediate methylation on 6mA (Fig. 10C). The resulting fractions were further purified and subjected to MS. Only four proteins-termed MTA1, MTA9, p1, and p2-were detected at higher abundance in fractions with high DNA methylase activity (Figs. 2C and 2D). p1 and p2 contain homeobox-like domains, suggesting a DNA binding function for an undetermined process (Fig. 10D). On the other hand, MTA1 and MTA9 are both MT-A70 proteins. Such domains are widely known to mediate m6A RNA methylation in eukaryotes (Liu et al., 2014). MTA1 and MTA9 received the large majority of peptide matches, relative to all other MT-A70 genes encoded by the Tetrahymena genome (Fig. 2D; Table 6). Curiously, although poly(A)-selected RNA transcripts were present from all MT-A70 genes (Fig. 2D), almost all peptides in fractions with high DNA methylase activity corresponded to MTA1 and MTA9. The poly(A)$^+$ RNA expression profiles of MTA1, MTA9, p1, and p2 are remarkably similar (Fig. 9K), peaking early in the sexual cycle. This coincides with a sharp increase in nuclear 6mA, as evidenced from immunostaining (Wang et al., 2017). Accumulation of MTA1, MTA9, p1, and p2 therefore correlates with the presence of 6mA *in vivo*.

[0204] We next investigated the phylogenetic relationship of MTA1 and MTA9 to other eukaryotic MT-A70 domain-containing proteins. Two widely studied mammalian MT-A70 proteins - METTL3 and METTL14 (Ime4 and Kar4 in yeast)- form a heterodimeric complex that is responsible for m6A methylation on mRNA. METTL3 is the catalytically active subunit, while METTL14 functions as an RNA-binding scaffold protein (Sledi arid Jinek, 2016; Wang et al., 2016a, 2016b). MTA1 and MTA9 derive from distinct monophyletic clades, outside of those that contain mammalian METTL3, METTL14, and *C. elegans* DAMT-1 (METTL4) (Fig. 2A). Thus, MTA1 and MTA9 are divergent MT-A70 family members that are phylogenetically distinct from all previously studied RNA and DNA N6-methyladenine MTases. We then asked whether MTA1 and MTA9 are also present in other eukaryotes with a similar occurrence of 6mA in ApT motifs as *Tetrahymena*. We queried the genomes of *Oxytricha*, green algae, and eight basal yeast species, all of which exhibit this distinct methylation pattern (as evidenced from Figs. 1A – 1E ; Figs. 9A-9E; Fu et al., 2015; Mondo et al., 2017). For all of these taxa, we can identify MT-A70 homologs that are monophyletic with MTA1 and MTA9 (Fig. 2B). On the other hand, MT-A70 homologs from multicellular eukaryotes, including *Arabidopsis*, *C. elegans*, *Drosophila*, and mammals, grouped exclusively with METTL3, METTL14, and METTL4 lineages, but not MTA1 or MTA9. None of these latter genomes exhibit a consensus ApT dinucleotide methylation motif for 6mA (Greer et al., 2015; Koziol et al., 2016; Liang et al., 2018; Liu et al., 2016; Wu et al., 2016; Xiao et al., 2018; Zhang et al., 2015). We note that the absence of an ApT dinucleotide motif is based on data from a limited number of cell types, developmental stages, and culture conditions tested in these studies. Nonetheless, within the scope of currently available data, the presence of MTA1 and MTA9 correlates with the distinctive genomic localization of 6mA within ApT motifs.

[0205] We then sought to determine whether MTA1 and/or MTA9 are bona fide 6mA methyltransferases. MTA1, but not MTA9, contains a catalytic DPPW motif (Fig. 10E) — a hallmark of N6-adenosine methyltransferases (Iyer et al., 2016). Surprisingly, recombinant full-length *Tetrahymena* MTA1 and MTA9 (Fig. 10G) showed no detectable DNA methyltransferase activity, individually or together (Fig. 2E). Examination of the MTA1 and MTA9 sequences revealed that neither protein possesses a predicted nucleic acid binding domain (Fig. 10D). In contrast, METTL3,

which catalyzes m6A methylation on RNA, contains two tandem CCCH-type zinc finger motifs, necessary for RNA binding (Huang et al., 2019; Wang et al., 2016a). Additional co-factors may thus be necessary for MTA1/7 to engage DNA substrates. Indeed, the p1 and p2 proteins that co-elute with MTA1/7 in nuclear extracts possess homeobox-like domains predicted to bind DNA. We then tested whether these accessory factors, in addition to MTA1/7, are necessary for 6mA methylation. Strikingly, mixing recombinant, full-length p1, p2, MTA1, and MTA9 resulted in robust 6mA methylation in vitro (Figs. 2E and 2F). This activity was abolished when each protein was omitted, indicated that all four are necessary for 6mA methylation. Furthermore, MTA1 harboring a D209A mutation in the catalytic DPPW motif showed no activity, even in the presence of MTA9, p1, and p2 (Fig. 2E). We also created double mutations in MTA1 (N370A, E371A), which lie in the conserved region that interacts with the 2' and 3'-hydroxyl groups of the ribose moiety in the SAM cofactor (Fig. 2E). This mutant protein also exhibited no 6mA methylase activity. Taken together, we find that four proteins— MTA1, MTA9, p1, and p2 — are necessary for 6mA methylation in vitro, with MTA1 the likely catalytic subunit. Henceforth, we refer to these four proteins as the putative MTA1 complex (MTA1c).

[0206] Purification of the MTA1c proteins from an *E. coli* overexpression system raises the possibility of methyltransferase activity arising from contaminating *Dam* methylase; however, we exclude this possibility for three reasons. (1) The DNA substrate used in this assay does not contain 5'-NATC-3' sites, which are recognized and methylated by *Dam* methylase (Horton et al., 2006). (2) Methyltransferase activity was only observed when all four recombinant proteins were incubated with DNA. If contaminating Dam methylase were present in one or more of these protein preparations, then background activity should be observed when subsets of these proteins are used in the assay. 3) Mutation of MTA1 catalytic residues leads to loss of methylation, which is also inconsistent with contaminating methyltransferase activity.

Table 5. Candidate genes in ciliates.

MT-A70 genes in *Oxytricha trifallax*

UniProt ID	Gene name in this study	OxyDB gene name
J9IF92_9SPIT	MTA1	Contig12701.0.0.g16
J9IGS7_9SPIT	TAMT-1	Contig17486.0.g100
J9J9V7_9SPIT	MTA1-B	Contig16314.0.g25
J9HW68_9SPIT	MTA9	Contig1237.1.g126
J9IMU5_9SPIT	MTA9-B	Contig17419.0.g36

MT-A70 genes in *Tetrahymena thermophila*

UniProt ID	Gene name in this study	*Tetrahymena* Genome Database gene name
Q22GC0_TETTS	MTA1	TTHERM_00704040
Q23TW8_TETTS	MTA2	TTHERM_00962190
I7LVP8_TETTS	MTA3 / TAMT-1-B	TTHERM_00136470
I7MGX6_TETTS	MTA4	TTHERM_00558100
Q23RE0_TETTS	MTA5 / TAMT-1	TTHERM_00388490
I7MIF9_TETTS	MTA9	TTHERM_00301770
Q22XT1_TETTS	MTA9-B	TTHERM_01005150

METTL16 homologs in *Oxytricha trifallax*

UniProt ID	OxyDB gene name
J9F3J7_9SPIT	Contig11945.0.g48
J9J5P9_9SPIT	Contig7462.0.g41
J9III0_9SPIT	Contig4244.0.g39

N6AMT1 homologs in *Oxytricha trifallax*

UniProt ID	OyDB gene name
J9IFV1_9SPIT	Contig7751.0.g12

Accessory factor genes in *Tetrahymena thermophila*

UniProt ID	Gene name in this study	*Tetrahymena* Genome Database gene name
Q22VV9_TETTS	p1	TTHERM_00161750
I7M8B9_TETTS	p2	TTHERM_00439330

ISWI homologs in *Oxytricha trifallax* and *Tetrahymena thermophila*

UniProt ID	OxyDB gene name	*Tetrahymena* Genome Database gene name
I7M280_TETTS		TTHERM_00137610
J9FBJ2_9SPIT	Contig11737.0.g12	

The Uniprot ID of each gene is listed. The *Oxytricha* macronuclear genome encodes five genes belonging to the MT-A70 family (Iyer et al., 2016; Swart et al., 2013). Such genes commonly function as RNA m6 A MTases in eukaryotes, having evolved from m.MunI-like MTases in bacterial restriction-modification systems (Iyer et al., 2016). An MT-A70 gene belonging to the METTL4 subclade, DAMT1, is a putative 6mA methyltransferase in *C. elegans* (Greer et al., 2015). However, none of the *Oxytricha* MT-A70 genes in this Table cluster together with METTL4 on a phylogenetic tree (Figures 2A and S2G). The *Oxytricha* genome also contains homologs of a structurally distinct RNA m6 A MTase, METTL16, which was reported

to methylate U6 snRNA (Table 5) (Pendleton et al., 2017; Warda et al., 2017). Another candidate, N6AMT1 – which does not contain an MT-A70 domain – was recently found to mediate DNA 6mA methylation in human cells (Xiao et al., 2018). An N6AMT1 homolog is also present in the *Oxytricha* genome. Accessory factors refer to the p1 and p2 proteins, which are necessary for 6mA methylation by MTA1 and MTA9 *in vitro.* The UniProt IDs of putative ISWI homologs in *Oxytricha* and *Tetrahymena* are also listed.

Table 6. Mass spectrometry analysis of MTA1, MTA9, p1, and p2 proteins.

Data from Low Salt Fraction

UniProt ID	Gene name in this study	% of protein covered by peptide data from LC-MS/MS experiment
Q22GC0_TETTS	MTA1	78.8%
I7MIF9_TETTS	MTA9	46.3%
Q22VV9_TETTS	p1	41.9%
I7M8B9_TETTS	p2	81.7%

Data from High Salt Fraction

UniProt ID	Gene name in this study	% of protein covered by peptide data from LC-MS/MS experiment
Q22GC0_TETTS	MTA1	69.9%
I7MIF9_TETTS	MTA9	72.2%
Q22VV9_TETTS	p1	55.3%
I7M8B9_TETTS	p2	93.4%

Percentage of each polypeptide that is covered by peptide data is calculated. “Low Salt Sample” and “High Salt Sample” correspond to partially purified nuclear extracts that elute as two distinct peaks of activity from a Q sepharose anion exchange column (Fig. 2C).

Example 4

MTA1c Preferentially Methylates ApT Dinucleotides in dsDNA

[0207] We next investigated the substrate preferences of MTA1c. First, in vitro transcription was performed to generate doublestranded RNA (dsRNA) and single-stranded RNA (ssRNA) from the input dsDNA substrate. We found that MTA1c methylates dsDNA but not dsRNA or ssRNA of the same sequence, indicating that it is selective for DNA over RNA (Fig. 10H). We then generated a series of dsDNA substrates by annealing oligonucleotide pairs of different length and sequence. All of

these substrates are bona fide Tetrahymena genomic DNA sequences. In each case, MTA1c can methylate the annealed dsDNA but not ssDNA (Figs. 2G and 101).

[0208] Since 6mA methylation mainly lies in ApT dinucleotides in vivo (Figs. 1D and 9D), we asked whether MTA1c preferentially methylates this motif. To test this, we used a 27 bp dsDNA substrate with two ApT dinucleotides in its native sequence (Fig. 2G). We disrupted one or both ApT motifs (Fig. 2G) by mutually swapping the 5' A with a neighboring base 5'-CAT-3' → 5'-ACT-3'. Disrupting both ApT dinucleotides resulted in >10-fold reduced methylation, while disrupting only one motif led to a 2- to 4-fold loss (Figs. 2G and 10K).

[0209] Given that 6mA occurs on both strands of genomic DNA *in vivo* (Figs. 1E and 9E), we asked whether pre-existing methylation of one strand affects MTA1c activity. DNA oligonucleotides were nonspecifically methylated with 6mA using *EcoGII* (Murray et al., 2018), a bacterial 6mA methyltransferase. After rigorous purification, samples were annealed to an unmethylated, complementary strand to yield hemimethylated dsDNA (Fig. 10F). MTA1c activity was 3- to 3.5-fold higher on hemimethylated substrates, relative to unmethylated dsDNA (Fig. 2G). This effect was similar between dsDNA substrates pre-methylated on the sense or antisense strand, consistent with the lack of an overt strand bias in 6mA locations in vivo (Figs. 1E and 9E). Importantly, the increase in MTA1c activity cannot be attributed to contaminating *EcoGII* in hemimethylated substrates, since no activity was observed in the absence of MTA1c (Fig. 10J). Thus, pre-existing 6mA methylation stimulates MTA1c, indicative of a positive feedback loop.

[0210] We then asked whether MTA1c activity is modulated not only by the dinucleotide motif sequence per se, but also by flanking sequences. This may manifest as the wide variation in frequency of DNA 4-mer containing a methylated ApT dinucleotide 5'-NA*TN-3' *in vivo* (Fig. 10M). To test this, we used a dsDNA substrate containing two ApT dinucleotides, both within a 5'-CATT-3'. Swapping of the ApT motif with the adjacent downstream DNA residue produced substrates containing 5'-TATA-3' (Fig. 10L). Substrates with this change at both locations had 4-fold less MTA1c activity, and an intermediate effect when only one dinucleotide was altered (Fig. 10L). These data indicate that 5'-CATT-3' is the preferred methylation substrate, consistent with the higher frequency of methylated 5'-CA*TT-3' versus 5'-TA*TA-3' in both *Tetrahymena* and *Oxytricha* genomic DNA (Fig. 10M). The

difference in frequency of methylated sequences cannot simply be attributed to the higher frequency of the 4nt 5'-CATT-3' motif versus 5'-TATA-3' in the genome, because the opposite trend is observed (Fig. 10N). Thus, MTA1c is sensitive to variation in DNA sequences flanking the ApT dinucleotide motif.

Example 5

MTA1 Is Necessary for 6mA Methylation *In Vivo*

[0211] Having established that MTA1c is a 6mA methyltransferase, we tested the role of MTA1c in mediating 6mA methylation *in vivo* in *Oxytricha*, for which we have ease of generating mutants. The genome-wide localization of 6mA is conserved between *Oxytricha* and *Tetrahymena* (Figs. 1A-1E and 9A-9F), implying similar underlying enzymatic machinery. Indeed, all four component genes-MTA1, MTA9, p1, and p2 - are clearly conserved between both species (Figs. 9G-9J). The DPPW catalytic motif is also completely conserved in *Tetrahymena* and *Oxytricha* MTA1 but not MTA9, suggesting that MTA1 is the likely catalytic subunit of MTA1c in both ciliates (Fig. 10E). To abrogate MTA1c function, we disrupted the *Oxytricha* MTA1 gene by inserting an ectopic DNA sequence 49 bp downstream of the start codon, resulting in a frameshift mutation and loss of the C-terminal MTase domain (Fig. 3A). *Oxytricha* has two MTA1 paralogs, named MTA1 and MTA1-B (Figs. 2A and 9G). We focused on MTA1 because MTA1-B is not expressed in vegetative *Oxytricha* cells (Swart et al., 2013), which we used to profile 6mA locations via SMRT-seq. Dot blot analysis confirmed a significant reduction in bulk 6mA levels in mutant lines (Fig. 3B). We then examined 6mA positions at high resolution using SMRT-seq to understand how the DNA methylation landscape is altered in mta1 mutants. Notably, these mutants exhibit genome-wide loss of 6mA, with complete abolishment of the dimethylated ApT motif, and reduction in frequency of all other methylated dinucleotide motifs (Figs. 3C-3E). These findings are consistent across all biological replicates and are robust to wide variation in SMRT-seq parameters for calling 6mA modifications (Figs. 11B-11D). It cannot be attributed to variation in sequencing coverage between wild-type and mutant lines. The loss of methylated ApT dinucleotides in mta1 mutants is consistent with our in vitro data suggesting that MTA1c primarily methylates ApT sites (Figs. 2G and 10K). The Inter Pulse Duration ratio (degree of polymerase slowing during PacBio sequencing due to presence of a

modified base) and estimated fractional methylation also decreased significantly at called 6mA sites in mtal mutants ($p < 2.2 \times 10^{-16}$, Wilcoxon rank-sum test) (Fig. 11A). MTA1 is therefore necessary for a significant proportion of *in vivo* 6mA methylation events in *Oxytricha*.

[0212] What are the phenotypic consequences of 6mA loss *in vivo*? It has been proposed that DNA methylation -including 6mA and cytosine methylation- is involved in nucleosome organization (Fu et al., 2015; Huff and Zilberman, 2014). We thus asked whether nucleosome organization is altered in mtal mutants. We quantified nucleosome "fuzziness," defined as the SD of MNase-seq read locations surrounding the called nucleosome peak (Lai and Pugh, 2017; Mavrich et al., 2008). A poorly positioned nucleosome consists of a shallow and wide peak of MNase-seq reads, manifested by a high fuzziness score. Nucleosomes were first grouped according to the change in flanking 6mA between wild-type and mtal mutant cells (Figs. 12A-12G). The nucleosomes that experience large changes in flanking 6mA exhibit significantly greater increase in fuzziness, compared to nucleosomes with little change in flanking 6mA (Figs. 12A and 12D). Such nucleosomes also exhibit changes in occupancy that are consistent with an increase in fuzziness (Figs. 12A and 12E). These results are robust to variation in MNase digestion (Figs. 14C and 14D). On the other hand, nucleosome linkers do not change in length or occupancy, even though 6mA is lost from these regions (Figs. 12B, 12C, 12F, and 12G). We conclude that 6mA exerts subtle effects on nucleosome organization *in vivo*.

Example 6

6mA Disfavors Nucleosome Occupancy across the Genome *In Vitro* but Not *In Vivo*

[0213] Multiple factors, including 6mA, DNA sequence, and chromatin remodeling complexes, may collectively contribute to nucleosome organization in vivo. The effect of 6mA could therefore be masked by these elements. We next sought to determine whether 6mA directly impacts nucleosome organization. To this end, we assembled chromatin *in vitro* using *Oxytricha* gDNA, which contains cognate 6mA. To obtain a matched negative control lacking DNA methylation, 98 complete chromosomes were amplified using PCR (Fig. 4A), purified and subsequently mixed together in stoichiometric ratios to obtain a "mini-genome" (Fig. 4B). These

chromosomes collectively reflect overall genome properties, including AT content, chromosome length, and transcriptional activity (Table 7). Native genomic DNA (containing 6mA) and amplified mini-genome DNA (lacking 6mA) were each assembled into chromatin *in vitro* using *Xenopus* or *Oxytricha* histone octamers (Figs. 13A-13F) and analyzed using MNase-seq. We computed nucleosome occupancy from the native genome and mini-genome samples across 199,795 overlapping DNA windows, spanning all base pairs in the 98 chromosomes. This allowed the direct comparison of nucleosome occupancy in each window of identical DNA sequence, with and without 6mA (Figs. 4C and 4D). Windows exhibit lower nucleosome occupancy with increasing 6mA, confirming the quantitative nature of this effect. Furthermore, similar trends were observed for both native *Oxytricha* and recombinant *Xenopus* histones, suggesting that the effects of 6mA on nucleosome organization arise mainly from intrinsic features of the histone octamer rather than from species-specific variants (Figs. 4C and 4D). These results are also robust to the extent of MNase digestion of reconstituted chromatin (Fig. 14A).

[0214] We then directly compared the impact of 6mA on nucleosome occupancy *in vitro* and *in vivo*. Loss of 6mA in vitro is achieved by mini-genome construction, while loss *in vivo* is achieved by the mta1 mutation. For each overlapping DNA window, we calculated the difference in nucleosome occupancy: (1) between native genome and mini-genome DNA *in vitro*, and (2) between wild-type and mta1 mutants *in vivo* (Fig. 4C). Nucleosome occupancy is indeed lower in the presence of 6mA methylation *in vitro* (Figs. 4C and 4D). In contrast, no change in nucleosome occupancy is observed *in vivo* (Figs. 4C and 4E). This result is consistent with our earlier analysis of linker occupancy in mta1 mutants (Figs. 12C and 12G). We note that highly methylated DNA windows show greater change in 6mA relative to mta1 mutants (Fig. 3D). Yet, these windows do not change in nucleosome occupancy *in vivo*. We conclude that 6mA methylation locally disfavors nucleosome occupancy *in vitro*, but that this intrinsic effect can be overcome by endogenous chromatin factors *in vivo*.

Table 7. Descriptive statistics of reference genomes.

	Native genomic DNA	Mini-genome DNA
Chromosome length (bp)	**2449 +/- 742** Min = 1155 Max = 6494	**2107 +/- 778** Min = 1201 Max = 4659
SMRT-seq coverage (x)	**177.4 +/- 117.0** Min = 75.1 Max = 1392.6	**205.3 +/- 136.1** Min = 77.8 Max = 918.4
Total number of 6mA marks in genome	46,322	2,344
6mA sites per chromosome	**12 +/- 8** Min = 0 Max = 73	**24 +/- 16** Min = 0 Max = 73
AT content (%)	**67.8 +/- 3.0** Min = 55.7 Max = 76.2	**66.5 +/- 2.7** Min = 60.2 Max = 72.2
RNAseq (FPKM)	**34.4 +/- 75.2** Min = 0.0 Max = 1444.5	**53.7 +/- 71.5** Min = 0.1 Max = 424.8

Properties of *Oxytricha* chromosomes in native genomic DNA and mini-genome DNA. "+/-" indicates one standard deviation above or below the mean.

Example 7

Modular Synthesis of Epigenetically Defined Chromosomes

[0215] The above experiments used kinetic signatures from SMRT-seq data to infer the presence of 6mA marks in genomic DNA. We next sought to confirm that 6mA is directly responsible for disfavoring nucleosomes *in vitro*, and to understand how this effect could be overcome by cellular factors. 6mA-containing oligonucleotides were annealed and subsequently ligated with DNA building blocks to form full-length chromosomes. Importantly, these chromosomes contain 6mA at all

locations identified by SMRT-seq *in vivo*. The representative chromosome, Contig1781.0, is 1.3 kb, contains a clearly defined TSS, and encodes a single highly transcribed gene with a predicted RING finger domain. The length and gene structure are characteristic of typical *Oxytricha* chromosomes (Fig. 5A). We independently validated the location of 6mA *in vivo* by sequencing chromosomal DNA immunoprecipitated with an anti-6mA antibody (Fig. 5A).

[0216] Four chromosome variants were synthesized, with cognate 6mA sites on neither, one, or both DNA strands (chromosomes 1-4 in Figs. 5B and 5C). Chromatin was assembled by salt dialysis with either *Oxytricha* or *Xenopus* nucleosomes and subsequently digested with MNase to obtain mononucleosomal DNA (Figs. 6A and 13G). Tiling qPCR was used to quantify nucleosome occupancy at ~50 bp increments along the entire length of the synthetic chromosome (Fig. 6B). The fully methylated locus exhibits a ~46% reduction in nucleosome occupancy relative to the unmethylated variant, while hemimethylated chromosomes containing half the number of 6mA marks showed intermediate nucleosome occupancy at the corresponding region (Fig. 6B). The reduction in nucleosome occupancy was confined to the methylated region and not observed across the rest of the chromosome. Similar trends were observed when chromatin was assembled using the NAP1 histone chaperone (Fig. 14F. top panel). indicating that this effect is not an artifact of the salt dialysis method. Furthermore, moving 6mA to an ectopic location (chromosome 5 in Figs. 5B and 5C) decreases nucleosome occupancy at that site (Fig. 6C). We conclude that 6mA directly disfavors nucleosome occupancy in a local, quantitative manner *in vitro*.

Example 8

Chromatin Remodelers Restore Nucleosome Occupancy over 6mA Sites

[0217] Nucleosome occupancy *in vivo* is influenced not only by DNA sequences but also by trans-acting factors such as ATP-dependent chromatin remodeling factors (Struhl and Segal, 2013). We used synthetic, methylated chromosomes to test how the well-studied chromatin remodeler ACF responds to 6mA in native DNA. ACF generates regularly spaced nucleosome arrays *in vitro* and in vivo (Clapier and Cairns, 2009; Ito et al., 1997). Its catalytic subunit ISWI is conserved across eukaryotes, including *Oxytricha* and *Tetrahymena* (Table 5).

Synthetic chromosomes were assembled into chromatin by salt dialysis as before and then incubated with ACF in the presence of ATP (Figs. 13H and 6D). We find that ACF partially -but not completely- restores nucleosome occupancy over the methylated locus in an ATP-dependent manner (Fig. 6D). This effect is observed when ACF was added to chromatin assembled by salt dialysis or the NAP1 histone chaperone (Figs. 6D and 14F). ACF also restores nucleosome occupancy over methylated loci in native genomic DNA (Figs. 6E and 13I), indicating that the effect is not restricted to a single chromosome. This result is robust to the extent of MNase digestion (Fig. 14B). Although the heterologous system used here may differ from endogenous chromatin assembly factors in *Oxytricha*, our experiment illustrates the principle that trans-acting factors can counteract or even overcome the effect of 6mA on nucleosome organization.

Example 9

Disruption of MTA1 Impacts Gene Expression and Sexual Development

[0218] Since *mta1* mutants exhibit genome-wide loss of 6mA, we assayed these cells for transcriptional changes by poly(A)$^+$ RNAseq. Only a small minority of genes show significant changes in gene expression (10% false discovery rate [FDR]; Fig. 7A). To examine the methylation status of these differentially expressed genes, we grouped them according to "starting" methylation level, as defined by the total number of 6mA marks near the TSS in wild-type cells. Genes exhibit two distinct transcriptional responses: those with low starting levels of 6mA exhibit a small change in 6mA between wild-type and mutant cells (Fig. 3D) and tend to be significantly upregulated in mutant lines ($p = 2.8 \times 10^{-9}$, Fisher's exact test; Fig. 7B). Surprisingly, genes with high starting 6mA are not enriched in differentially expressed genes ($p > 0.1$, Fisher's exact test), even though they exhibit greater loss of 6mA in mutants (Fig. 3D). Steady-state RNA-seq levels are therefore largely robust to drastic changes in 6mA levels. Since most, but not all, 6mA is lost from mtal mutants (Fig. 3C), it is also possible that residual DNA methylation across the genome sufficiently buffers genes from changes in transcription.

[0219] Because the aforementioned phenotypic changes were assayed in vegetative *Oxytricha* cells, we asked whether MTA1 may play roles outside of this developmental state. MTA1 transcript levels are markedly upregulated in the sexual

cycle, as assayed by poly(A). RNA-seq (Fig. 7C). Strikingly, mtal mutants fail to complete the sexual cycle when induced to mate and display complete lethality (Fig. 7D). Our data do not exclude the possibility that m6A RNA methylation, in addition to 6mA DNA methylation, is also impacted by MTA1 loss during development. Further studies would clarify the role of MTA1 in these pathways.

Example 10

DISCUSSION

[0220] The present disclosure has identified MTA1c as a conserved, hitherto undescribed 6mA methyltransferase. It consists of two MT-A70 proteins (MTA1/MTA9) and two homeobox-like proteins (p1/p2). The composition of MTA1c provides immediate insights into how it specifically methylates DNA (Fig. 7F). MTA1 likely mediates transfer of the methyl group from SAM to the acceptor adenine moiety, given that it contains conserved amino acid residues implicated in catalysis and SAM binding (Fig. 10E). Indeed, we show that these residues are necessary for its activity (Fig. 2E). While MTA1 constitutes the catalytic center, it lacks a CCCH-type zinc finger domain that is necessary for RNA binding in the canonical m6A methyltransferase METTL3. Instead, nucleic acid binding is likely assumed by the homeobox-like domains in p1 and p2, which are known to specifically engage dsDNA through helix-turn-helix motifs.

[0221] The observation that MTA1c is more active in the presence of pre-methylated DNA templates is reminiscent of the CpG methyltransferase DNMT1. Yet, MTA1c and DNMT1 exhibit distinct protein domain architectures. Further biochemical studies are required to elucidate the molecular basis of this property. A distinct MT-A70 protein, named TAMT-1, was recently reported to act as a 6mA methyltransferase in Tetrahymena, (Luo et al., 2018), suggesting that multiple enzymes mediate 6mA deposition. It remains to be determined how MTA1c and TAMT-1 collectively mediate DNA methylation at various developmental stages, and whether cross-talk occurs between these enzymes.

[0010] In addition to identifying the ciliate 6mA methyltransferase, we investigated the function of 6mA in vitro by building epigenetically defined chromosomes. We show that 6mA directly disfavors nucleosome occupancy in a local, quantitative manner, independent of DNA sequence (Fig. 7E). Our

experiments do not reveal exactly how 6mA disfavors nucleosome occupancy. Early studies suggest that 6mA destabilizes dA:dT base pairing, leading to a decrease in the melting temperature of DNA (Engel and von Hippel, 1978). Whether this or some other property of 6mA contributes to lowered nucleosome stability awaits further investigation.

[0222] Intriguingly, nucleosome organization exhibits only subtle changes after genome-wide loss of 6mA (Fig. 7E). Only a small set of genes (<10%) is transcriptionally dysregulated. It is possible that residual 6mA in mtal mutants could mask relevant phenotypes. Nonetheless, our results caution against interpreting 6mA function solely based on correlation with genomic elements. We also find that 6mA intrinsically disfavors nucleosomes *in vitro*, but —crucially —this effect can be overridden by distinct factors *in vitro* and *in vivo*. We propose that phased nucleosome arrays are first established in vivo, which then restrict MTA1-mediated methylation to linker regions due to steric hindrance. This in turn decreases the fuzziness of flanking nucleosomes, reinforcing chromatin organization. Therefore, 6mA tunes nucleosome organization *in vivo*. Our data do not support the hypothesis that nucleosome phasing is established by predeposited 6mA.

[0223] More broadly, our work showcases the utility of *Oxytricha* chromosomes for advancing chromatin biology. By extending current technologies (Muller et al., 2016), it should be feasible to introduce both modified nucleosomes and DNA methylation in a site-specific manner on full-length chromosomes. Such "designer" chromosomes will serve as powerful tools for studying DNA-templated processes such as transcription within a fully native DNA environment.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Ammermann, D., Steinbruck, G., Baur, R., and Wohlert, H. (1981). Methylated bases in the DNA of the ciliate Stylonychia mytilus. Eur. J. Cell Biol. 24, 154-156.

An, W., and Roeder, R.G. (2004). Reconstitution and transcriptional analysis of chromatin in vitro. Methods Enzymol. 377, 460-474.

Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T.R. (2013). Highfidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res. 23, 169-180.

Beh, L.Y., Muller, M.M., Muir, T.W., Kaplan, N., and Landweber, L.F. (2015). DNA-guided establishment of nucleosome patterns within coding regions of a eukaryotic genome. Genome Res. 25, 1727-1738.

Beh et al., Identification of a DNA N6-Adenine Methyltransferase Complex and Its Impact on Chromatin Organization, Cell (2019), https://doi.org/10.1016/j.ce11.2019.04.028.

Bern, M., Kil, Y.J., and Becker, C. (2012). Byonic: Advanced Peptide and Protein Identification Software. Curr. Protoc. Bioinformatics. 13, 13.20.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. 19, 19.10.1-19.10.21.

Bracht, J.R., Fang, W., Goldman, A.D., Dolzhenko, E., Stein, E.M., and Landweber, L.F. (2013). Genomes on the edge: programmed genome instability in ciliates. Cell 152, 406-416.

Bromberg, S., Pratt, K., and Hattman, S. (1982). Sequence specificity of DNA adenine methylase in the protozoan Tetrahymena thermophila. J. Bacteriol. 150, 993-996.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84, 843-851.

Cassidy-Hanley, D.M. (2012). Tetrahymena in the Laboratory: Strain Resources, Methods for Culture, Maintenance, and Storage. Methods Cell Biol. 109,237-276.

Chen, X., Bracht, J.R., Goldman, A.D., Dolzhenko, E., Clay, D.M., Swart, E.C., Perlman, D.H., Doak, T.G., Stuart, A., Amemiya, C.T., et al. (2014). The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. Cell 158, 1187-1198.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. 78, 273-304.

Cummings, D.J., Tait, A., and Goddard, J.M. (1974). Methylated bases in DNA from Paramecium aurelia. Biochim. Biophys. Acta 374, 1-11.

Debelouchina, G.T., Gerecht, K., and Muir, T.W. (2017). Ubiquitin utilizes an acidic surface patch to alter chromatin structure. Nat. Chem. Biol. 13, 105-110.

Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., et al. (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol. 4, e286.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976-989.

Engel, J.D., and von Hippel, P.H. (1978). Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. J. Biol. Chem. 253, 927-934.

Fang, W., Wang, X., Bracht, J.R., Nowacki, M., and Landweber, L.F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell 151, 1243-1255.

Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F., Bateman, A., and Eddy, S.R. (2015). HMMER web server 2015 update. Nucleic Acids Res. 43 (W1), W30-W38.

Fioravanti, A., Fumeaux, C., Mohapatra, S.S., Bompard, C., Brilli, M., Frandi, A., Castric, V., Villeret, V., Viollier, P.H.P., and Biondi, E.G. (2013). DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in Caulobacter crescentus and other Alphaproteobacteria. PloS Genet. 9, e1003541.

Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C., et al. (2015). N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell 161, 879-892.

Fyodorov, D.V., and Kadonaga, J.T. (2003). Chromatin assembly in vitro with purified recombinant ACF and NAP-1. Methods Enzymol. 371, 499-515.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15, 1451-1455.

Goecks, J., Nekrutenko, A., and Taylor, J.; Galaxy Team (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 11, R86.

Gorovsky, MA., Hattman, S., and Pleger, G.L. (1973). (6 N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. J. Cell Biol. 56, 697-701.

Gottschling, D.E., and Cech, T.R. (1984). Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex. Cell 38, 501-510.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA Methylation on N6-Adenine in C. elegans. Cell 161, 868-878.

Haberle, V., Forrest, A.R.R., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015). CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. 43, e51.

Hattman, S., Kenny, C., Berger, L, and Pratt, K. (1978). Comparative study of DNA methylation in three unicellular eucaryotes. J. Bacteriol. 135, 1156-1157.

Horton, J.R., Liebert, K., Bekes, M., Jeltsch, A., and Cheng, X. (2006). Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J. Mol. Biol. 358, 559-570.

Huang, Y., Niu, B., Gao, Y., Fu, L., and U, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680-682.

Huang, J., Dong, X., Gong, Z., Qin, L -Y., Yang, S., Zhu, Y.-L, Wang, X., Zhang, D., Zou, T., Yin, P., et al. (2019). Solution structure of the RNA recognition domain of METTL3-METTL14 N6-methyladenosine methyltransferase. Protein Cell 10, 272-284.

Huff, J.T., and Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156,1286-1297.

Ito, T., Bulger, M., Pazin, M.J., Kobayashi, R., and Kadonaga, J.T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. Cell 90, 145-155.

Iyer, LM., Zhang, D., and Aravind, L (2016). Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. BioEssays 38, 27-40.

Karrer, K.M., and VanNuland, T.A. (1999). Nucleosome positioning is independent of histone H1 in vivo. J. Biol. Chem. 274,33020-33024.

Katoh, K., Rozewicki, J., and Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. 26, 1351-1359.

Khurana, J.S., Wang, X., Chen, X., Perlman, D.H., and Landweber, LF. (2014). Transcription-independent functions of an RNA polymerase II subunit, Rpb2, during genome rearrangement in the ciliate, Oxytricha trifallax. Genetics 197, 839-849.

Khurana, J.S., Clay, D.M., Moreira, S., Wang, X., and Landweber, L.F. (2018). Small RNA-mediated regulation of DNA dosage in the ciliate Oxytricha. RNA 24,18-29.

Koziol, M.J., Bradshaw, C.R., Allen, G.E., Costa, A.S.H., Frezza, C., and Gurdon, J.B. (2016). Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat. Struct. Mol. Biol. 23, 24-30.

Kuraku, S., Zmasek, C.M., Nishimura, 0., and Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41, W22-W28.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. 18, 548-562.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9,357-359.

Laughlin, T.J., Henry, J.M., Phares, E.F., Long, M.V., and Olins, D.E. (1983). Methods for the Large-Scale Cultivation of an Oxytricha (Ciliophora: Hypotrichide). J. Protozool. 30, 63-64.

Lauth, M.R., Spear, B.B., Neumann, J., and Prescott, D.M. (1976). DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of Oxytricha. Cell 7,67-74.

Lawn, R.M., Heumann, J.M., Herrick, G., and Prescott, D.M. (1978). The genesize DNA molecules in Oxytricha. Cold Spring Harb. Symp. Quant. Biol. 42, 483-492.

Liang, Z., Shen, L., Cui, X., Bao, S., Geng, Y., Yu, G., Liang, F., Xie, S., Lu, T., Gu, X., and Yu, H. (2018). DNA N6-Adenine Methylation in Arabidopsis thaliana. Dev. Cell 45,406-416.

Lieleg, C., Ketterer, P., Nuebler, J., Ludwigsen, J., Gerland, U., Dietz, H., Mueller-Planitz, F., and Korber, P. (2015). Nucleosome spacing generated by ISWI and CHD1 remodelers is constant regardless of nucleosome density. Mol. Cell. Biol. 35,1588-1605.

Liu, Y., Tavema, S.D., Muratore, T.L, Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev. 21, 1530-1545.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L, Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. 10, 93-95.

Liu, J., Zhu, Y., Luo, G.-Z., Wang, X., Yue, Y., Wang, X., Zong, X., Chen, K., Yin, H., Fu, Y., et al. (2016). Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat. Commun. 7,13052.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25,402-408.

Lugar, K., Rcchatcincr, T.J., and Richmond, T.J. (1900). Proparation of nucleosome core particle from recombinant histones. Methods Enzymol. 304, 3-19.

Luo, G.-Z., Blanco, M.A., Greer, E.L., He, C., and Shi, Y. (2015). DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? Nat. Rev. Mol. Cell Biol. 16, 705-710.

Luo, G.-Z., Hao, Z., Luo, L, Shen, M., Sparvoli, D., Zheng, Y., Zhang, Z., Weng, X., Chen, K., Cui, Q., et al. (2018). N6-methyldeoxyadenosine directs nucleosome positioning in Tetrahymena DNA. Genome Biol. 19, 200.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, LP., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. 18, 1073-1083.

Miao, W., Xiong, J., Bowen, J., Wang, W., Liu, Y., Braguinets, 0., Grigull, J., Pearlman, R.E., Orias, E., and Gorovsky, MA. (2009). Microarray analyses of gene expression during the Tetrahymena thermophila life cycle. PLoS ONE 4,e4429.

Miller, MA., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA. pp. 1-8.

Mondo, S.J., Dannebaum, R.O., Kuo, R.C., Louie, K.B., Bewick, A.J., LaButti, K., Haridas, S., Kuo, A., Salamov, A., Ahrendt, S.R., et al. (2017). Widespread adenine N6-methylation of active genes in fungi. Nat. Genet. 49, 964-968.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621-628.

M.M., Fierz, B., Bittova, L., Liszczak, G., and Muir, T.W. (2016). A two-state activation mechanism controls the histone methyltransferase Suv39h1. Nat. Chem. Biol. 12, 188-193.

Murray, i.A., Morgan, R.D., Luyten, Y., Fomenkov, A., Correa, i.R., jr., Dai, Allaw, M.B., Zhang, X., Cheng, X., and Roberts, R.J. (2018). The non-specific adenine DNA methyltransferase M.EcoGII. Nucleic Acids Res. 46, 840-848.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646-4658.

Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 451, 153-158.

Pendleton, K.E., Chen, B., Liu, K., Hunter, 0.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell 169, 824-835.

Pratt, K., and Hattman, S. (1981). Deoxyribonucleic acid methylation and chromatin organization in Tetrahymena thermophila. Mol. Cell. Biol. 1, 600-608.

Prescott, D.M. (1994). The DNA of ciliated protozoa. Microbiol. Rev. 58, 233-267.

Rae, P.M., and Spear, B.B. (1978). Macronuclear DNA of the hypotrichous ciliate Oxytricha fallax. Proc. Natl. Acad. Sci. USA 75, 4992-4996.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896-1906.

Schaffer, A.A., Aravind, L, Madden, T.L, Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., and Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 29, 2994-3005.

Schiffers, S., Ebert, C., Rahimoff, R., Kosmatchev, 0., Steinbacher, J., Bohne, A.-V., Spada, F., Michalakis, S., Nickelsen, J., Mailer, M., and Carell, T. (2017). Quantitative LC-MS Provides No Evidence for m6 dA or m4 dC in the

Genome of Mouse Embryonic Stem Cells and Tissues. Angew. Chem. Int. Ed. Engl. 56, 11268-11271.

Sledz, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. eLife 5. Published online September 14, 2016. https://doi.org/10.7554/eLife.18434.

Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc. Natl. Acad. Sci. USA 96, 14967-14972.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. 20, 267-273.

Swart, E.G., Bracht, J.R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J.S., Goldman, AD., Nowacki, M., Schotanus, K., et al. (2013). The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol. 11, e1001473.

Tavema, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell 110, 701-711.

Wada, R.K., and Spear, B.B. (1980). Nucleosomal organization of macronuclear chromatin in Oxytricha fallax. Cell Differ. 9, 261-268.

Wang, P., Doxtader, K.A., and Nam, Y. (2016a). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Mol. Cell 63, 306-317.

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., et al. (2016b). Structural basis of N(6)-adenosine methylation by the METTL3-METTL14 complex. Nature 534, 575-578.

Wang, Y., Chen, X., Sheng, Y., Liu, Y., and Gao, S. (2017). N6-adenine DNA methylation is associated with the linker DNA of H2A2-containing well-positioned nucleosomes in Pol II-transcribed genes in Tetrahymena. Nucleic Acids Res. 45, 11594-11606.

Warda, A.S., Kretschmer, J., Heckert, P., Lenz, C., Urlaub, H., Hobartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N6-methyladenosine (m6A) methyltransferase that targets pre-mRNAs and various noncoding RNAs. EMBO Rep. 18, 2004-2014.

Wei, Y., Mizzen, C.A., Cook, R.G., Gorovsky, M.A., and Allis, C.D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proc. Natl. Acad. Sci. USA 95, 7480-7484.

Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., et al. (2016). DNA methylation on N(6)-adenine in mammalian embryonic stem cells. Nature 532, 329-333.

Xiao, R., and Moore, D.D. (2011). DamIP: Using Mutant DNA Adenine Methyltransferase to Study DNA-Protein Interactions In Vivo. Curr. Protoc. Mol. Biol. 21. https://doi.org/10.1002/0471142727.mb2121s94.

Xiao, C.-L., Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y., Yu, G., Liu, J., Xie, S.-Q., Luo, F., et al. (2018). N6-Methyladenine DNA Modification in the Human Genome. Mol. Cell 71, 306-318.

Xiong, J., Lu, X., Lu, Y., Zeng, H., Yuan, D., Feng, L, Chang, Y., Bowen, J., Gorovsky, M., Fu, C., and Miao, W. (2011). Tetrahymena Gene Expression Database (TGED): a resource of microarray data and co-expression analyses for Tetrahymena. Sci. China Life Sci. 54, 65-67.

Xiong, J., Lu, X., Zhou, Z., Chang, Y., Yuan, D., Tian, M., Zhou, Z., Wang, L, Fu, C., Orias, E., and Miao, W. (2012). Transcriptome analysis of the model protozoan, Tetrahymena thermophila, using Deep RNA sequencing. PLoS ONE 7, e30630.

Yao, B., Cheng, Y., Wang, Z., Li, Y., Chen, L, Huang, L., Zhang, W., Chen, D., Wu, H., Tang, B., and Jin, P. (2017). DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. Nat. Commun. 8, 1122.

Yao, B., Li, Y., Wang, Z., Chen, L., Poidevin, M., Zhang, C., Lin, L, Wang, F., Bao, H., Jiao, B., et al. (2018). Active N 6 -Methyladenine Demethylation by DMAD Regulates Gene Expression by Coordinating with Polycomb Protein in Neurons. Mol. Cell 71, 848-857.

Yerlici, V.T., and Landweber, L.F. (2014). Programmed Genome Rearrangements in the Ciliate Oxytricha. Microbiol. Spectr. 2. Published online December 2014. 10.1128/microbiolspec.MDNA3-0025-2014.

Zhang, Z., and Pugh, B.F. (2011). High-resolution genome-wide mapping of the primary structure of chromatin. Cell 144, 175-186.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). N6-methyladenine DNA modification in Drosophila. Cell 161, 893-906.

Zhou, C., Wang, C., Liu, H., Zhou, Q., Liu, Q., Guo, Y., Peng, T., Song, J., Zhang, J., Chen, L., et al. (2018). Identification and analysis of adenine N6-methylation sites in the rice genome. Nat. Plants 4, 554-563.

[0224] The embodiments described in this disclosure can be combined in various ways. Any aspect or feature that is described for one embodiment can be incorporated into any other embodiment mentioned in this disclosure. While various novel features of the inventive principles have been shown, described and pointed out as applied to particular embodiments thereof, it should be understood that various omissions and substitutions and changes may be made by those skilled in the art without departing from the spirit of this disclosure. Those skilled in the art will appreciate that the inventive principles can be practiced in other than the described embodiments, which are presented for purposes of illustration and not limitation.

What is claimed is:

1. A method of treating or ameliorating the effects of a disease characterized by an abnormal level of m6dA in a subject, comprising administering to the subject an amount of MTA1c or any components thereof effective to modulate m6dA levels in the subject.
2. The method according to claim 1, wherein the modulation comprises restoring m6dA levels to normal or near-normal ranges in the subject.
3. The method according to claim 1, wherein the disease is a cancer.
4. The method according to claim 3, wherein the cancer is gastric cancer or liver cancer.
5. The method according to claim 4, further comprising administering to the subject one or more of anti-gastric cancer and anti-liver cancer drugs.
6. The method according to claim 1, furthering comprising co-administering to the subject an epigenetic agent.
7. The method according to claim 6, wherein the epigenetic agent is selected from the group consisting of methylation inhibiting drugs, Bromodomain inhibitors, histone acetylase (HAT) inhibitors, protein methyltransferase inhibitors, histone methylation inhibitors, histone deacetlyase (HDAC) inhibitors, histone acetylases, histone deacetlyases, and combinations thereof.
8. A pharmaceutical composition comprising MTA1c or any components thereof that is effective to modulate m6dA levels in a subject in need thereof and a pharmaceutically acceptable carrier, diluent, adjuvant or vehicle.
9. A method of modifying a nucleic acid from a cell, the cell derived from a multicellular eukaryote, comprising the steps of:
   (a) obtaining the nucleic acid from the cell; and
   (b) contacting the nucleic acid with MTA1c or any components thereof under conditions effective to methylate the nucleic acid.
10. The method according to claim 9, wherein the methylated nucleic acid is effective to modulate nucleosome organization and transcription.
11. The method according to claim 9, wherein the modification is a DNA N6-adenine methylation.

12. The method according to claim 11, wherein the DNA N6-adenine methylation is one or more of dimethylated AT (5'-A*T-3'/3'-TA*-5'), dimethylated TA (5'-TA*-3'/3'-A*T-5'), dimethylated AA (5'-A*A*-3'/3'-TT-5'), methylated AT (5'-A*T-3'/3'-TA-5'), methylated AA (5'-A*A-3'/3'-TT-5'), methylated AC (5'-A*C-3'/3'-TG-5'), methylated AG (5'-A*G-3'/3'-TC-5'), methylated TA (5'-TA*-3'/3'-AT-5'), methylated AA (5'-AA*-3'/3'-TT-5'), methylated CA (5'-CA*-3'/3'-GT-5'), and methylated GA (5'-GA* -3'/3'-CT-5').
13. The method according to claim 9, wherein the MTA1c or any components thereof comprises a mutation effective to abrogate dimethylation of the nucleic acid.
14. The method according to claim 13, wherein the mutation comprises loss of a C-terminal methyltransferase domain.
15. The method according to claim 9, wherein the MTA1c or any components thereof is obtained from ciliates, algae, or basal fungi.
16. The method according to claim 9, wherein the MTA1c or any components thereof is obtained from *Oxytricha* or *Tetrahymena*.
17. A cell line obtained from a multicellular eukaryote comprising a nucleic acid encoding MTA1c or any components thereof and/or an MTA1c protein complex or any components thereof.
18. The eukaryotic cell according to claim 17, wherein the nucleic acid encoding MTA1c or any components thereof is operably linked to a recombinant expression vector.
19. A method of identifying protein binding sites on DNA comprising the steps of:
    (a) providing DNA;
    (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA;
    (c) contacting the DNA with one or more proteins;
    (d) contacting the DNA with an enzyme effective to hydrolyze the DNA in positions where no protein binding occurs;
    (e) removing the DNA bound protein; and
    (f) isolating and sequencing the DNA fragments.
20. The method according to claim 19, wherein the one or more proteins comprise histone octamers.

Fig. 1A

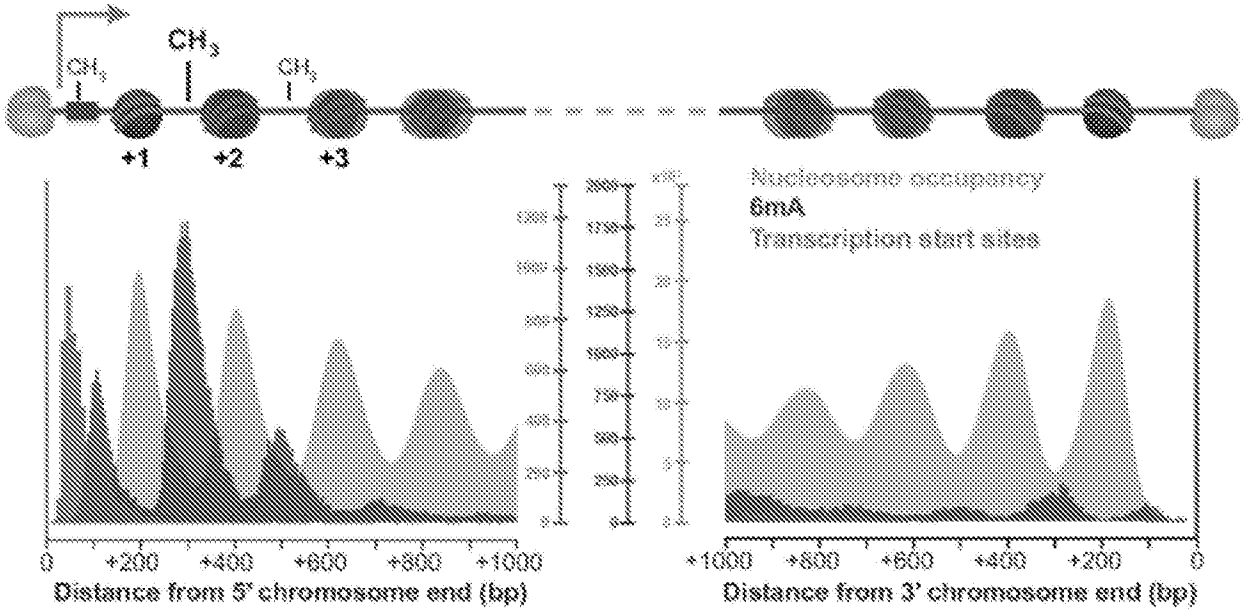

CH3
CH3
CH3
+1
+2
+3
Nucleosome occupancy
6mA
Transcription start sites
0
+200
+400
+600
+800
+1000
Distance from 5' chromosome end (bp)
+1000
+800
+600
+400
+200
0
Distance from 3' chromosome end (bp)

Fig. 1B

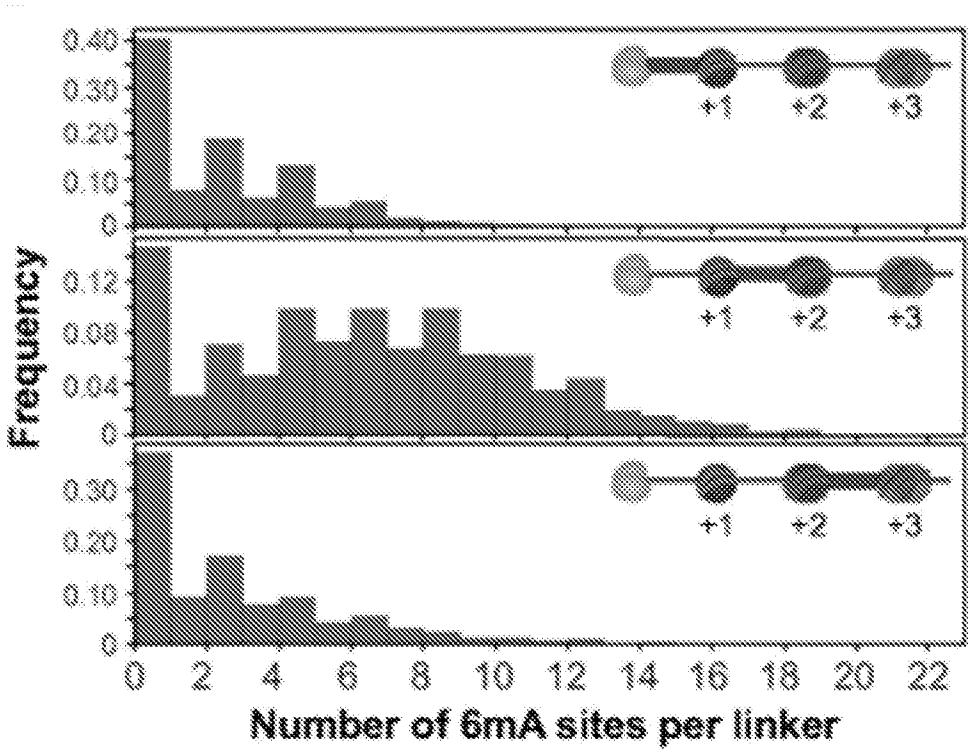

Frequency
0.40
0.30
0.20
0.10
0
0.12
0.08
0.04
0
0.30
0.20
0.10
0
+1
+2
+3
0
2
4
6
8
10
12
14
16
18
20
22
Number of 6mA sites per linker

Fig. 1C

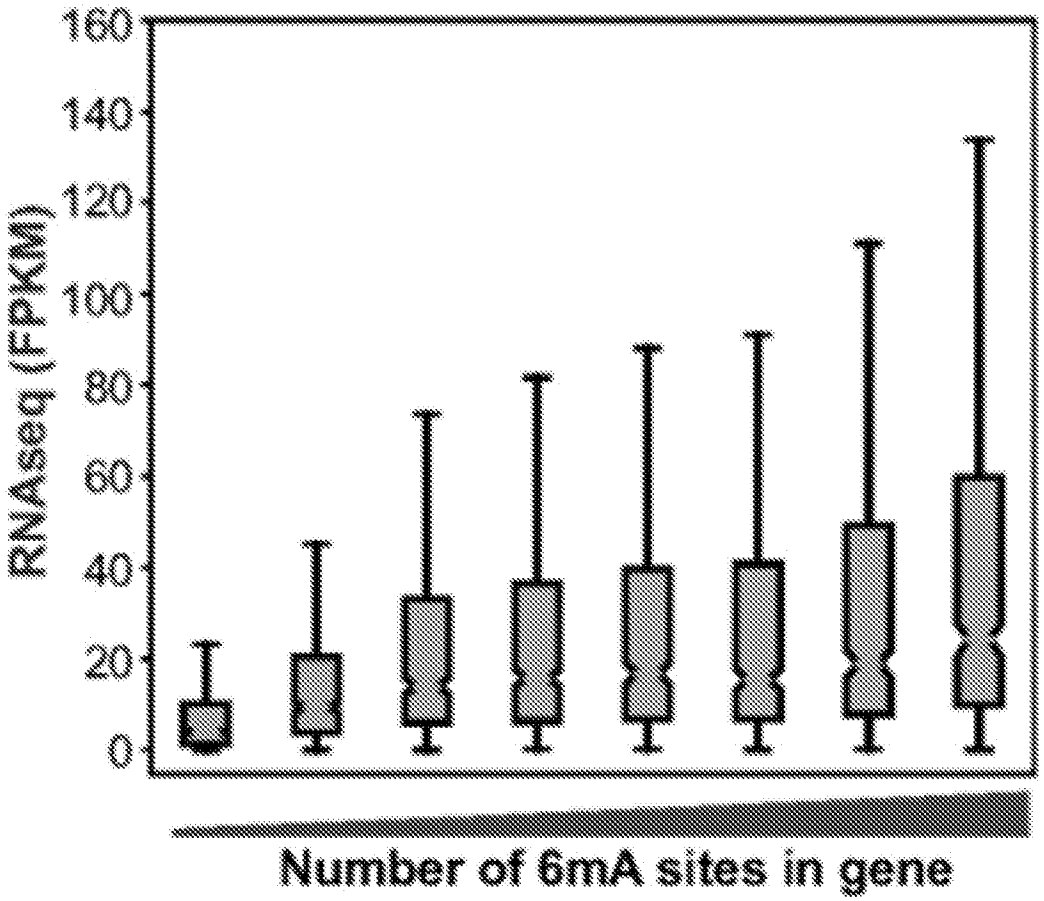

160
140
120
100
80
60
40
20
0
RNAseq (FPKM)
Number of 6mA sites in gene

Fig. 1D


Probability
1.0
0.5
0.0
*

Fig. 1E

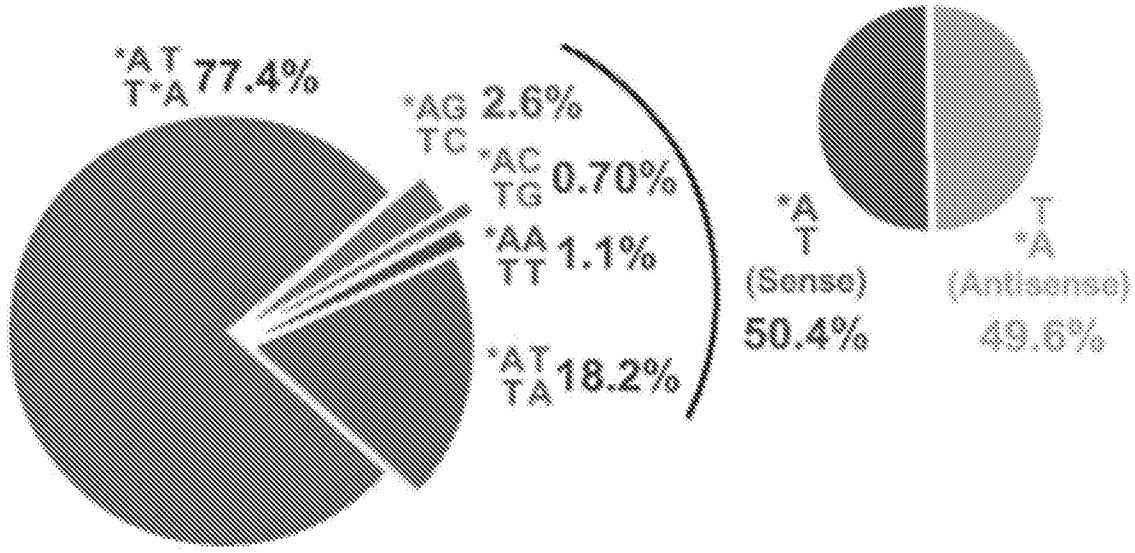

*AT
T*A 77.4%
*AG
TC 2.6%
*AC
TG 0.70%
*AA
TT 1.1%
*AT
TA 18.2%
*A
T
(Sense)
50.4%
T
*A
(Antisense)
49.6%

Fig. 2A

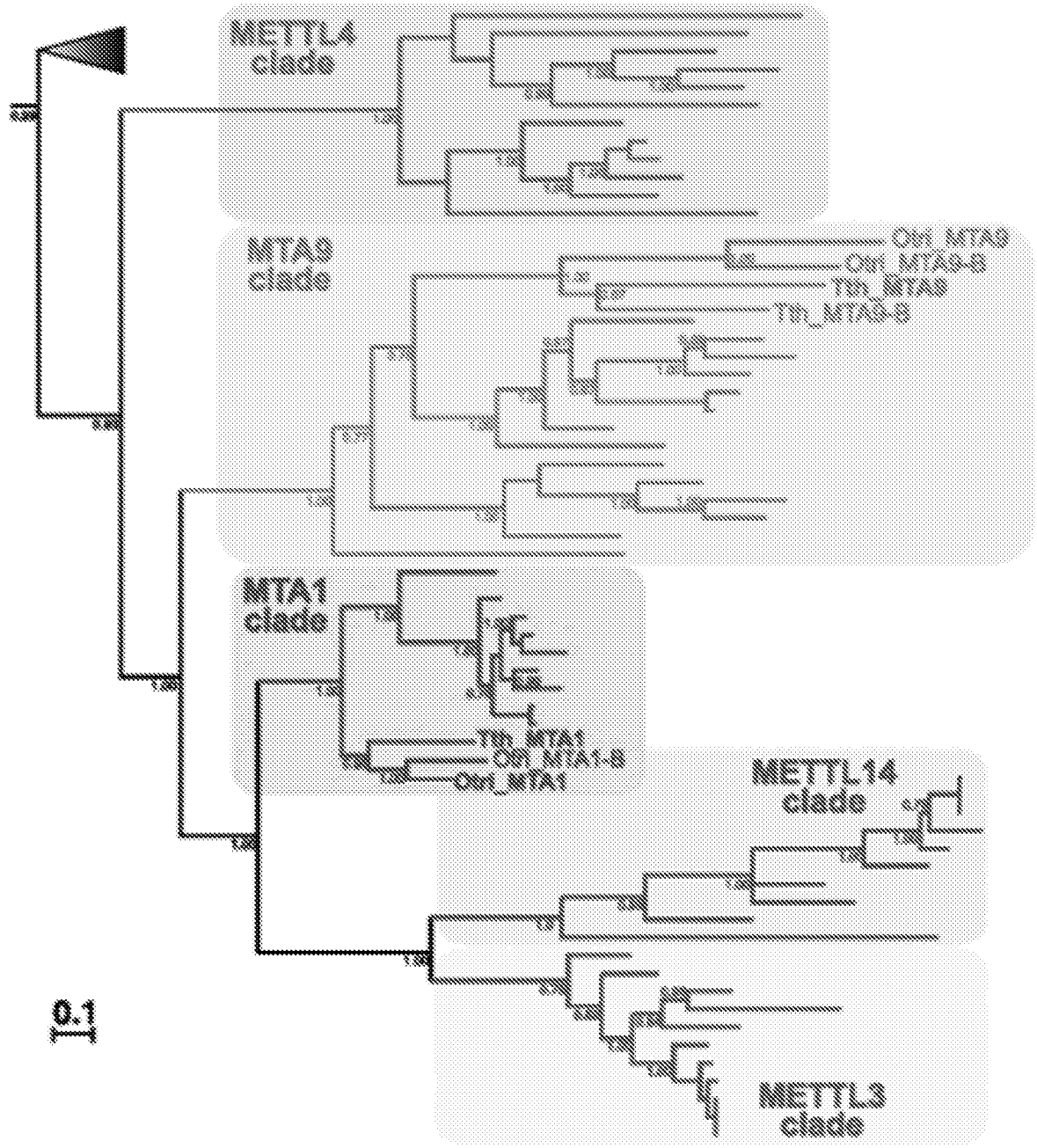

METTL4 clade
MTA9 clade
Ore_MTA9
Ore_MTA9-B
Tbr_MTA9
Tbr_MTA9-B
MTA1 clade
Tbr_MTA1
Ore_MTA1-B
Ore_MTA1
METTL14 clade
METTL3 clade
0.1

Fig. 2B

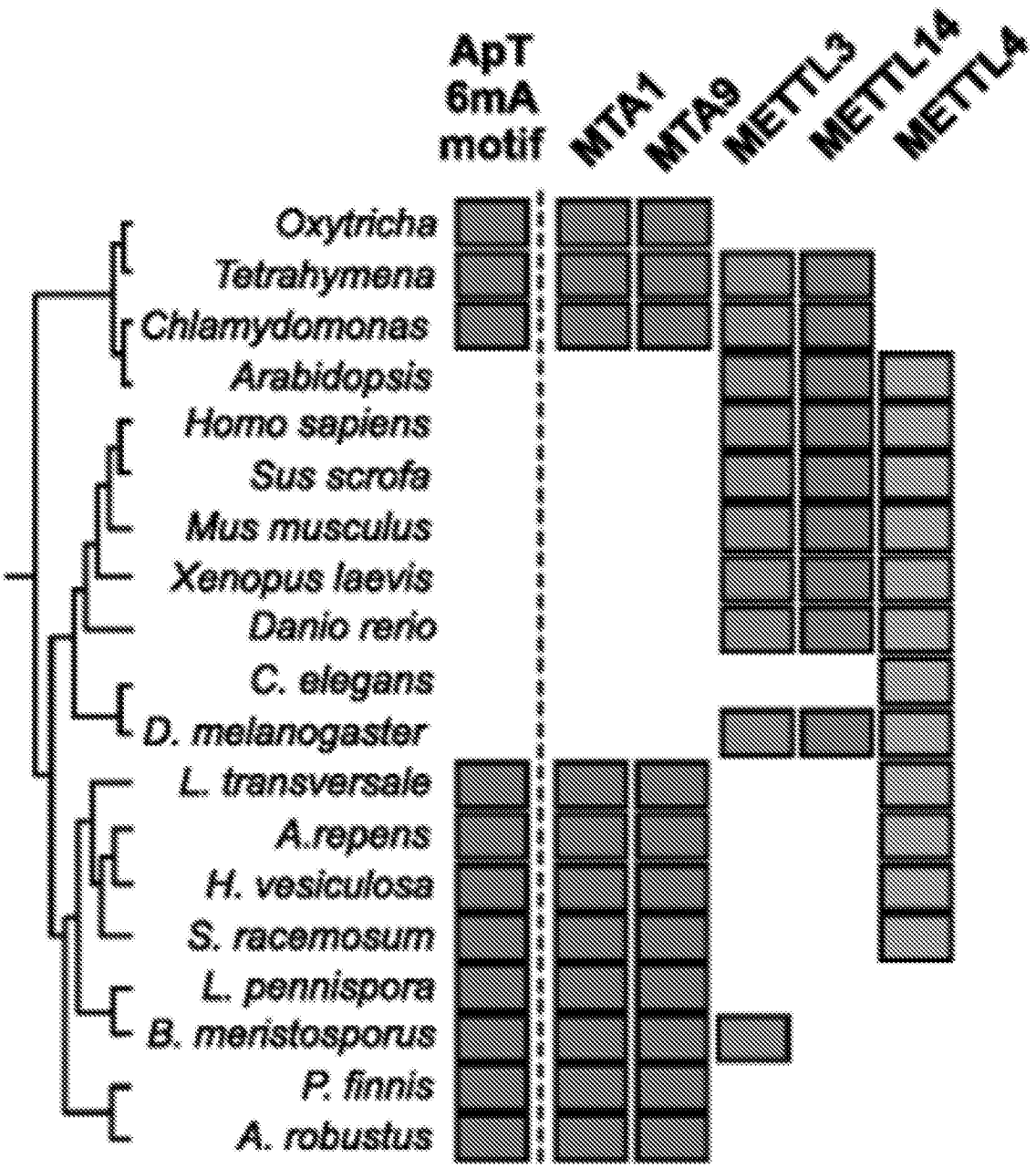

ApT
6mA
motif
MTA1
MTA9
METTL3
METTL14
METTL4
Oxytricha
Tetrahymena
Chlamydomonas
Arabidopsis
Homo sapiens
Sus scrofa
Mus musculus
Xenopus laevis
Danio rerio
C. elegans
D. melanogaster
L. transversale
A.repens
H. vesiculosa
S. racemosum
L. pennispora
B. meristosporus
P. finnis
A. robustus

Fig. 2C

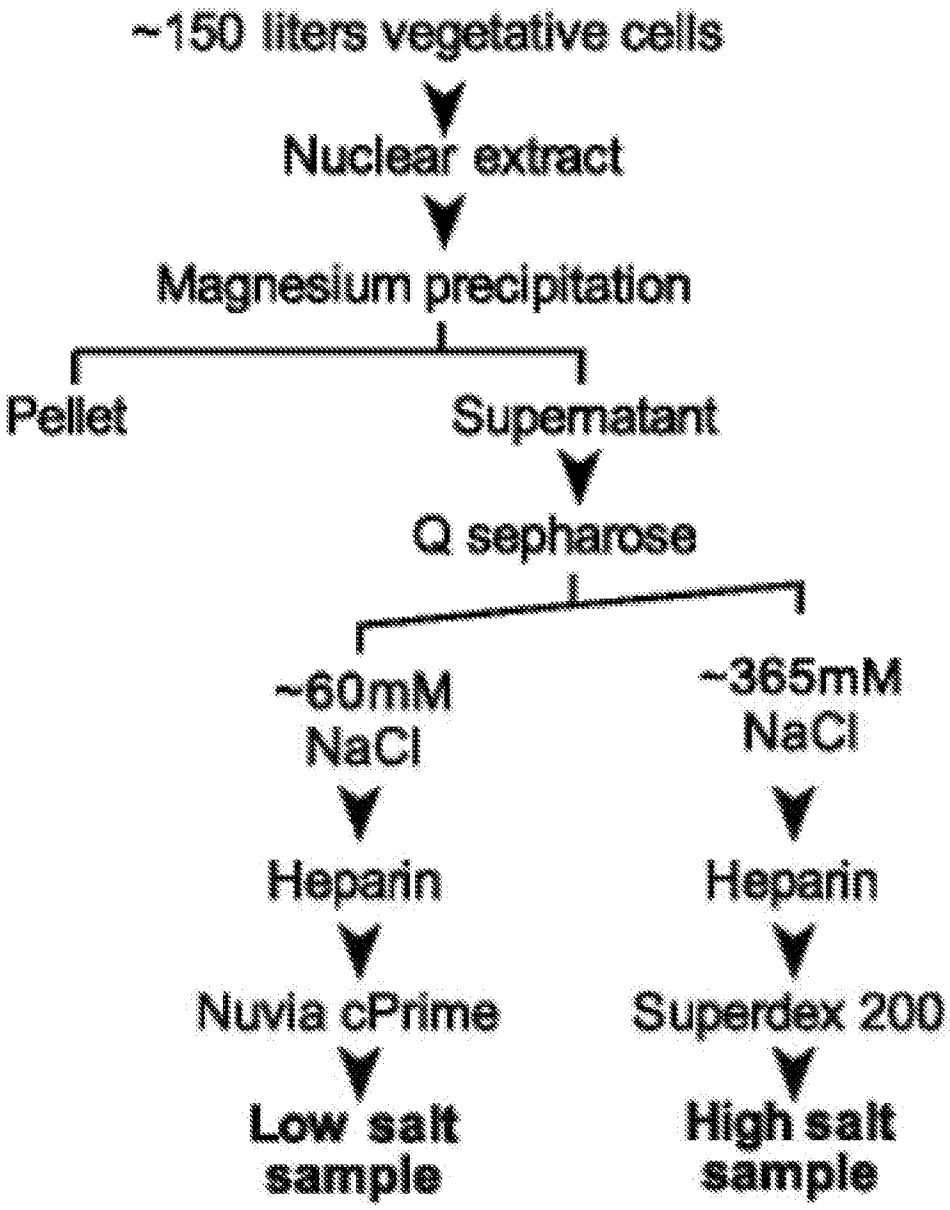

~150 liters vegetative cells
Nuclear extract
Magnesium precipitation
Pellet
Supernatant
Q sepharose
~60mM NaCl
~365mM NaCl
Heparin
Heparin
Nuvia cPrime
Superdex 200
Low salt sample
High salt sample

Fig. 2D

RNAseq (FPKM)		Low Salt sample			High Salt sample			DNA methylase activity
		Low	Mid	High	Low	Mid	High	
43.10	MTA1	1	18	49	1	36	101	Total spectral counts (LC-MS/MS)
25.27	MTA9	0	2	29	5	29	90	
47.25	p1	0	6	24	8	30	98	
16.54	p2	1	9	31	5	11	44	
23.65	MTA2	0	3	1	0	0	0	
7.74	MTA3	0	0	0	0	0	0	
18.01	MTA4	0	0	0	0	0	0	
1.43	MTA5	0	0	5	0	0	0	
18.89	MTA9-B	0	0	0	0	0	0	

Fig. 2E

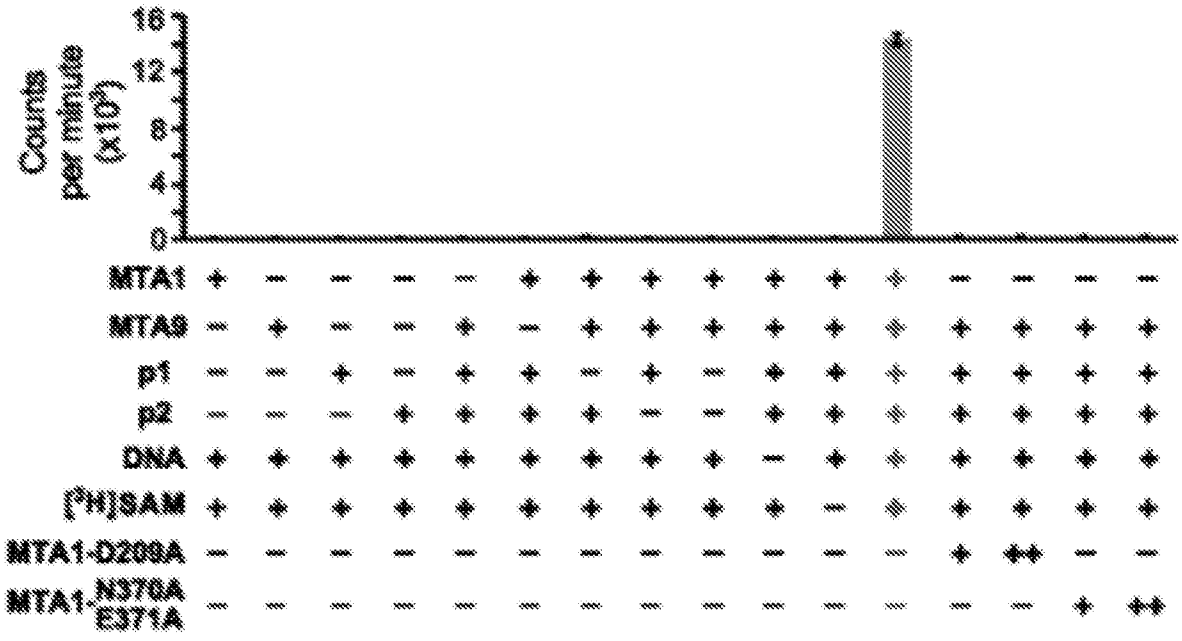


Fig. 2F

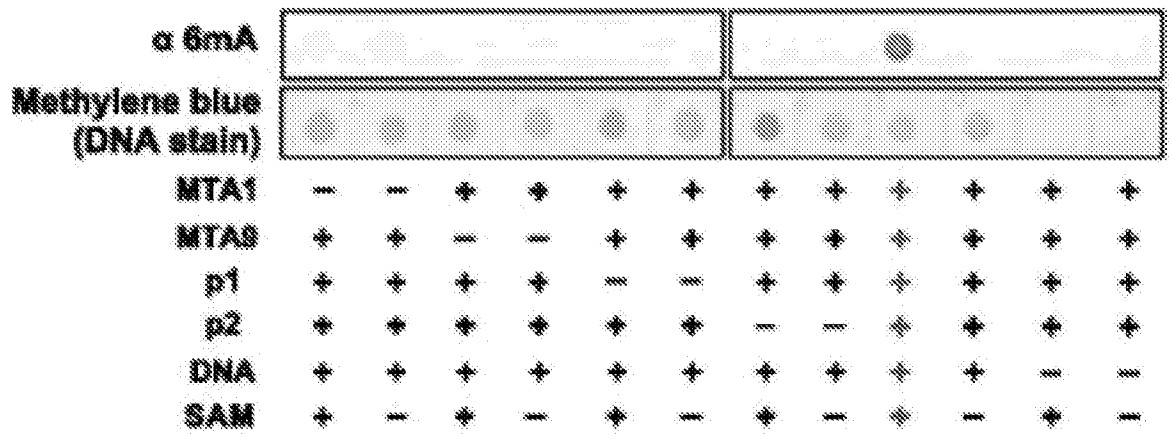

α 6mA
Methylene blue (DNA stain)
MTA1
MTA9
p1
p2
DNA
SAM

Fig. 2G

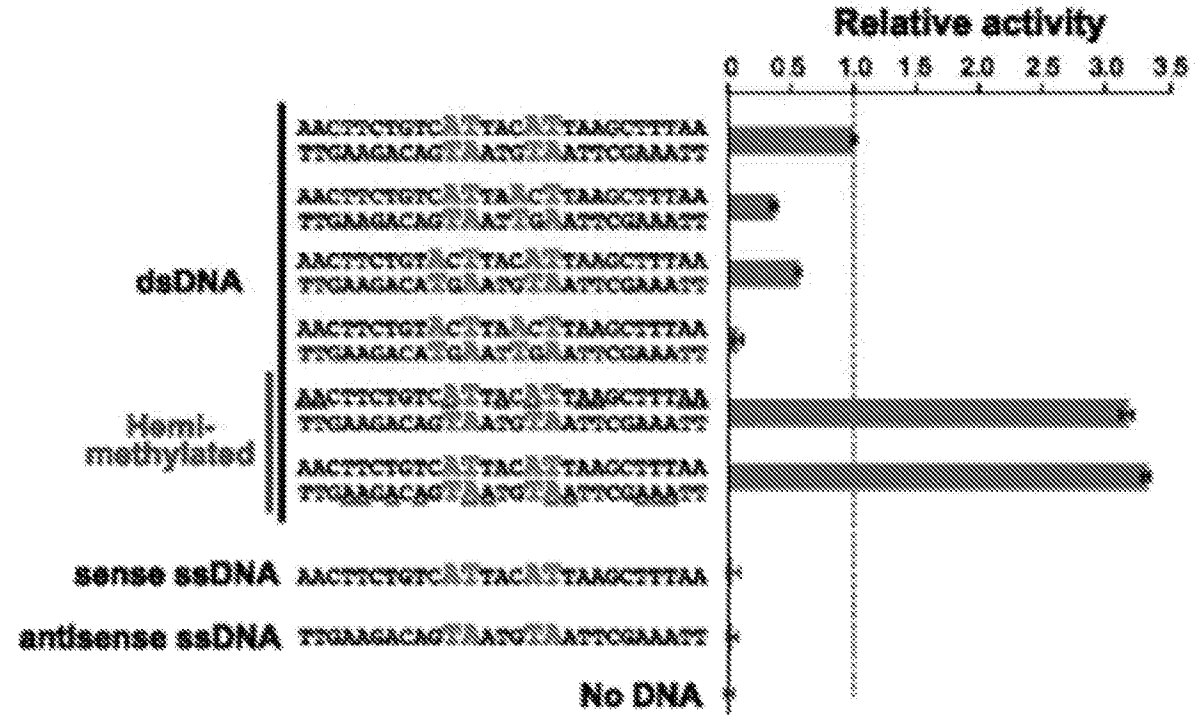

Relative activity
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
dsDNA
Hemi-methylated
sense ssDNA
antisense ssDNA
No DNA

Fig. 3A

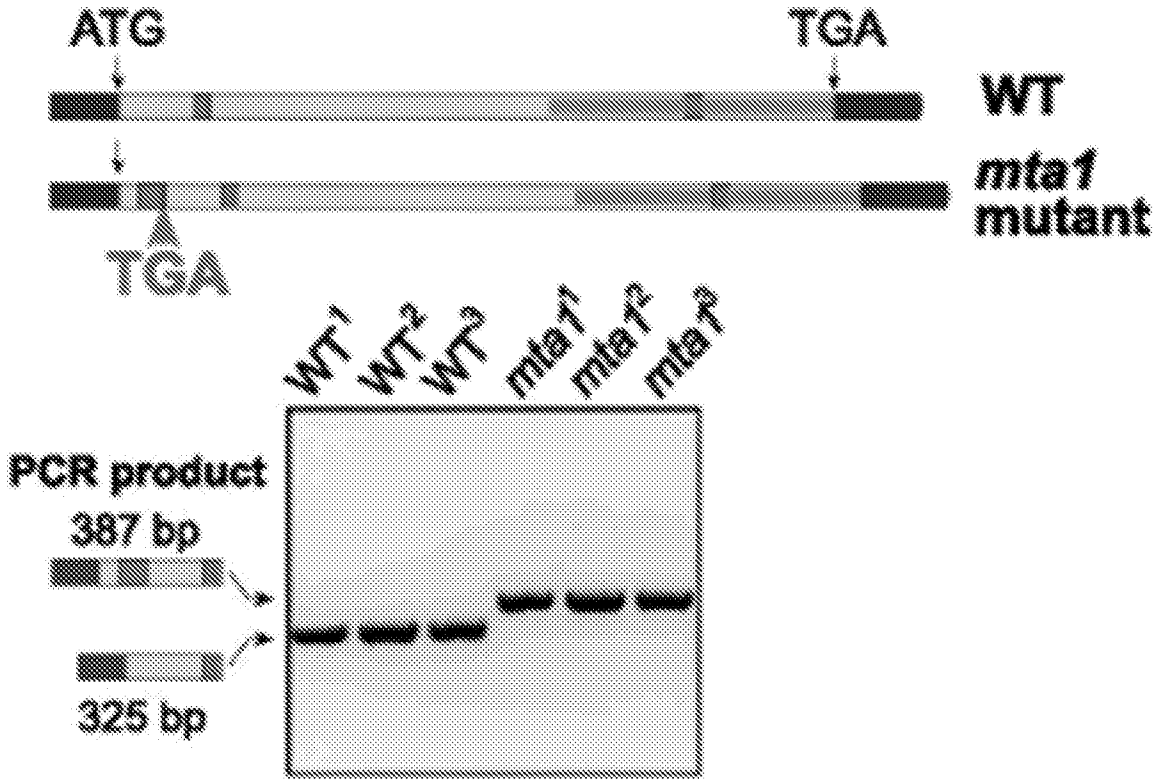

ATG
TGA
WT
mta1
mutant
TGA
WT1 WT2 WT3 mta1-1 mta1-2 mta1-3
PCR product
387 bp
325 bp

Fig. 3B

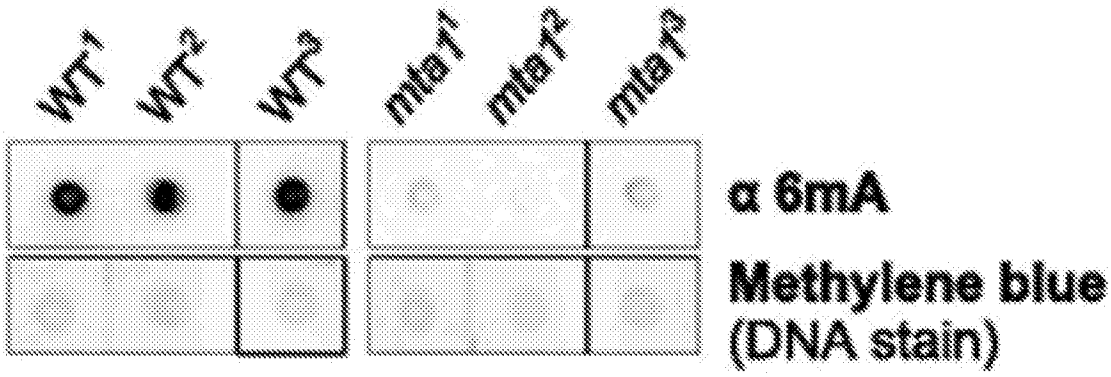

WT1 WT2 WT3 mta1-1 mta1-2 mta1-3
α 6mA
Methylene blue
(DNA stain)

Fig. 3C

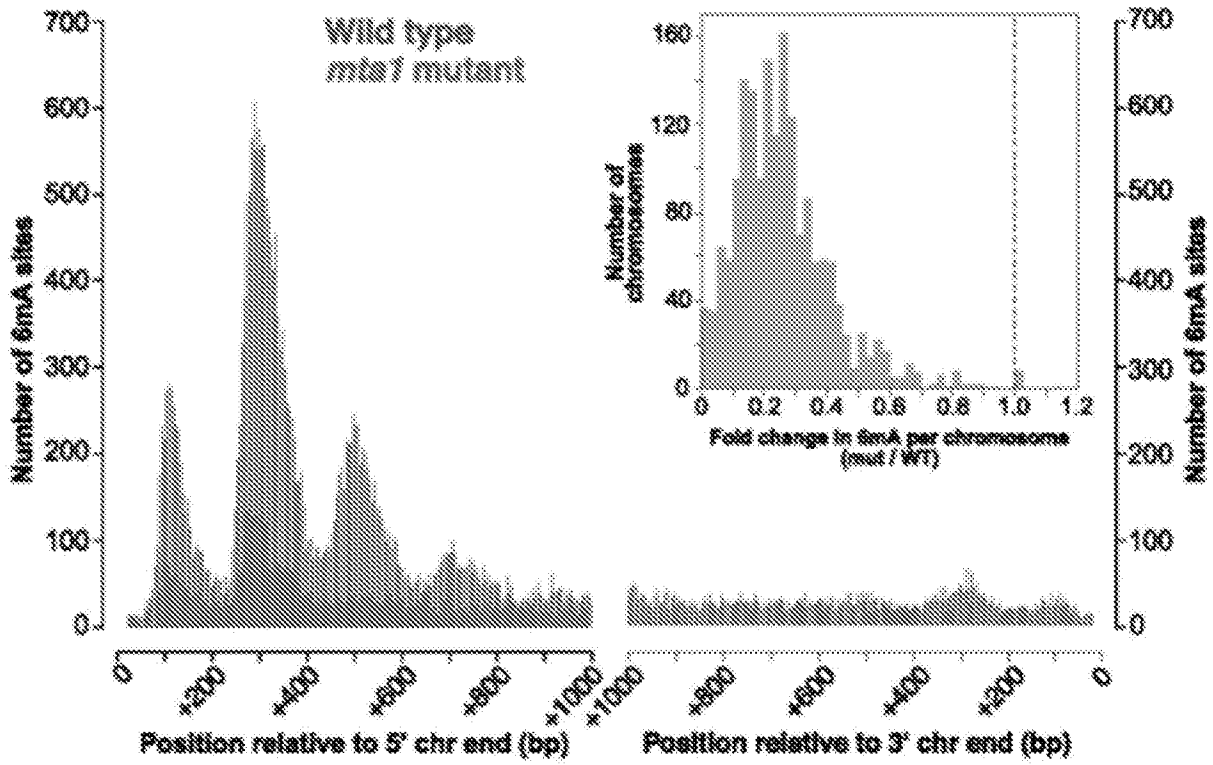
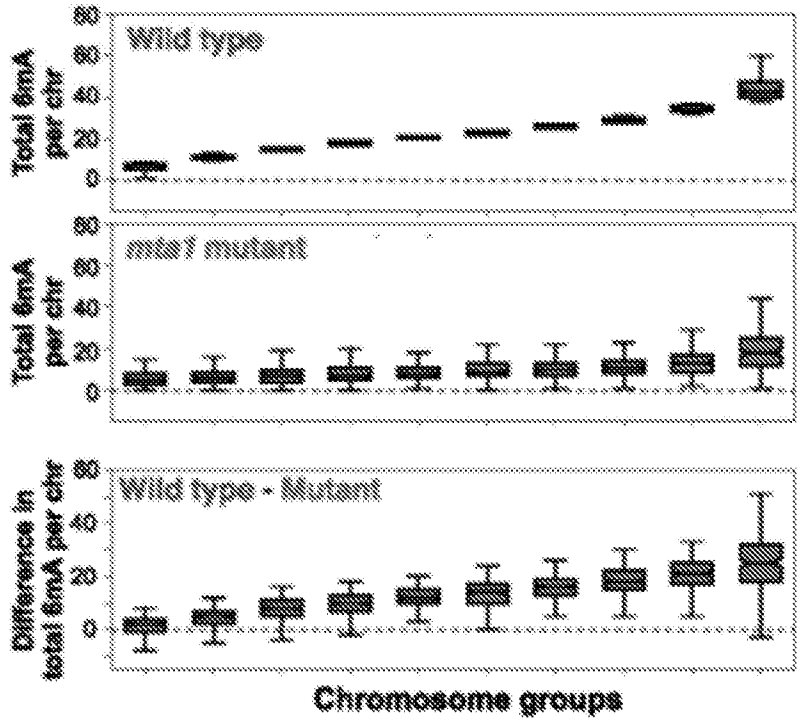

Wild type
mta1 mutant
Number of 6mA sites
Position relative to 5' chr end (bp)
Position relative to 3' chr end (bp)
Number of chromosomes
Fold change in 6mA per chromosome (mut / WT)

Fig. 3D


Wild type
mta1 mutant
Wild type - Mutant
Total 6mA per chr
Difference in total 6mA per chr
Chromosome groups

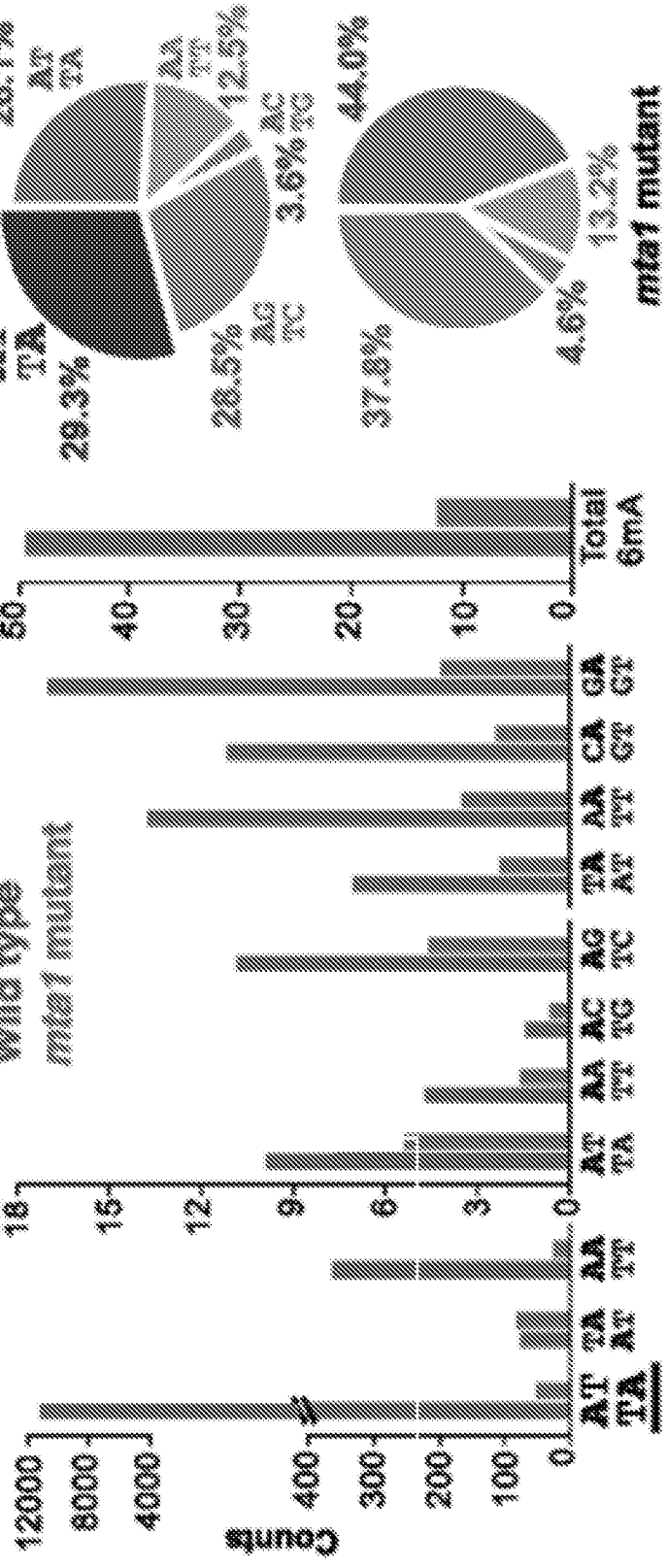

Wild type
AT
TA
29.3%
26.1%
AT
TA
AA
TT
12.5%
AG
TC
28.5%
AC
TG
3.6%
mta1 mutant
37.8%
44.0%
4.6%
13.2%
Wild type
mta1 mutant
Counts
x10³
Total 6mA

Fig. 3E

Fig. 4A

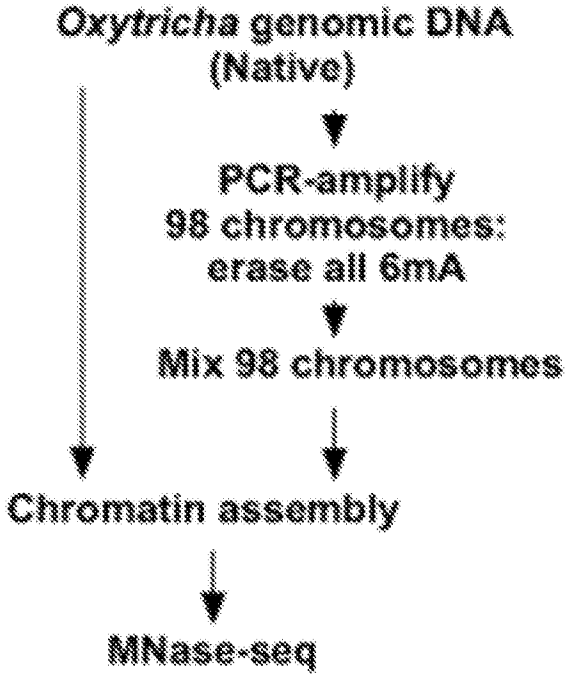

Oxytricha genomic DNA
(Native)
PCR-amplify
98 chromosomes:
erase all 6mA
Mix 98 chromosomes
Chromatin assembly
MNase-seq

Fig. 4B

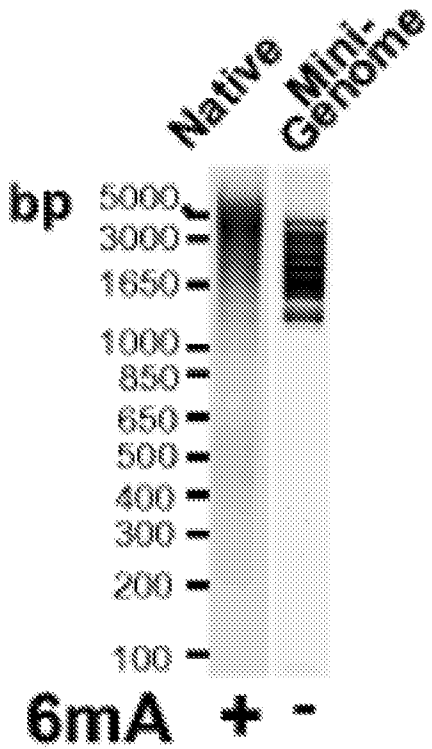

Native
Mini-Genome
bp
5000
3000
1650
1000
850
650
500
400
300
200
100
6mA + -

Fig. 4C

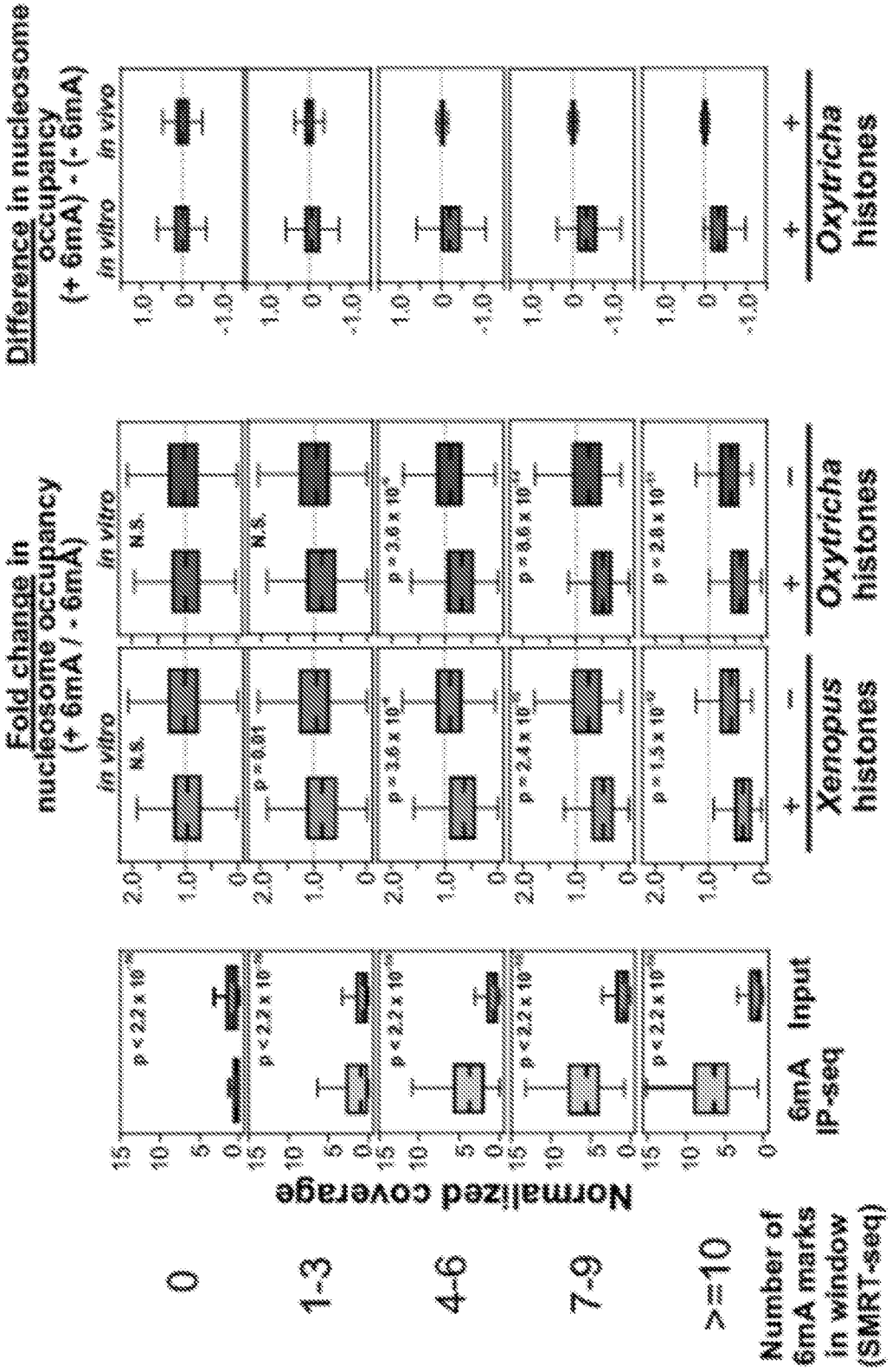

Number of 6mA marks in window (SMRT-seq)
0
1-3
4-6
7-9
>=10
Normalized coverage
6mA IP-seq
Input
Fold change in nucleosome occupancy (+ 6mA / - 6mA)
in vitro
in vivo
N.S.
p = 3.61
p = 3.6 x 10^-4
p = 2.4 x 10^-10
p = 1.5 x 10^-10
p = 8.6 x 10^-11
p = 2.8 x 10^-11
Xenopus histones
Oxytricha histones
Difference in nucleosome occupancy (+ 6mA) - (- 6mA)
in vitro
in vivo
Oxytricha histones

Fig. 4D

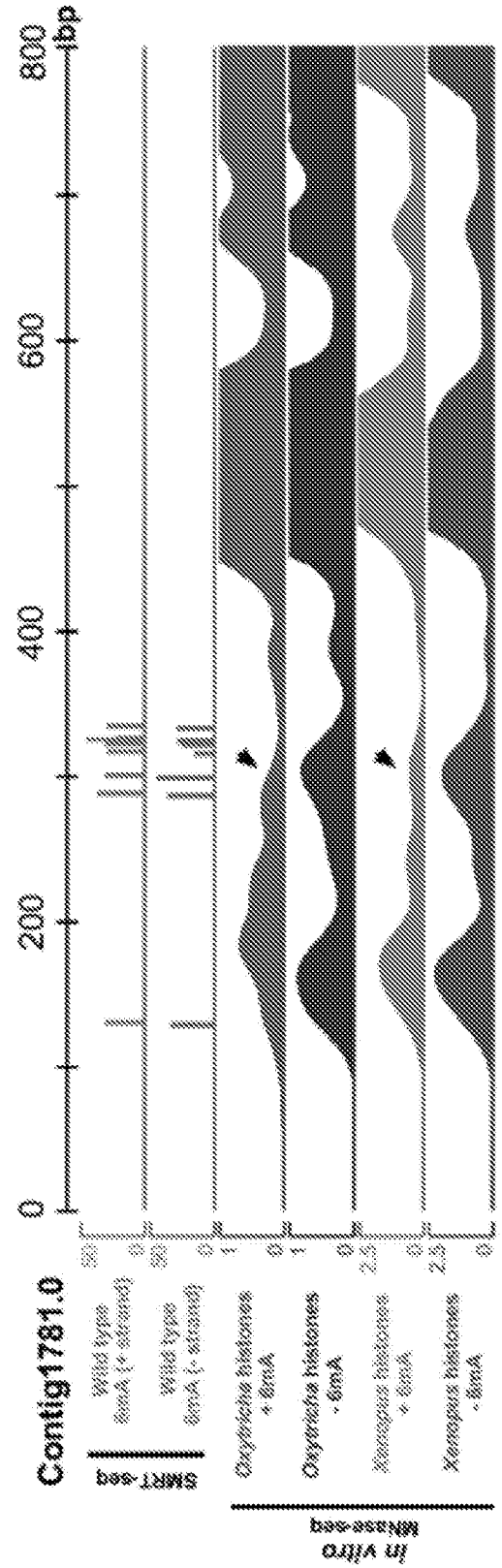

Contig1781.0
0
200
400
600
800
bp
SMRT-seq
Wild type 6mA (+ strand)
Wild type 6mA (- strand)
In vitro MNase-seq
Oxytricha histones + 6mA
Oxytricha histones - 6mA
Xenopus histones + 6mA
Xenopus histones - 6mA

Fig. 4E

0
200
400
600
800
bp
SMRT-seq
Wild type 6mA (+ strand)
Wild type 6mA (- strand)
mta1 mutant 6mA (+ strand)
mta1 mutant 6mA (- strand)
In vivo MNase-seq
Wild type
mta1 mutant

Fig. 5A

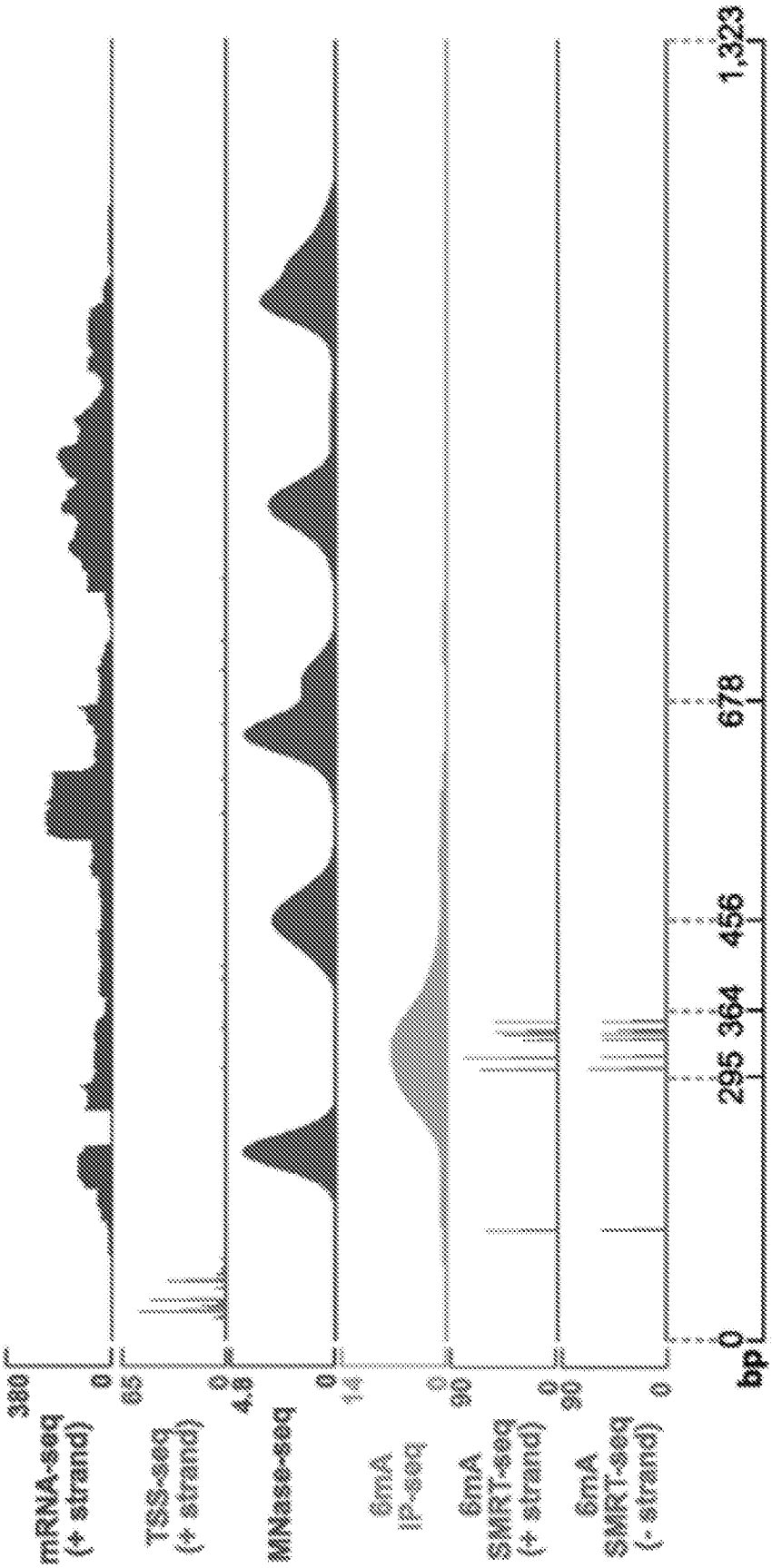
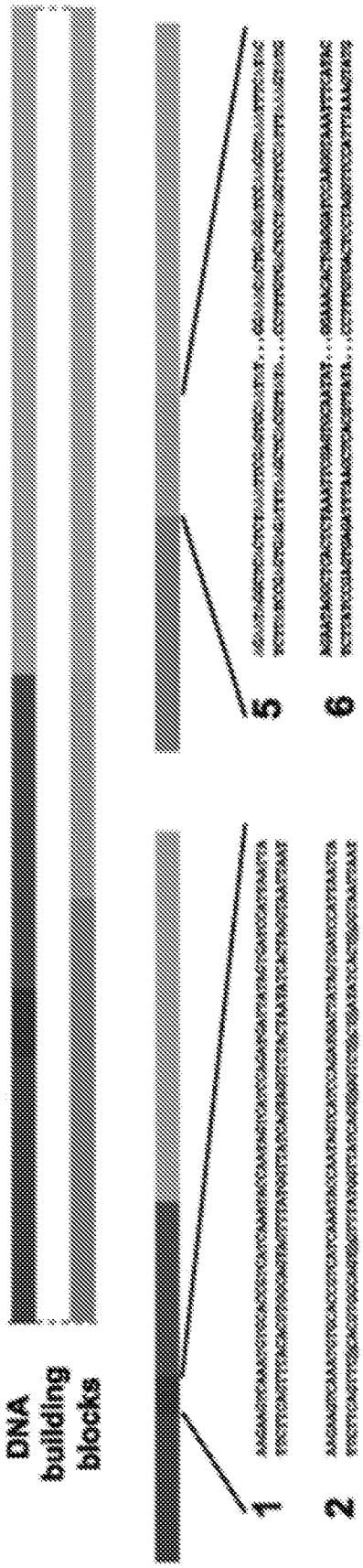

Contig1781.0 (1,323bp)
mRNA-seq (+ strand)
TSS-seq (+ strand)
MNase-seq
6mA IP-seq
6mA SMRT-seq (+ strand)
6mA SMRT-seq (- strand)
300
0
65
0
4.8
0
14
0
90
0
80
0
bp
0
295
364
456
678
1,323

Fig. 5B


DNA building blocks
1
2
3
4
Precise 6mA sites
5
6
Enzymatic 6mA sites

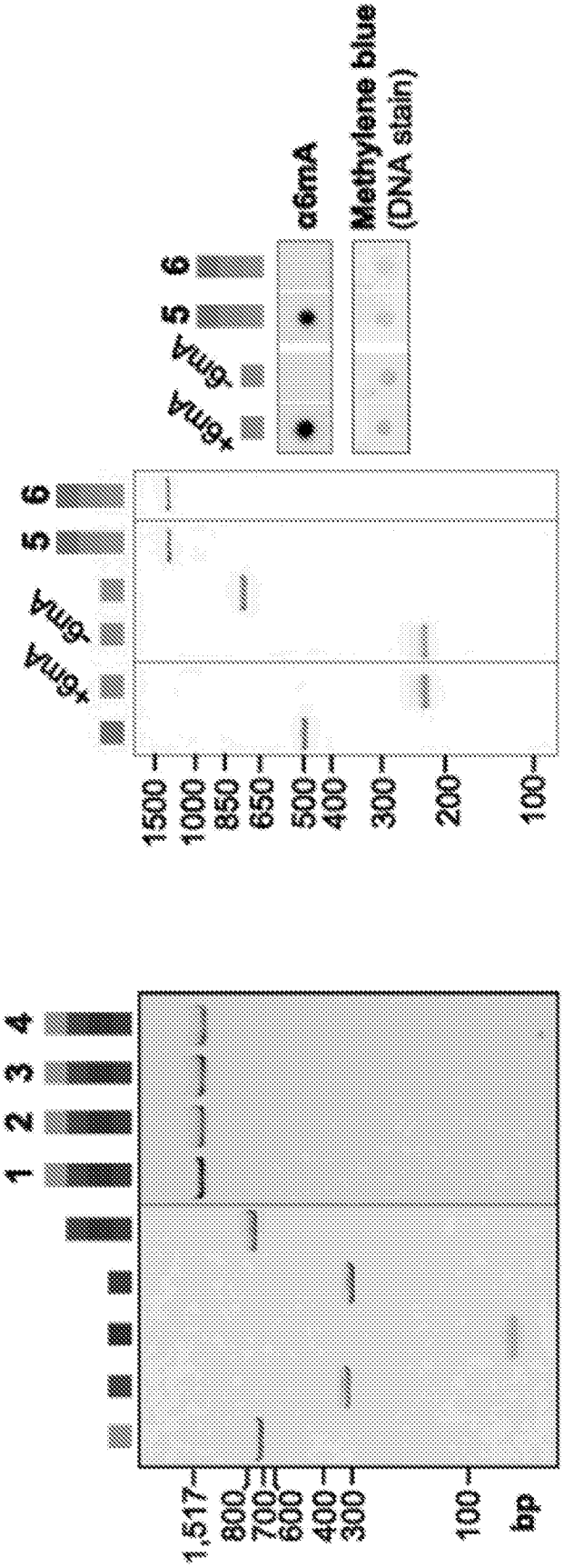

1 2 3 4
1,517
800
700
600
400
300
100
bp
5 6
+6mA
-6mA
1500
1000
850
650
500
400
300
200
100
5 6
+6mA
-6mA
α6mA
Methylene blue
(DNA stain)

Fig. 5C

Fig. 6A

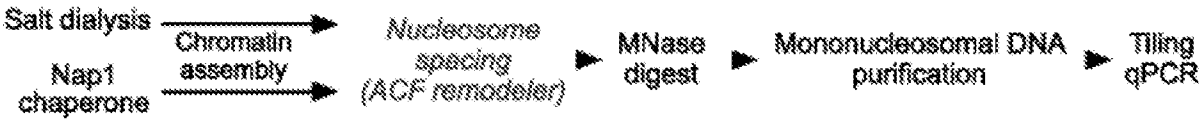

Salt dialysis
Nap1 chaperone
Chromatin assembly
Nucleosome spacing (ACF remodeler)
MNase digest
Mononucleosomal DNA purification
Tiling qPCR

Fig. 6B

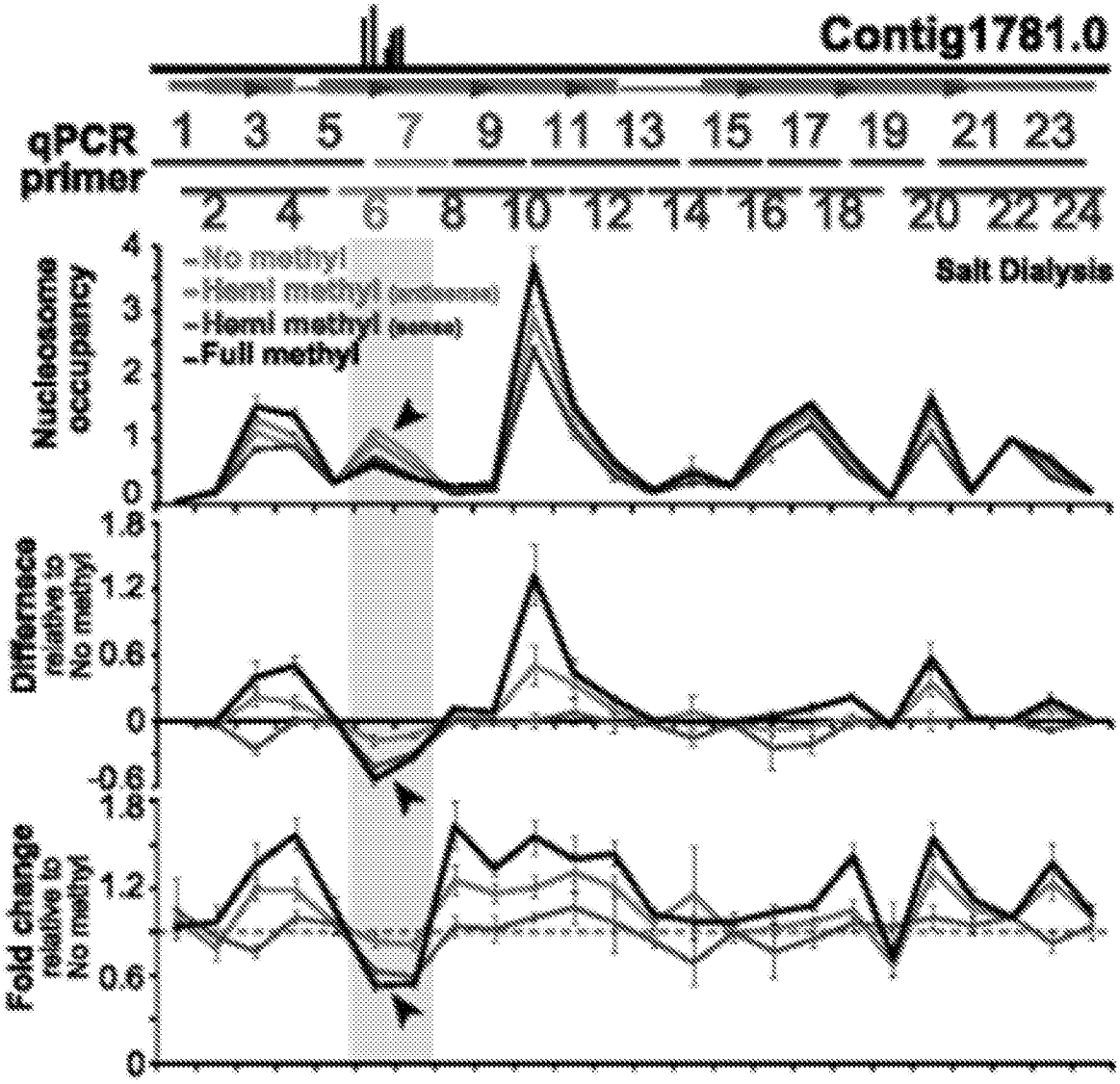

Contig1781.0
qPCR primer
1 3 5 7 9 11 13 15 17 19 21 23
2 4 6 8 10 12 14 16 18 20 22 24
Nucleosome occupancy
No methyl
Hemi methyl
Hemi methyl
Full methyl
Salt Dialysis
Difference relative to No methyl
Fold change relative to No methyl

Fig. 6C

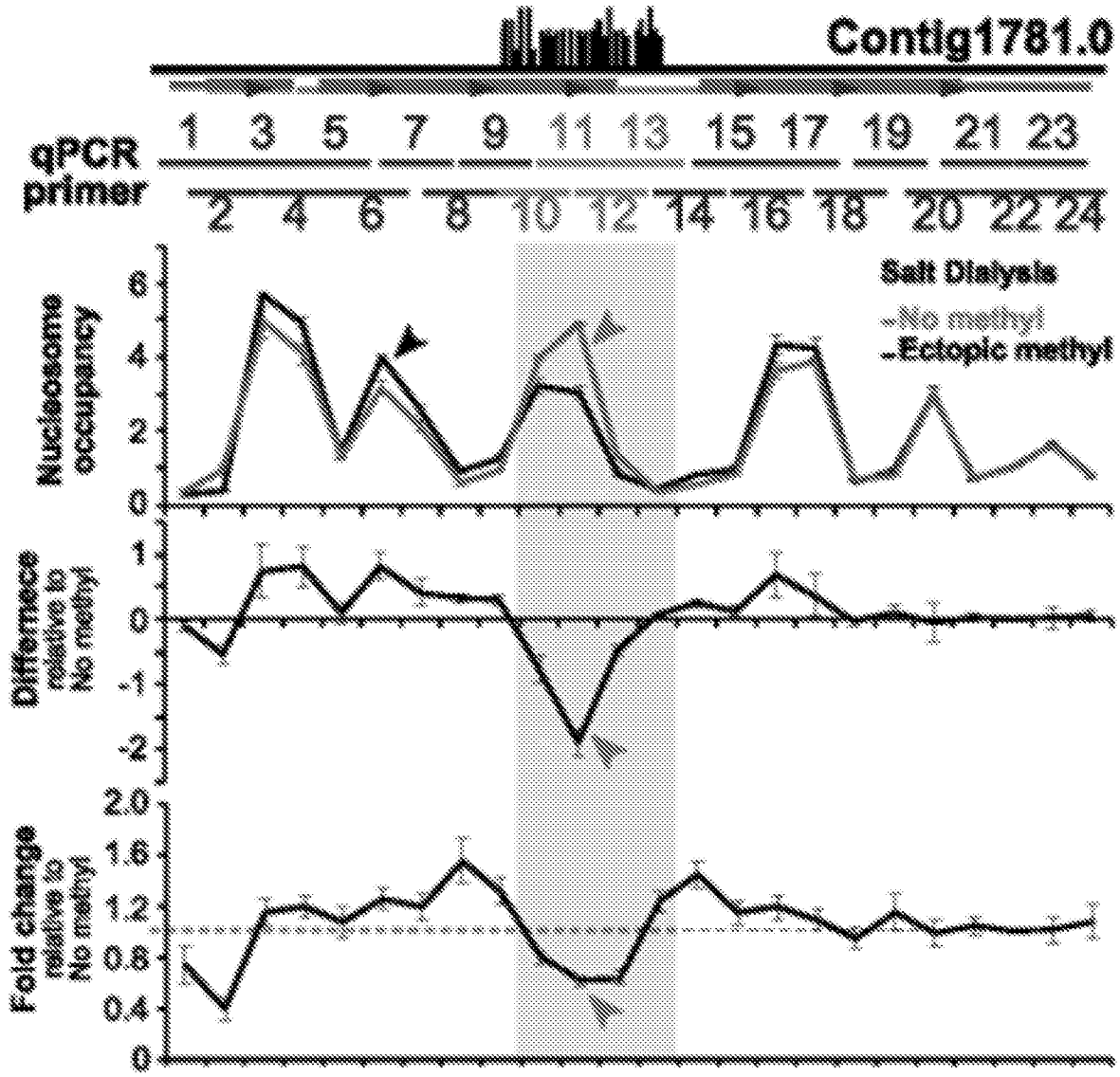

Contig1781.0
qPCR primer
1 3 5 7 9 11 13 15 17 19 21 23
2 4 6 8 10 12 14 16 18 20 22 24
Salt Dialysis
-No methyl
-Ectopic methyl
Nucleosome occupancy
0 2 4 6
Difference relative to No methyl
-2 -1 0 1
Fold change relative to No methyl
0 0.4 0.8 1.2 1.6 2.0

Fig. 6D

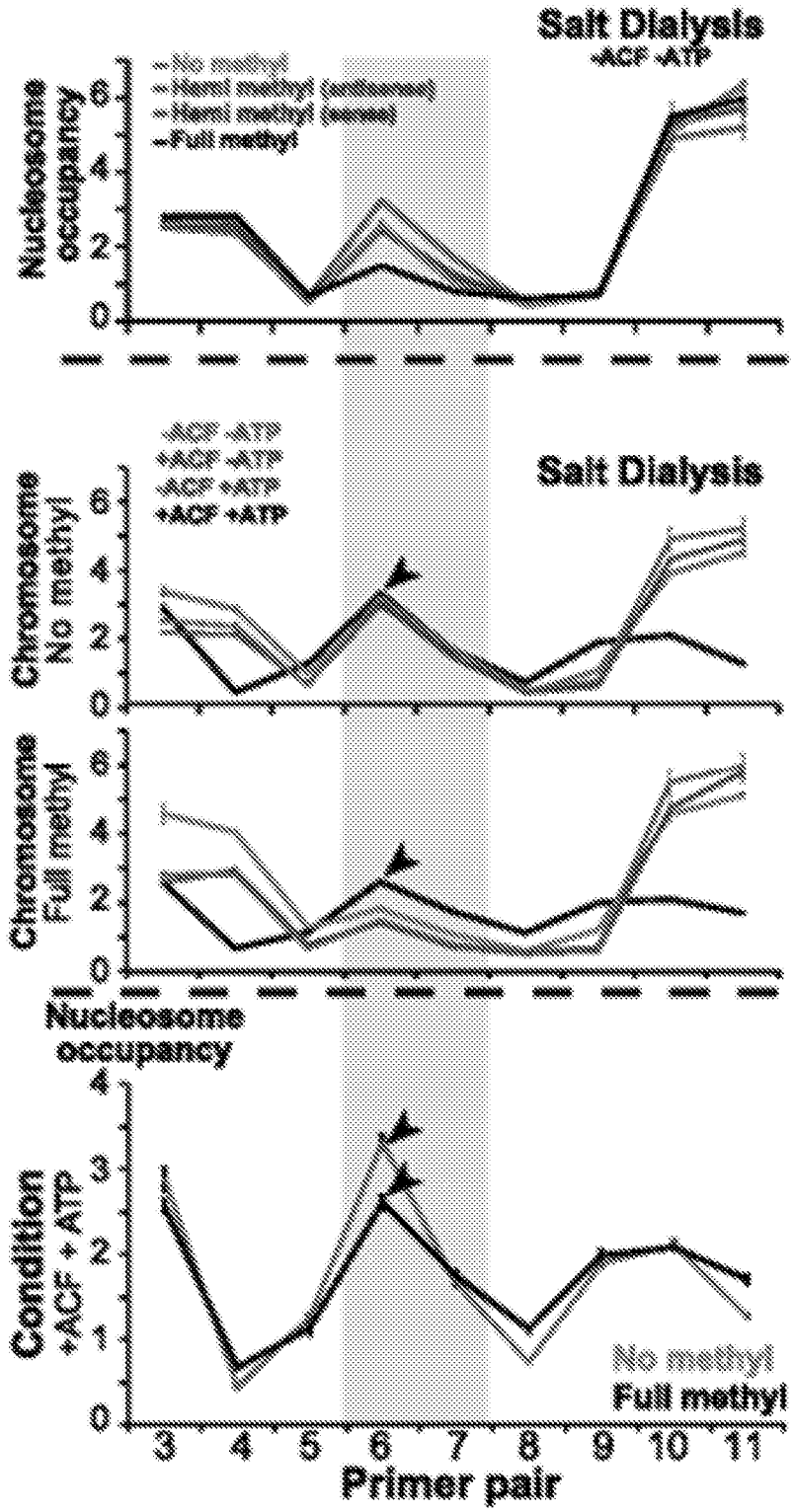

Salt Dialysis
+ACF +ATP
Nucleosome occupancy
No methyl
Hemi methyl (cytosine)
Hemi methyl (amine)
Full methyl
Salt Dialysis
-ACF -ATP
+ACF -ATP
-ACF +ATP
+ACF +ATP
Chromosome No methyl
Chromosome Full methyl
Nucleosome occupancy
Condition +ACF + ATP
No methyl
Full methyl
Primer pair

Fig. 6E

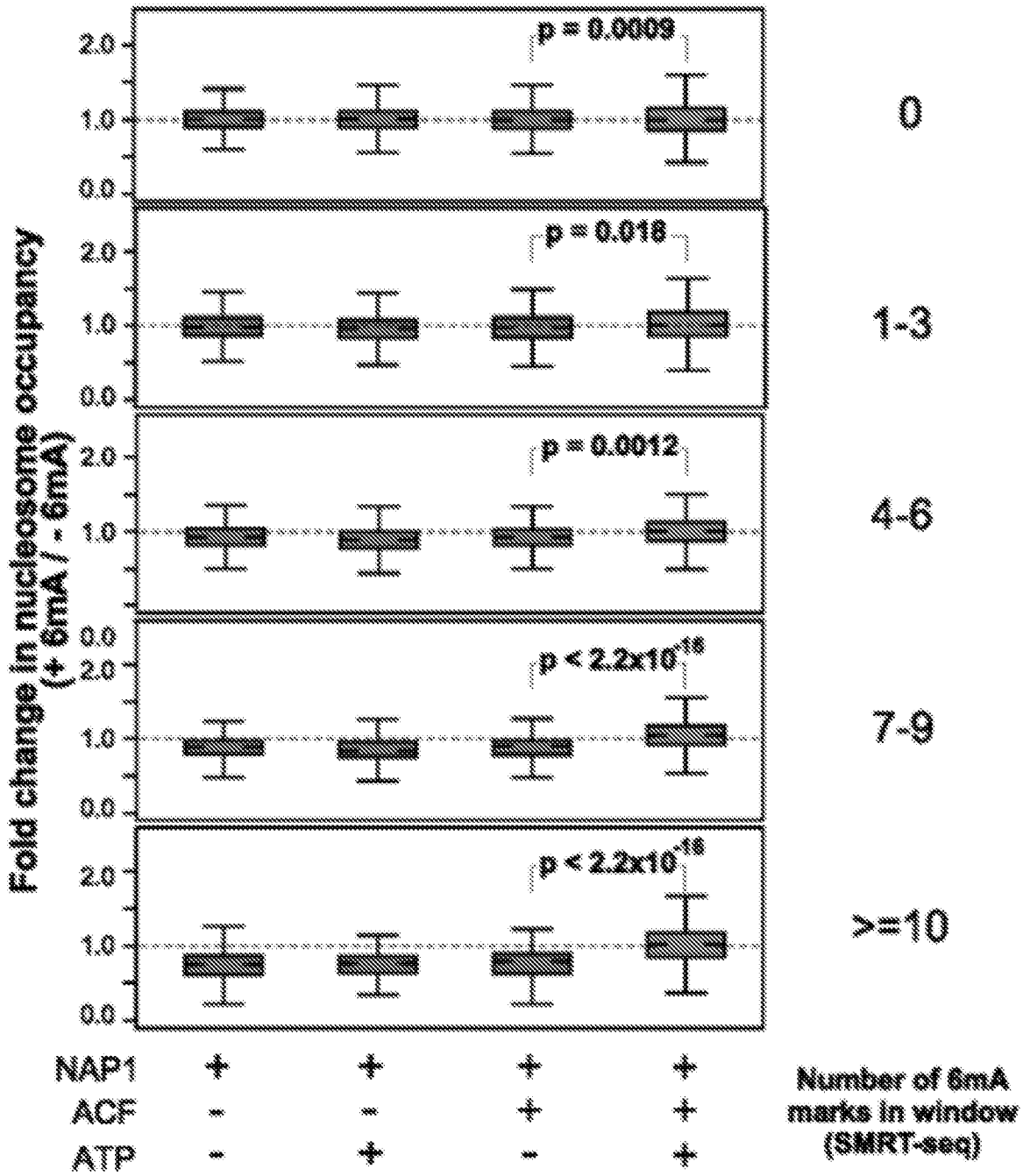

p = 0.0009
p = 0.018
p = 0.0012
p < 2.2x10^-16
p < 2.2x10^-16
0
1-3
4-6
7-9
>=10
Fold change in nucleosome occupancy (+ 6mA / - 6mA)
2.0
1.0
0.0
NAP1 + + + +
ACF - - + +
ATP - + - +
Number of 6mA marks in window (SMRT-seq)

Fig. 7A

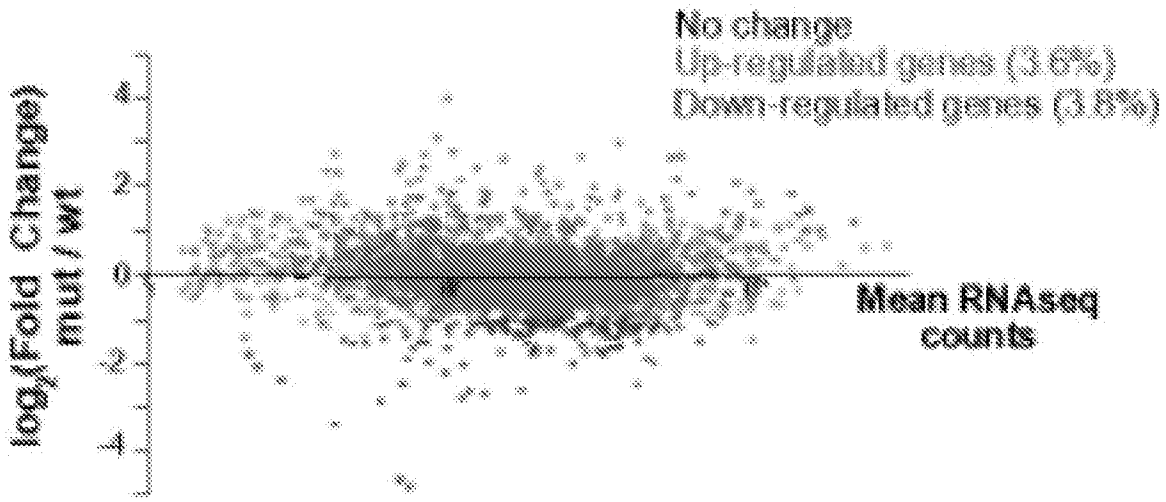

No change
Up-regulated genes (3.6%)
Down-regulated genes (3.6%)
log2Fold Change) mut / wt
Mean RNAseq counts

Fig. 7B

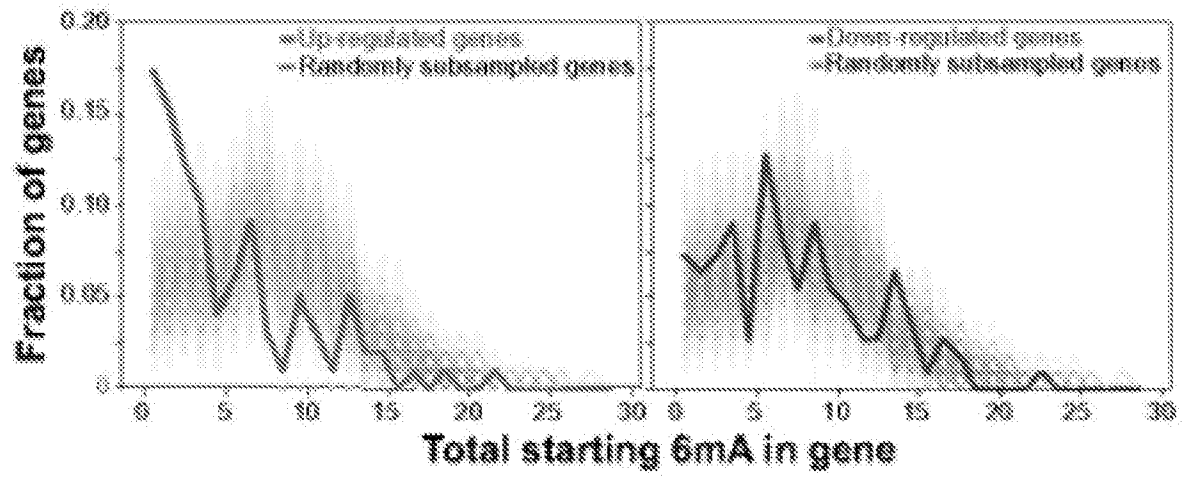

Up-regulated genes
Randomly subsampled genes
Down-regulated genes
Randomly subsampled genes
Fraction of genes
Total starting 6mA in gene

Fig. 7C

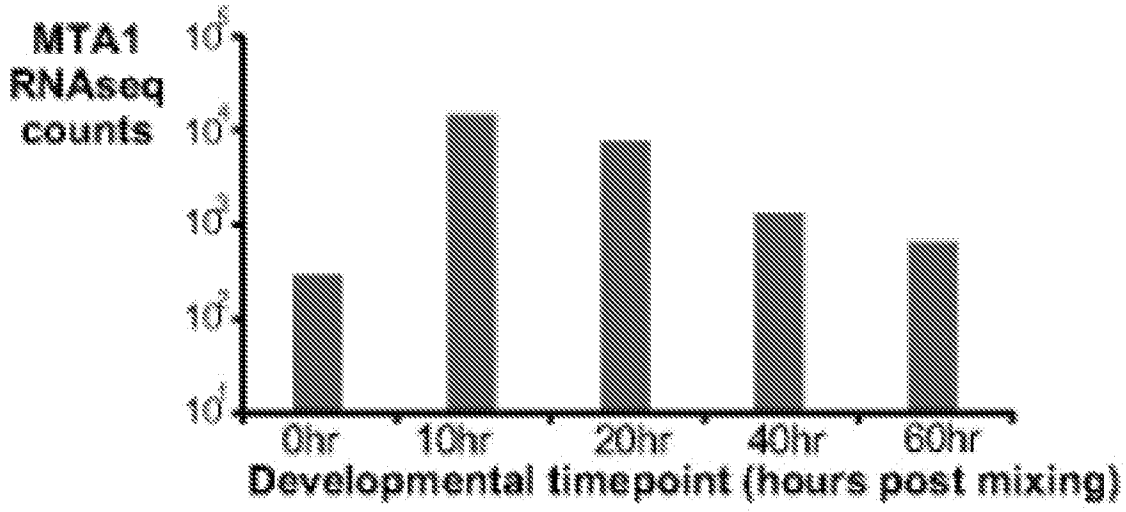

MTA1
RNAseq
counts
10^5
10^4
10^3
10^2
10^1
0hr
10hr
20hr
40hr
60hr
Developmental timepoint (hours post mixing)

Fig. 7D

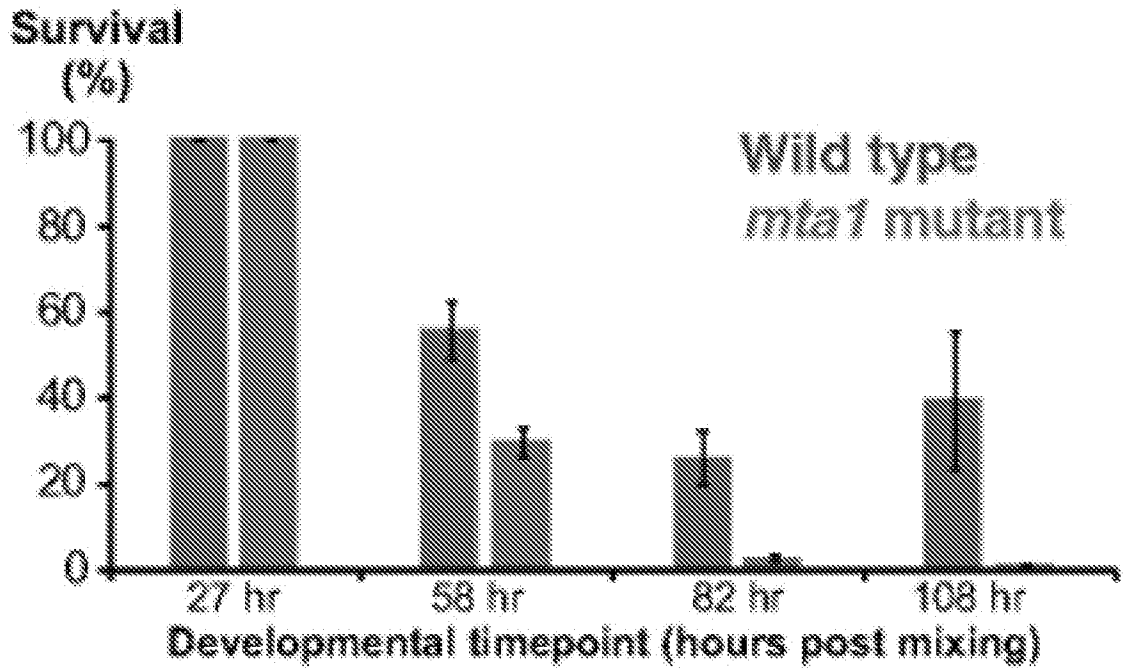

Survival
(%)
100
80
60
40
20
0
Wild type
mta1 mutant
27 hr
58 hr
82 hr
108 hr
Developmental timepoint (hours post mixing)

Fig. 7E

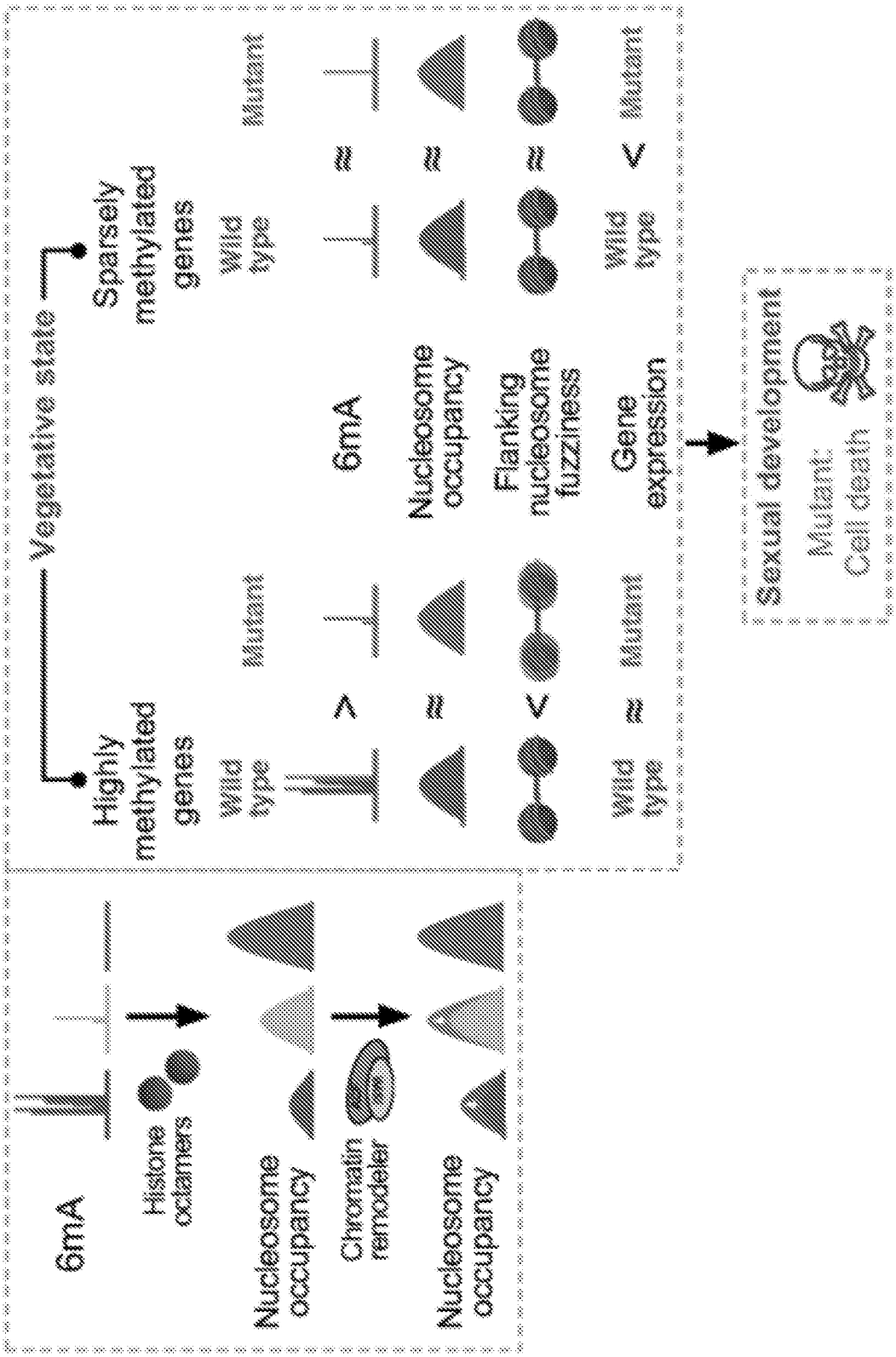

Impact of 6mA
in vitro
in vivo
6mA
Histone octamers
Nucleosome occupancy
Chromatin remodeler
Nucleosome occupancy
Vegetative state
Highly methylated genes
Sparsely methylated genes
Wild type
Mutant
Wild type
Mutant
6mA
Nucleosome occupancy
Flanking nucleosome fuzziness
Gene expression
Wild type
Mutant
Wild type
Mutant
Sexual development
Mutant:
Cell death

Fig. 7F

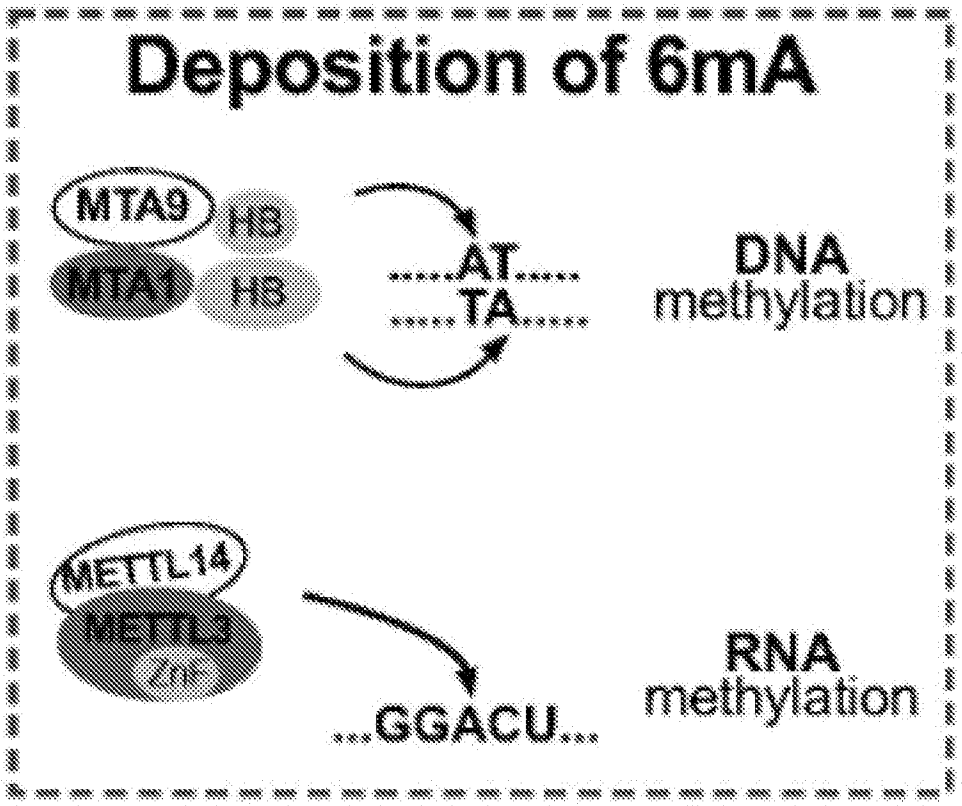

Deposition of 6mA
MTA9
HB
MTA1
HB
.....AT.....
.....TA.....
DNA
methylation
METTL14
METTL3
Zn
...GGACU...
RNA
methylation

Fig. 8A

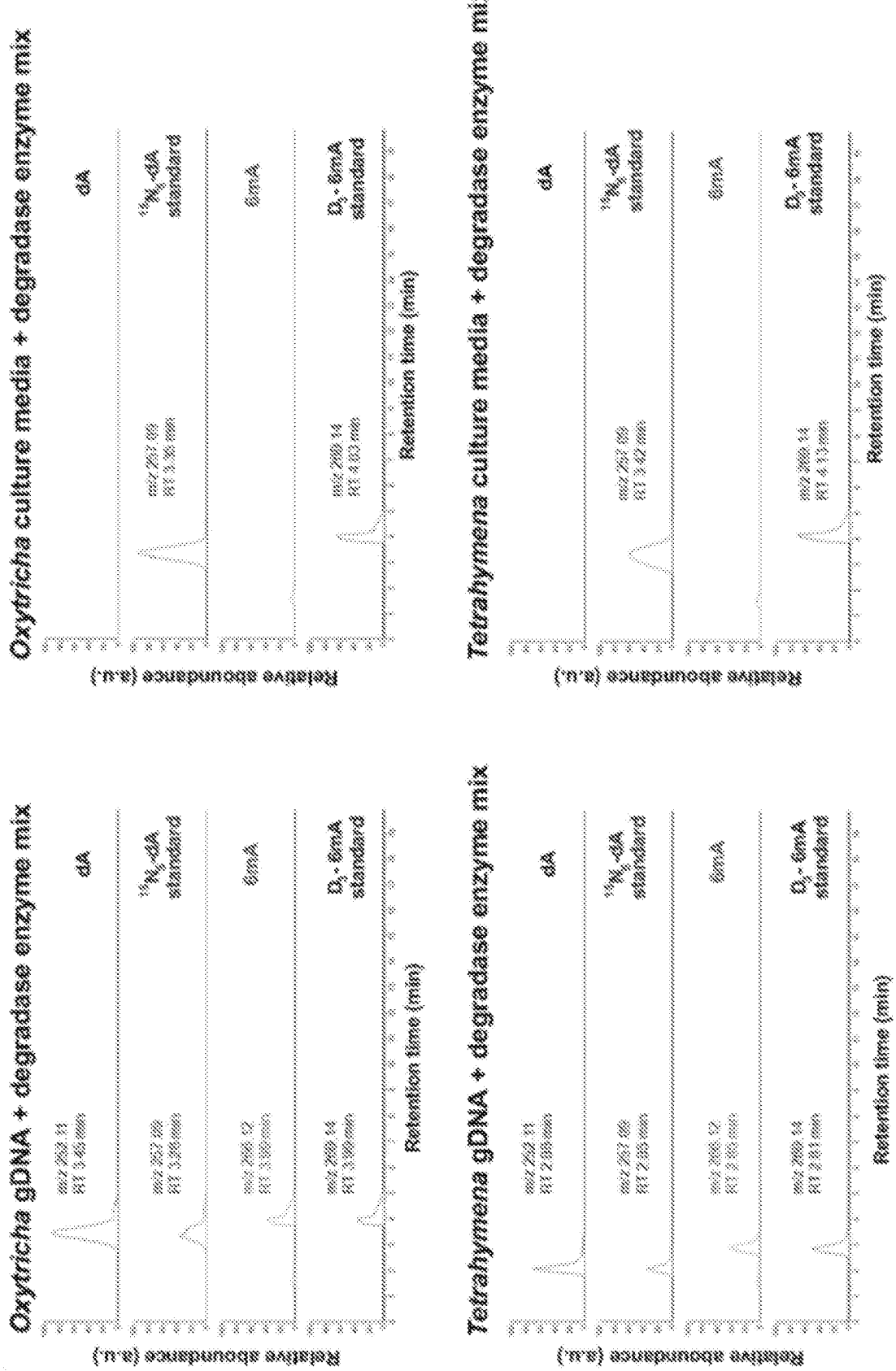

Oxytricha gDNA + degradase enzyme mix
dA
m/z 252.11
RT 3.40 min
$^{15}N_5$-dA standard
m/z 267.09
RT 3.36 min
6mA
m/z 266.12
RT 3.38 min
D_3-6mA standard
m/z 269.14
RT 3.39 min
Relative abundance (a.u.)
Retention time (min)
Oxytricha culture media + degradase enzyme mix
dA
m/z 252.11
RT 3.40 min
$^{15}N_5$-dA standard
m/z 267.09
RT 3.36 min
6mA
m/z 266.12
RT 3.38 min
D_3-6mA standard
m/z 269.14
RT 3.39 min
Relative abundance (a.u.)
Retention time (min)
Tetrahymena gDNA + degradase enzyme mix
dA
m/z 252.11
RT 2.88 min
$^{15}N_5$-dA standard
m/z 267.09
RT 2.85 min
6mA
m/z 266.12
RT 2.86 min
D_3-6mA standard
m/z 269.14
RT 2.87 min
Relative abundance (a.u.)
Retention time (min)
Tetrahymena culture media + degradase enzyme mix
dA
m/z 252.11
RT 3.40 min
$^{15}N_5$-dA standard
m/z 267.09
RT 3.37 min
6mA
m/z 266.12
RT 3.38 min
D_3-6mA standard
m/z 269.14
RT 3.39 min
Relative abundance (a.u.)
Retention time (min)

Fig. 8A (Continued)

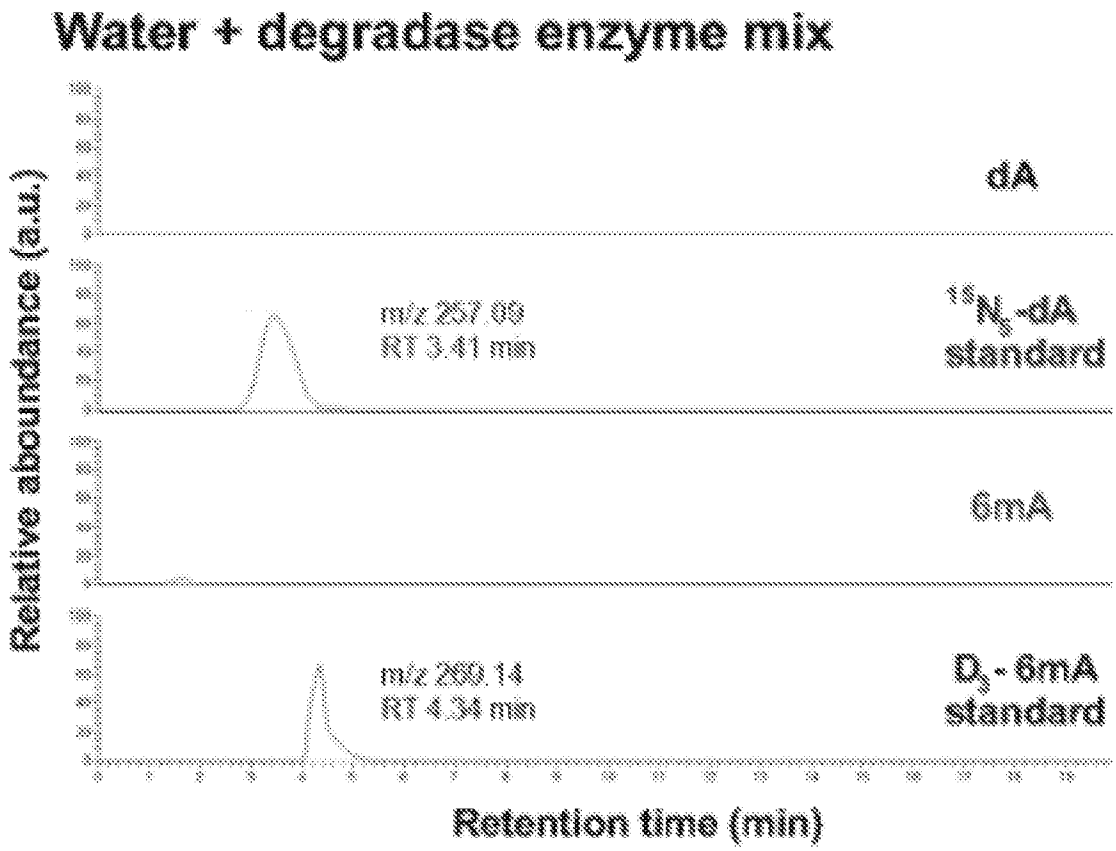


Fig. 8B

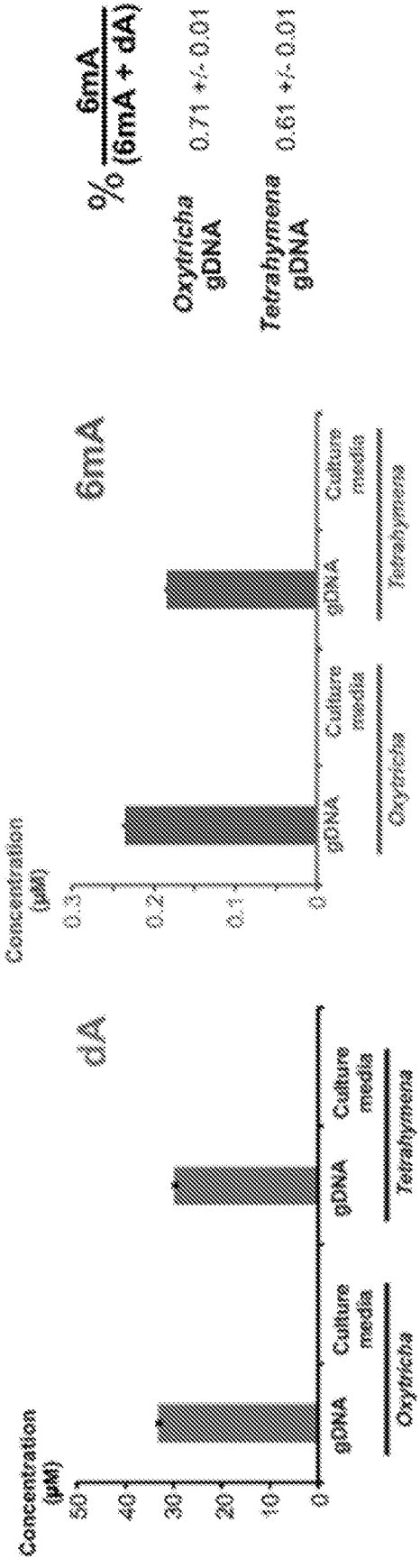

dA
6mA
Concentration (μM)
50
40
30
20
10
0
0.3
0.2
0.1
gDNA
Culture media
Oxytricha
Tetrahymena
% 6mA / (6mA + dA)
Oxytricha gDNA 0.71 +/- 0.01
Tetrahymena gDNA 0.61 +/- 0.01

Fig. 9A

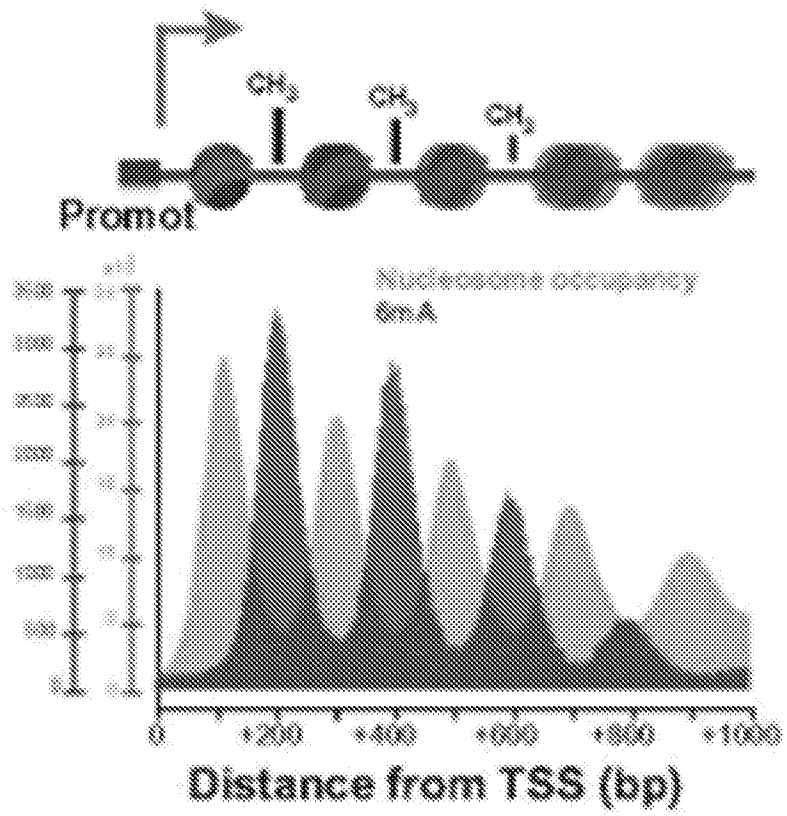

Promot
Nucleosome occupancy
6mA
0
+200
+400
+600
+800
+1000
Distance from TSS (bp)

Fig. 9B

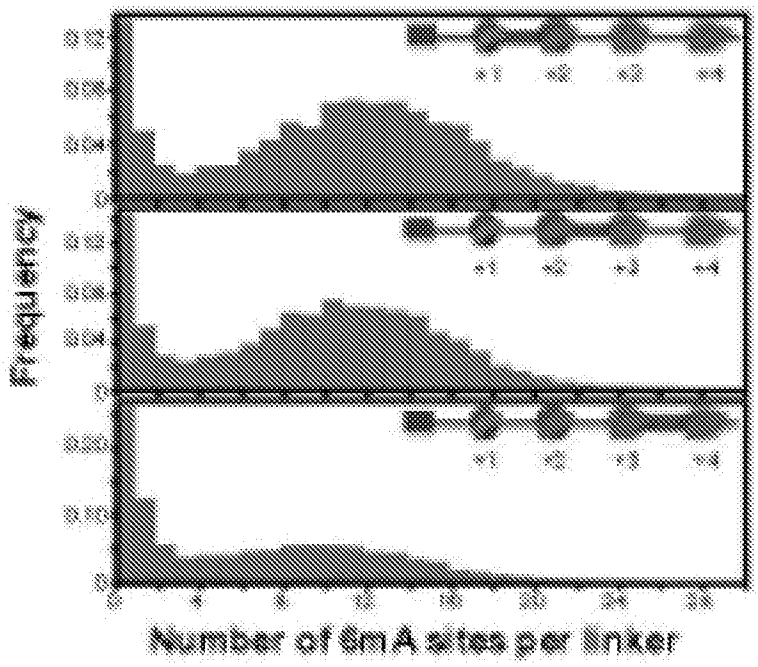

Frequency
Number of 6mA sites per linker

Fig. 9C

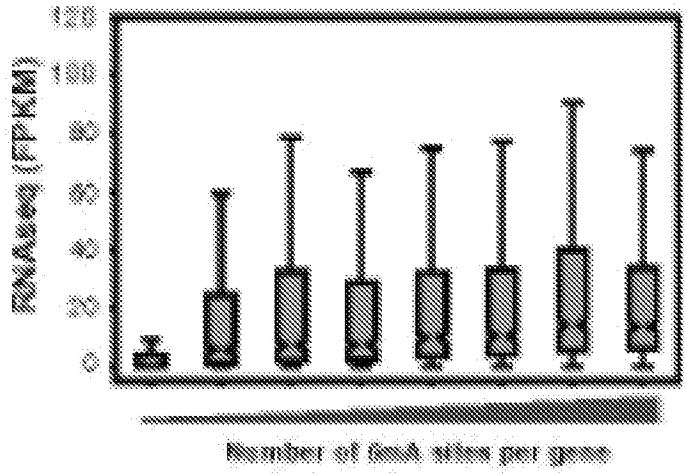


Fig. 9D

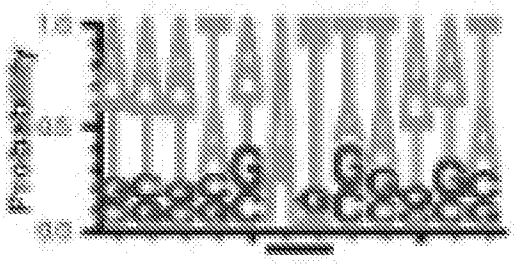


Fig. 9E

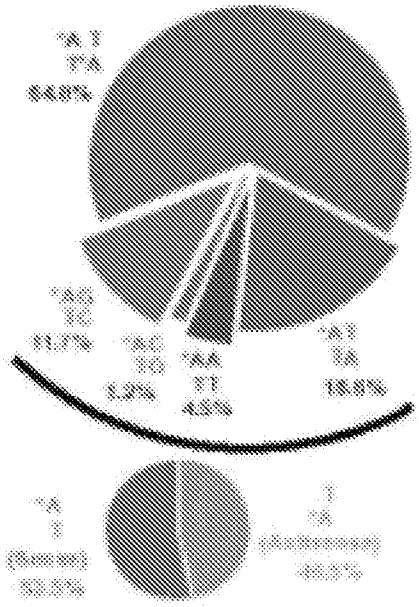


Fig. 9F

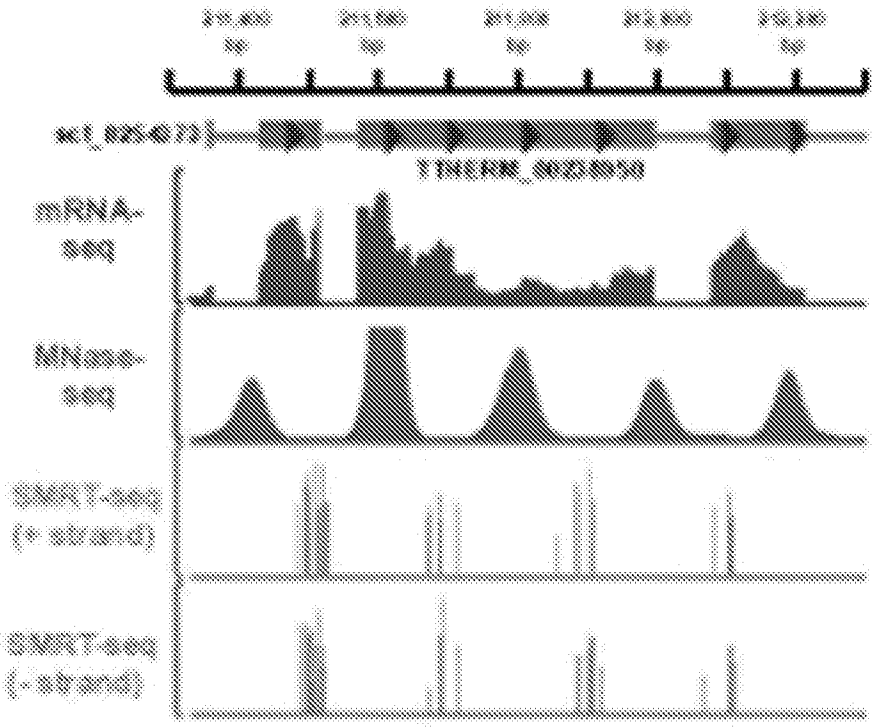


Fig. 9G

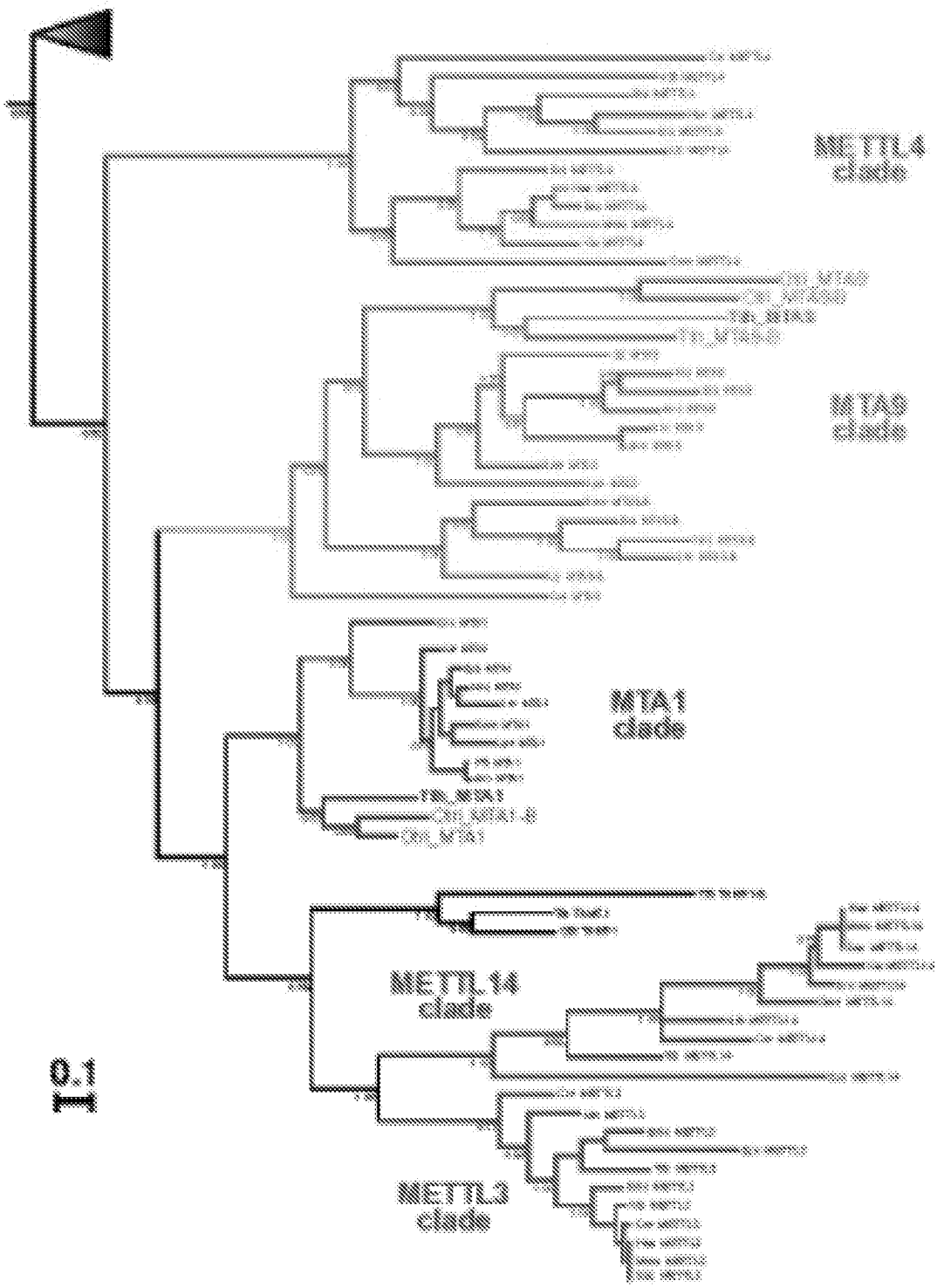

METTL4
clade
MTA3
clade
MTA1
clade
METTL14
clade
METTL3
clade
0.1

Fig. 9H

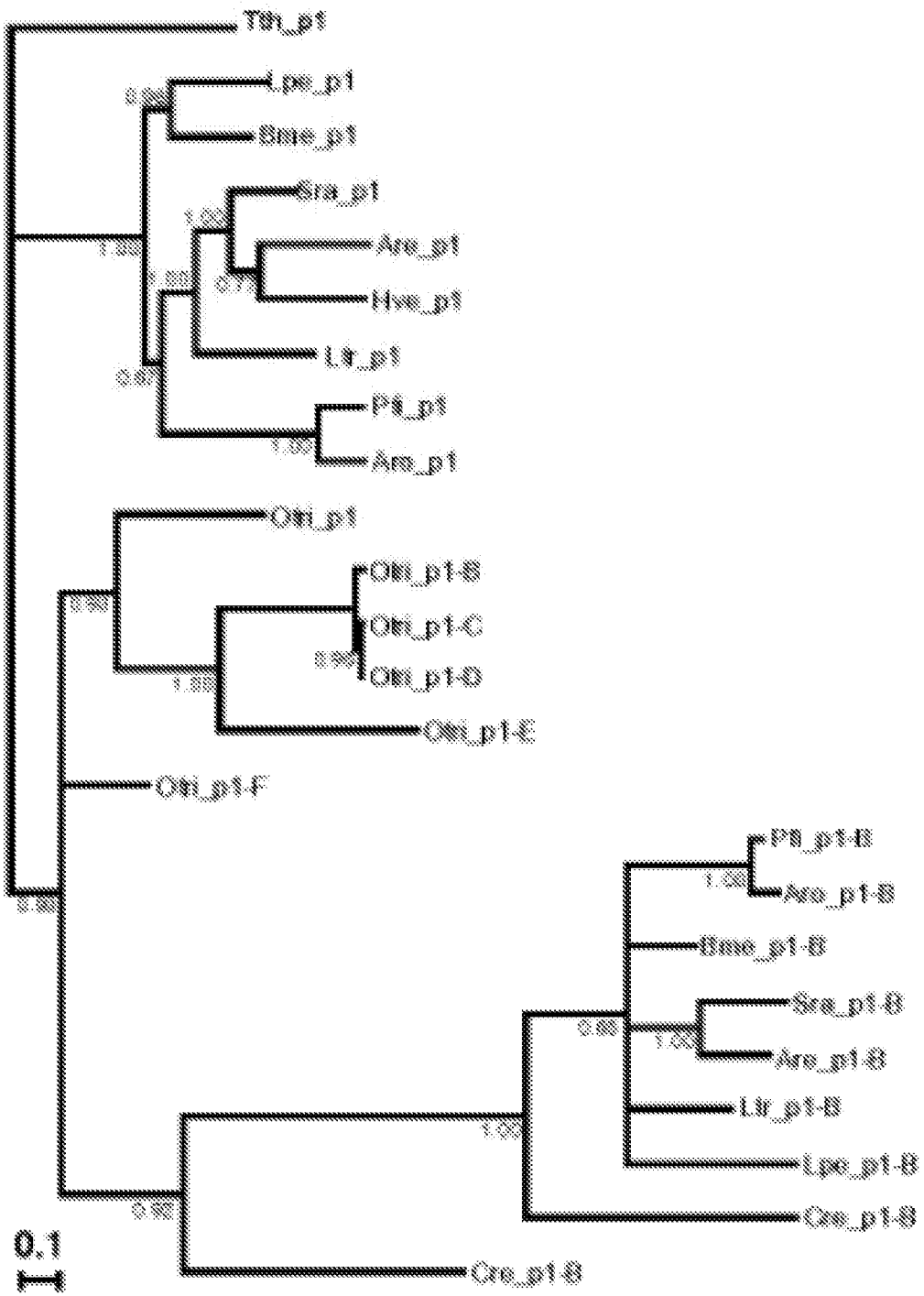

Tth_p1
Lpe_p1
Sme_p1
Sra_p1
Are_p1
Hse_p1
Ltr_p1
Pfl_p1
Ara_p1
Oti_p1
Oti_p1-B
Oti_p1-C
Oti_p1-D
Oti_p1-E
Oti_p1-F
Pfl_p1-B
Are_p1-B
Sme_p1-B
Sra_p1-B
Are_p1-B
Ltr_p1-B
Lpe_p1-B
Ore_p1-B
Ore_p1-B
0.1

Fig. 9I

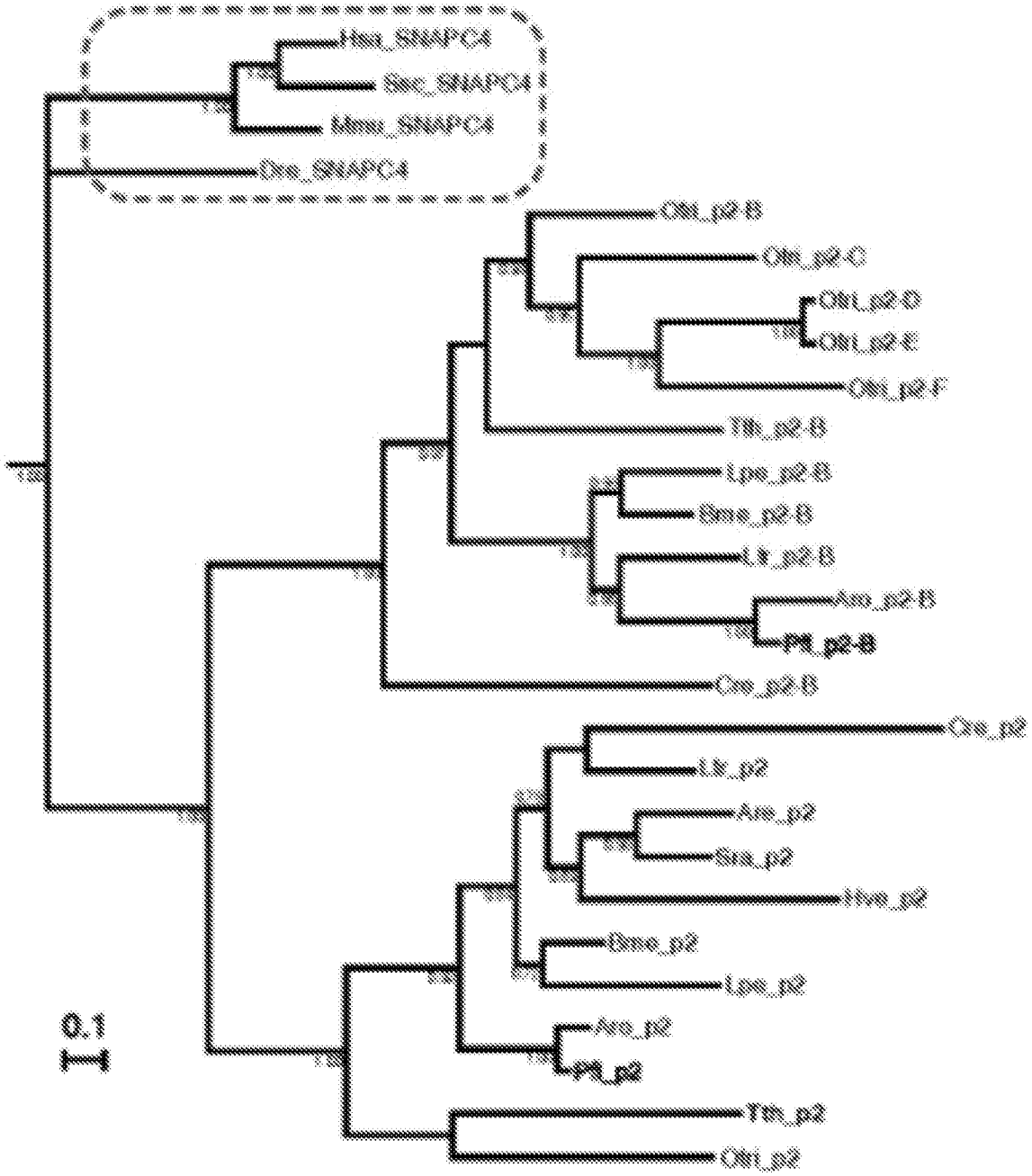

Hsa_SNAPC4
Ssc_SNAPC4
Mmu_SNAPC4
Dre_SNAPC4
Ola_p2-B
Ola_p2-C
Ola_p2-D
Ola_p2-E
Ola_p2-F
Tni_p2-B
Lpe_p2-B
Gmo_p2-B
Lc_p2-B
Aca_p2-B
Pm_p2-B
Cse_p2-B
Dre_p2
Lc_p2
Aca_p2
Sra_p2
Hse_p2
Gmo_p2
Lpe_p2
Aca_p2
Pm_p2
Tni_p2
Ola_p2
0.1

Fig. 9J

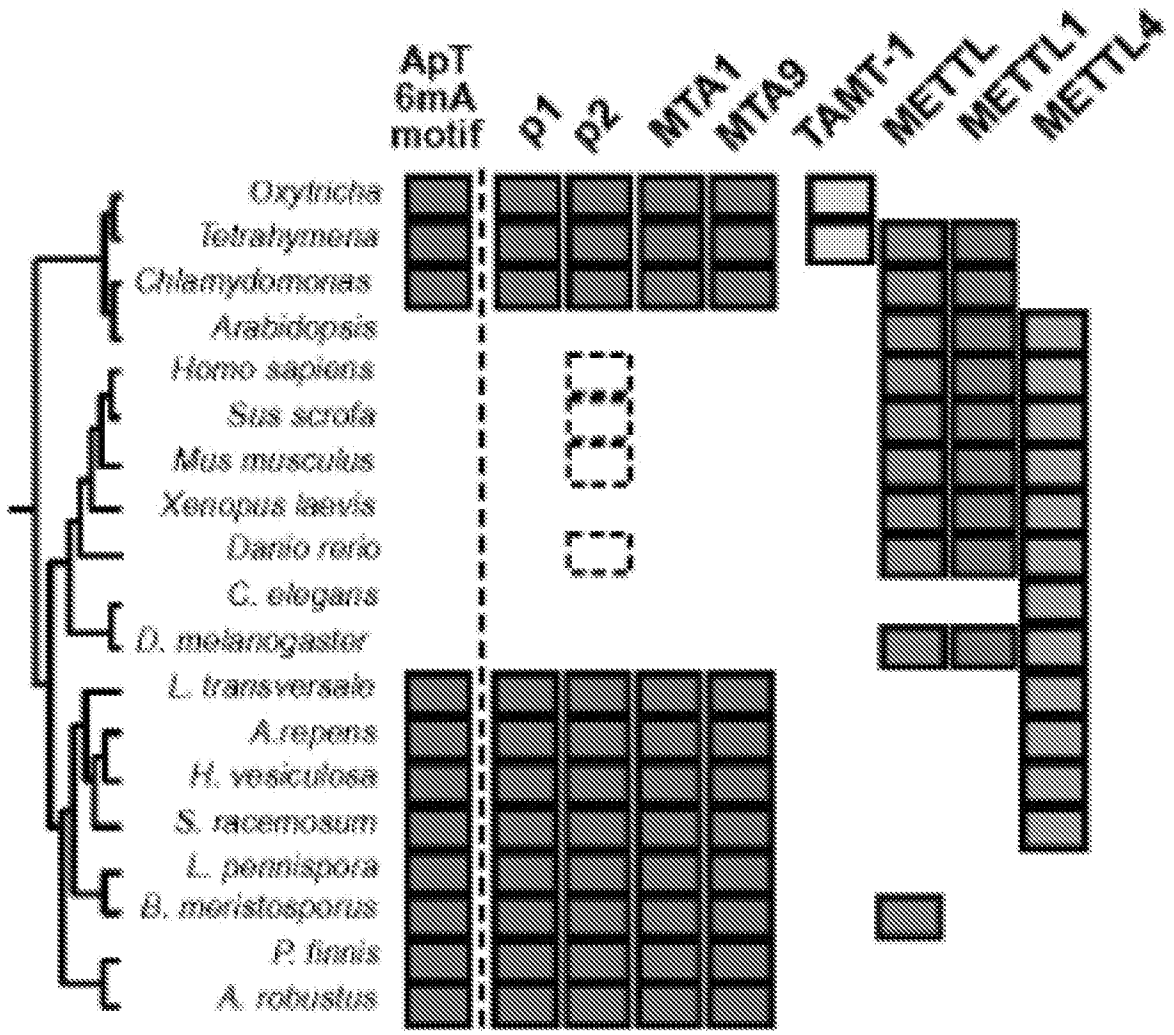

ApT 6mA motif
p1
p2
MTA1
MTA9
TAMT-1
METTL
METTL1
METTL4
Oxytricha
Tetrahymena
Chlamydomonas
Arabidopsis
Homo sapiens
Sus scrofa
Mus musculus
Xenopus laevis
Dario rerio
C. elegans
D. melanogaster
L. transversalis
A.repens
H. vesiculosa
S. racemosum
L. pennispora
B. meristosporus
P. finnis
A. robustus

Fig. 9K

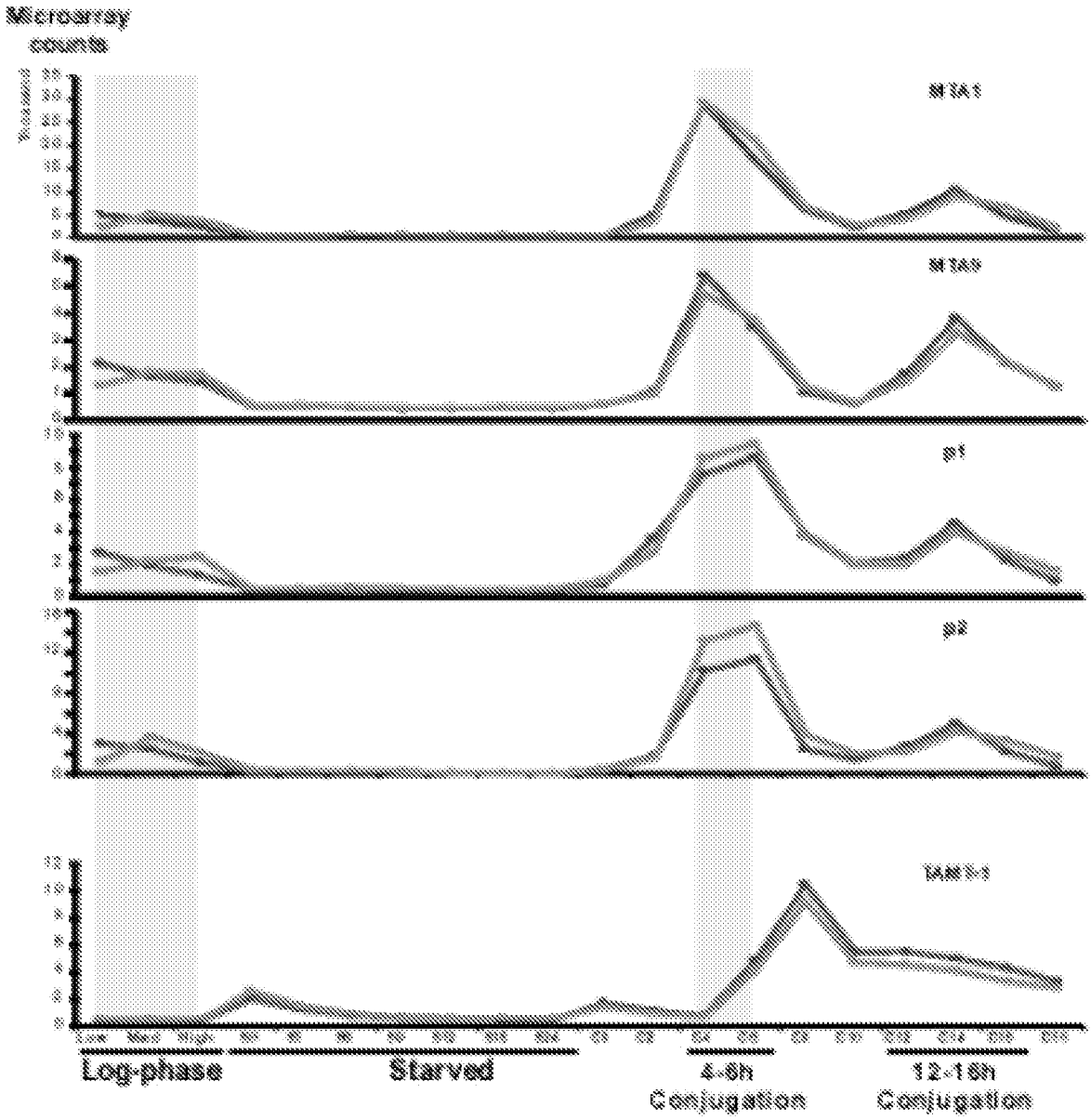

Microarray counts
Log-phase
Starved
4-6h Conjugation
12-18h Conjugation

Fig. 10A

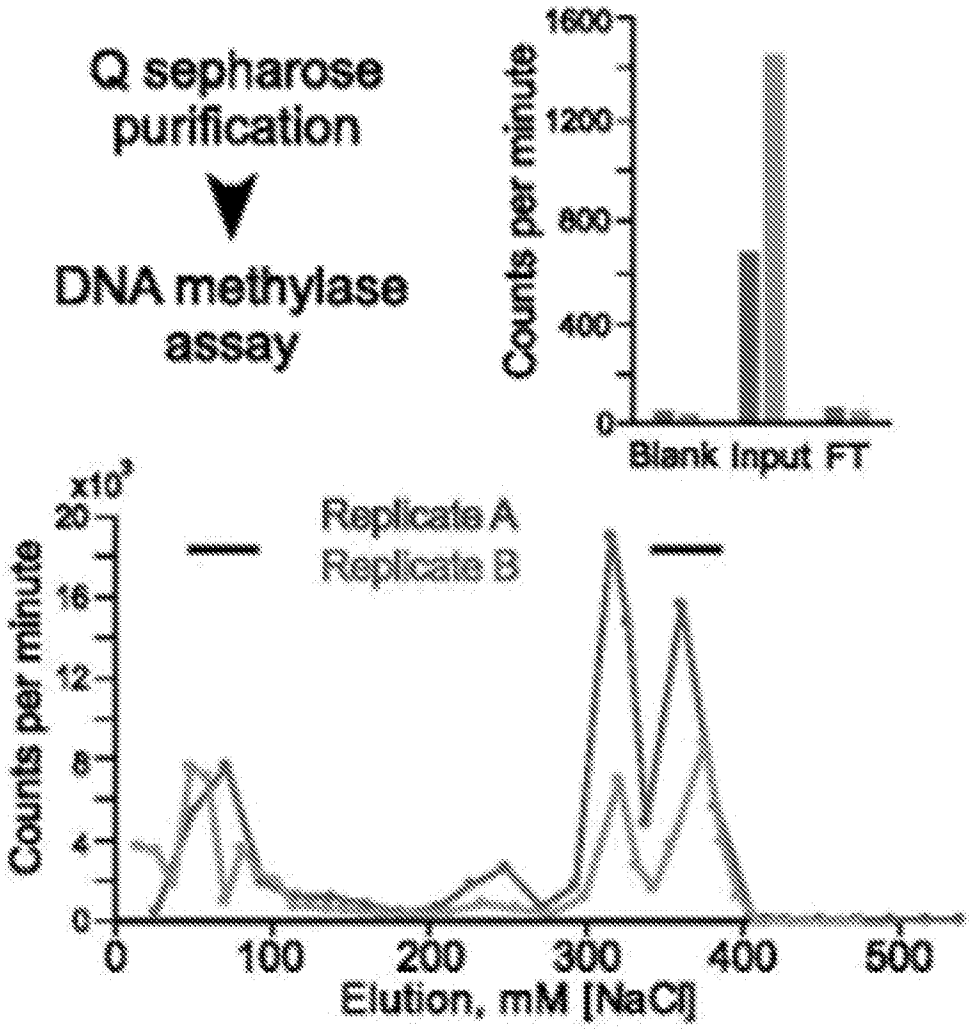

Q sepharose purification
DNA methylase assay
Counts per minute
1600
1200
800
400
0
Blank Input FT
x10³
Counts per minute
Replicate A
Replicate B
0
100
200
300
400
500
Elution, mM [NaCl]

Fig. 10B

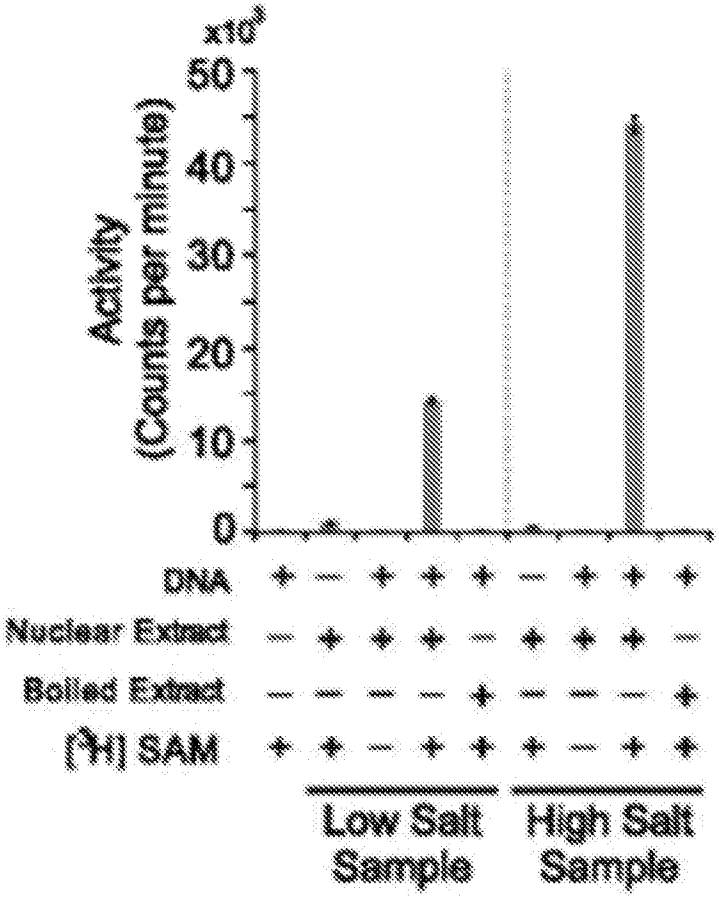


Fig. 10C

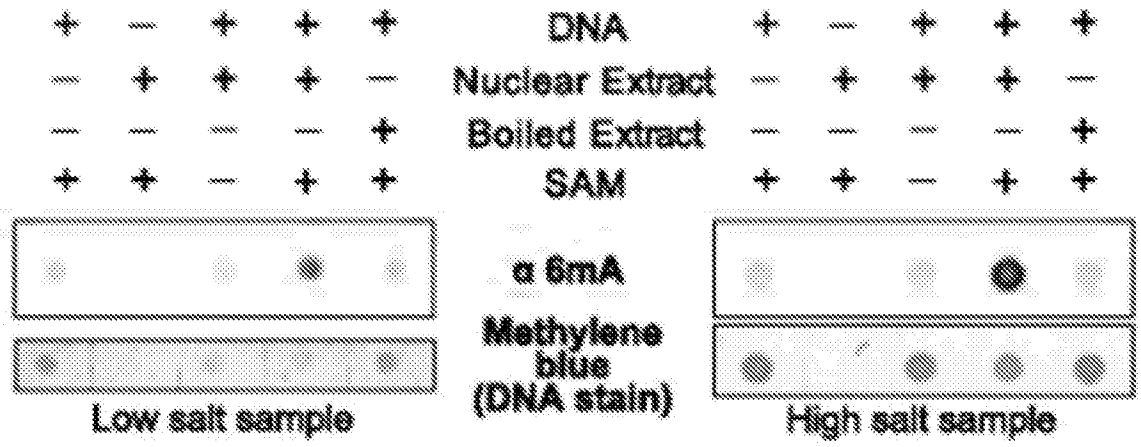


Fig. 10D

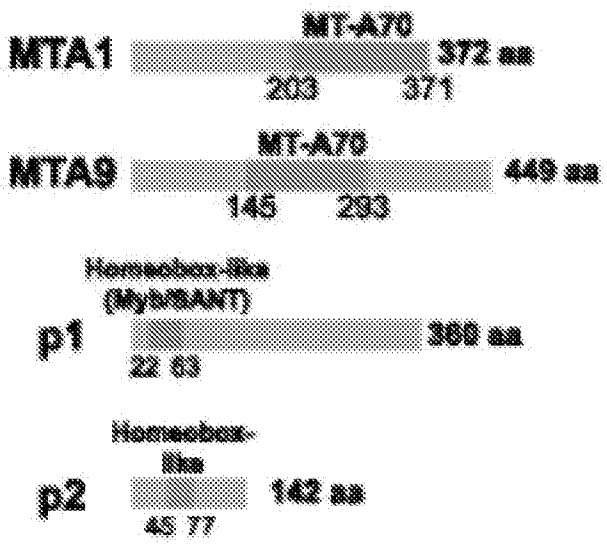


Fig. 10E

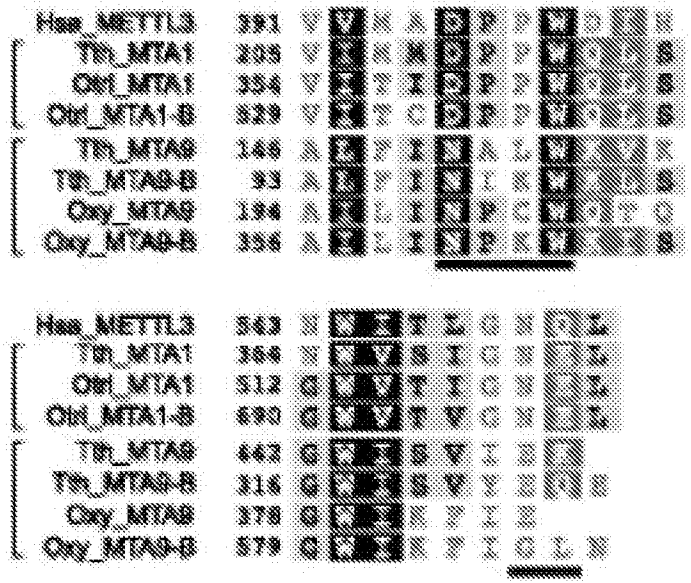


Fig. 10F

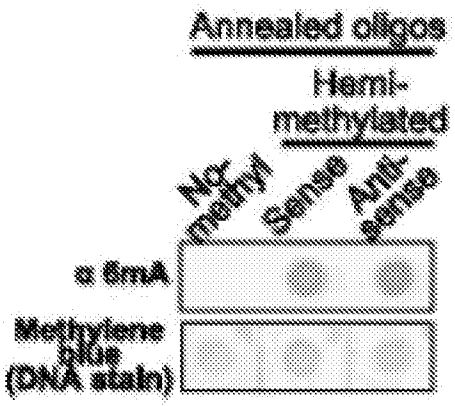


Fig. 10G

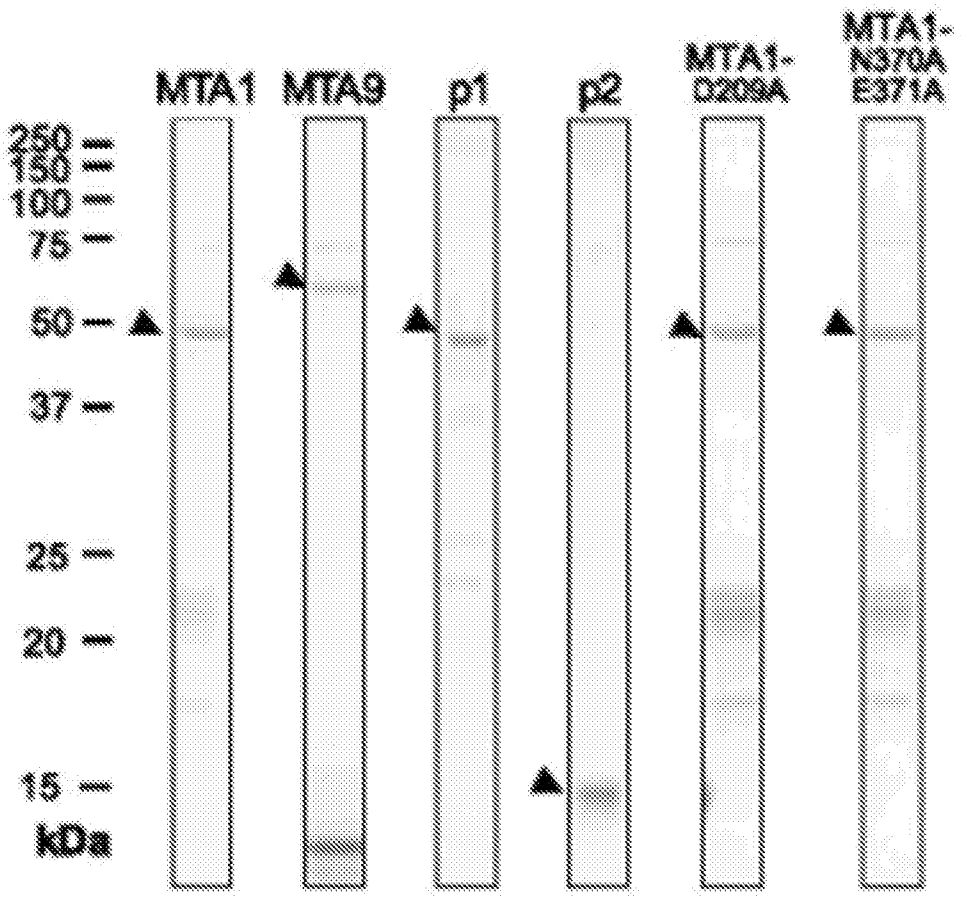
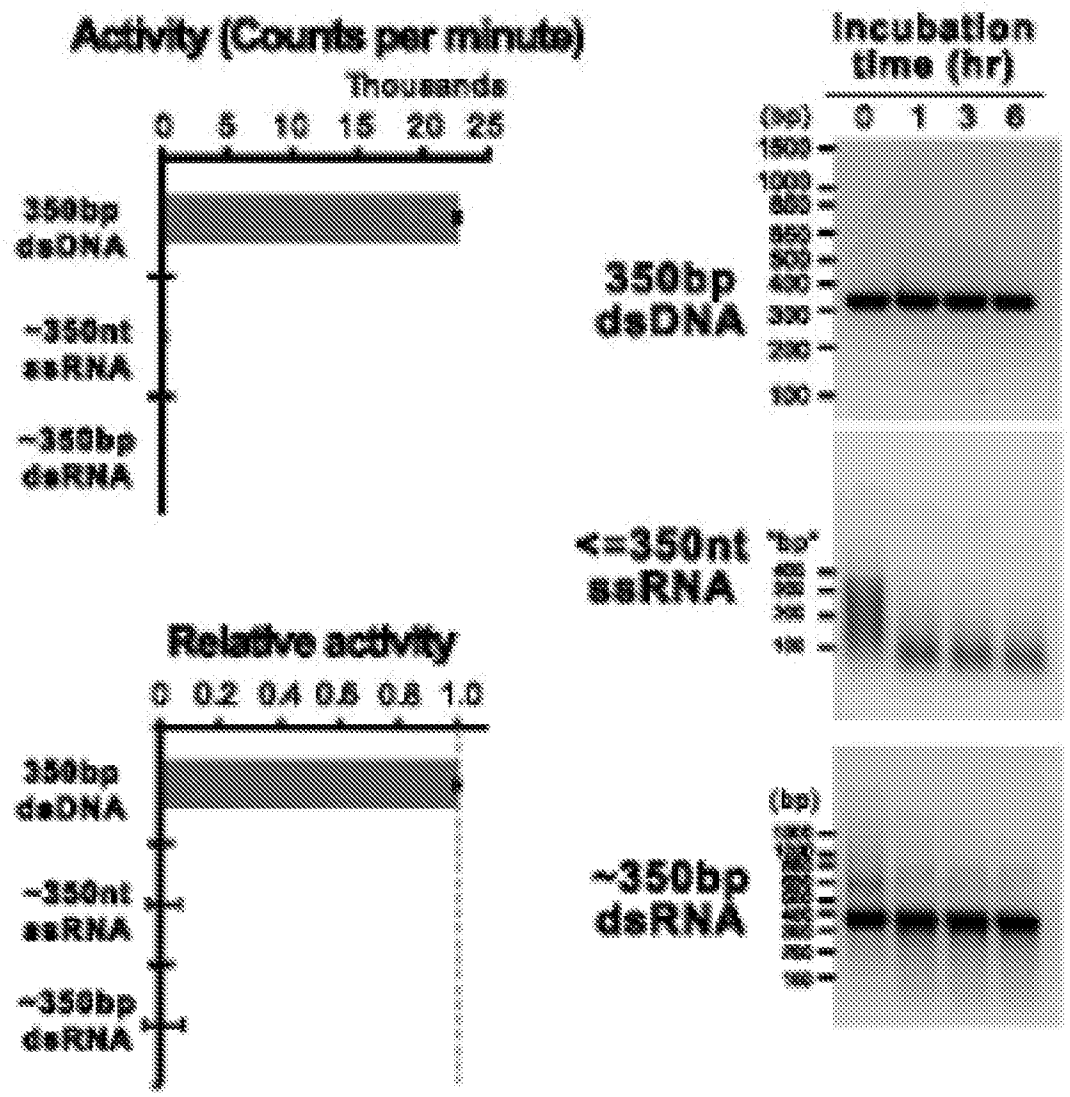


Fig. 10H


Activity (Counts per minute)
Thousands
0 5 10 15 20 25
350bp dsDNA
~350nt ssRNA
~350bp dsRNA
Relative activity
0 0.2 0.4 0.6 0.8 1.0
350bp dsDNA
~350nt ssRNA
~350bp dsRNA
Incubation time (hr)
0 1 3 6
350bp dsDNA
<=350nt ssRNA
~350bp dsRNA

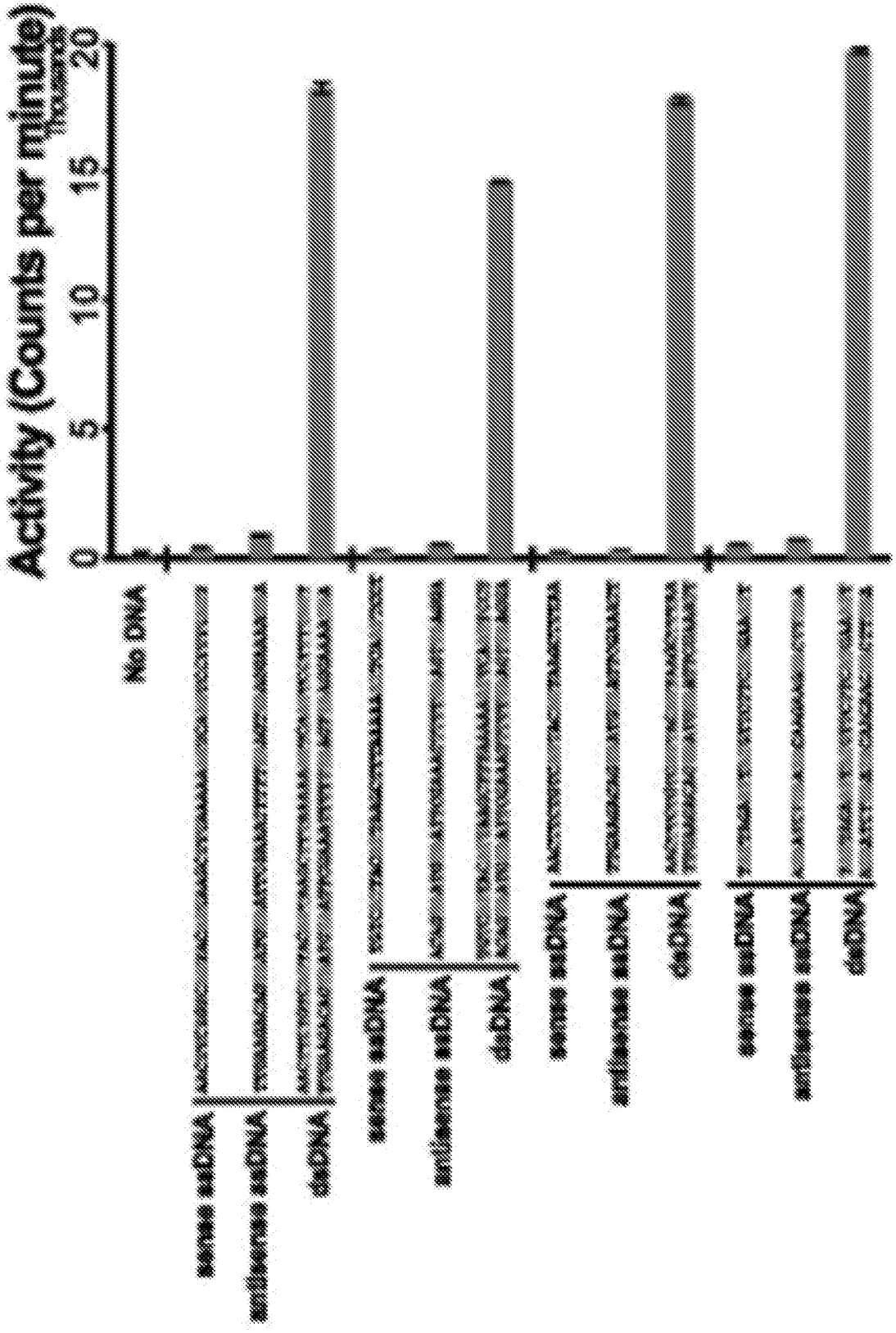
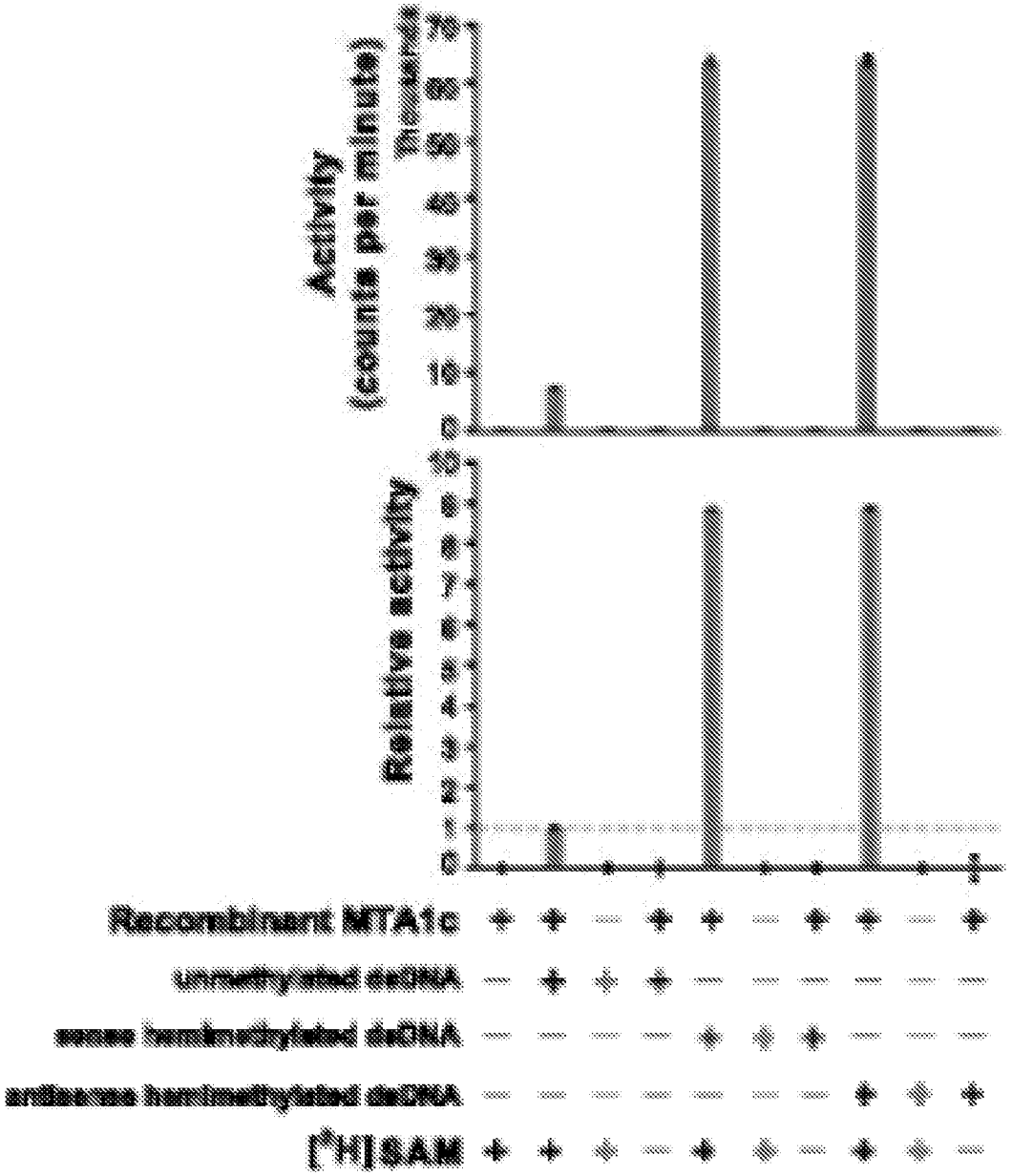

Activity (Counts per minute)
Thousands
0
5
10
15
20
No DNA
sense ssDNA
antisense ssDNA
dsDNA
sense ssDNA
antisense ssDNA
dsDNA
sense ssDNA
antisense ssDNA
dsDNA
sense ssDNA
antisense ssDNA
dsDNA

Fig. 10I

Fig. 10J


Activity
(counts per minute)
Relative activity
Recombinant MTA1c
unmethylated dsDNA
hemimethylated dsDNA
[3H]SAM

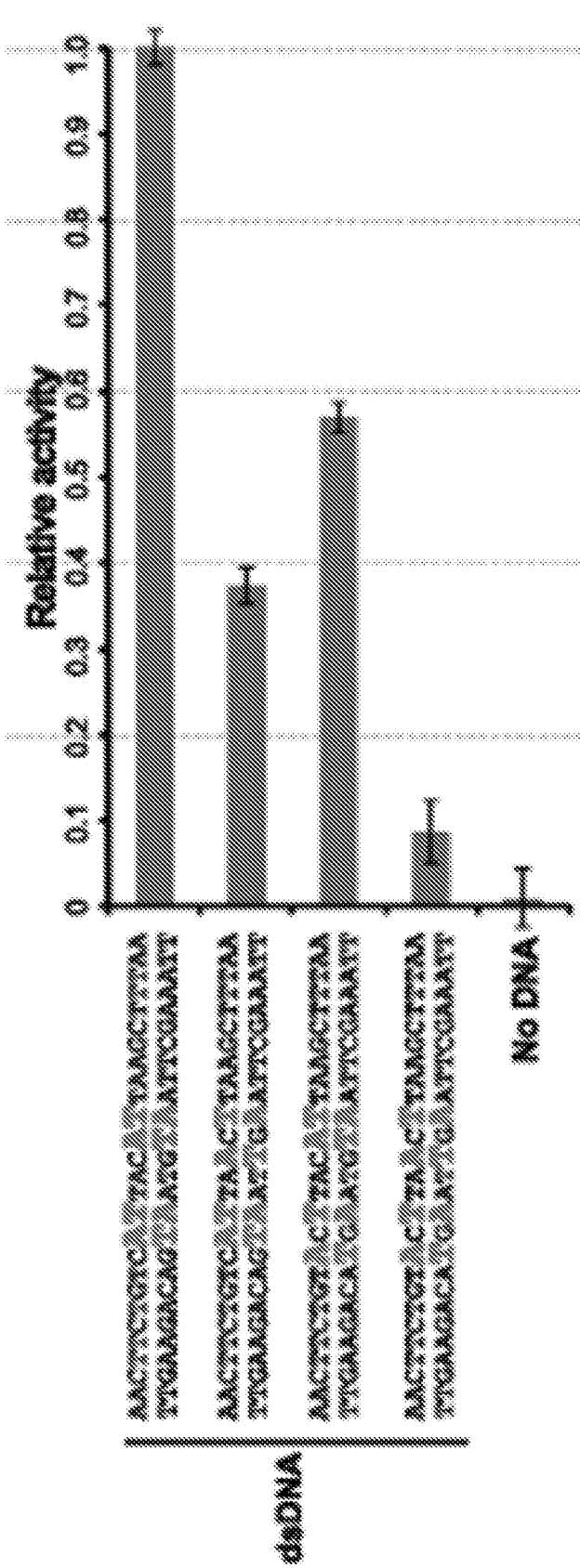


Fig. 10K

Fig. 10L

Fig. 10M

	Frequency (%)	
Motif	*Oxytricha*	*Tetrahymena*
AA*TA	0.68	9.54
AA*TT	5.57	13.30
AA*TC	10.81	7.38
AA*TG	9.98	7.68
TA*TA	0.18	2.38
TA*TT	0.75	12.18
TA*TC	1.61	4.74
TA*TG	0.64	4.67
CA*TA	0.51	4.03
CA*TT	10.42	7.55
CA*TC	13.70	4.32
CA*TG	6.60	2.50
GA*TA	1.31	4.25
GA*TT	11.62	8.00
GA*TC	12.33	3.13
GA*TG	13.33	4.47

Fig. 10N

	Frequency (%)	
Motif	*Oxytricha*	*Tetrahymena*
AATA	11.49	14.62
AATT	14.17	9.68
AATC	6.41	3.53
AATG	5.88	4.61
TATA	8.14	16.73
TATT	11.49	14.62
TATC	4.98	4.74
TATG	4.18	5.76
CATA	4.18	5.76
CATT	5.88	4.61
CATC	3.36	2.16
CATG	2.46	1.59
GATA	4.98	4.74
GATT	6.41	3.53
GATC	2.64	1.16
GATG	3.36	2.16

Fig. 11A

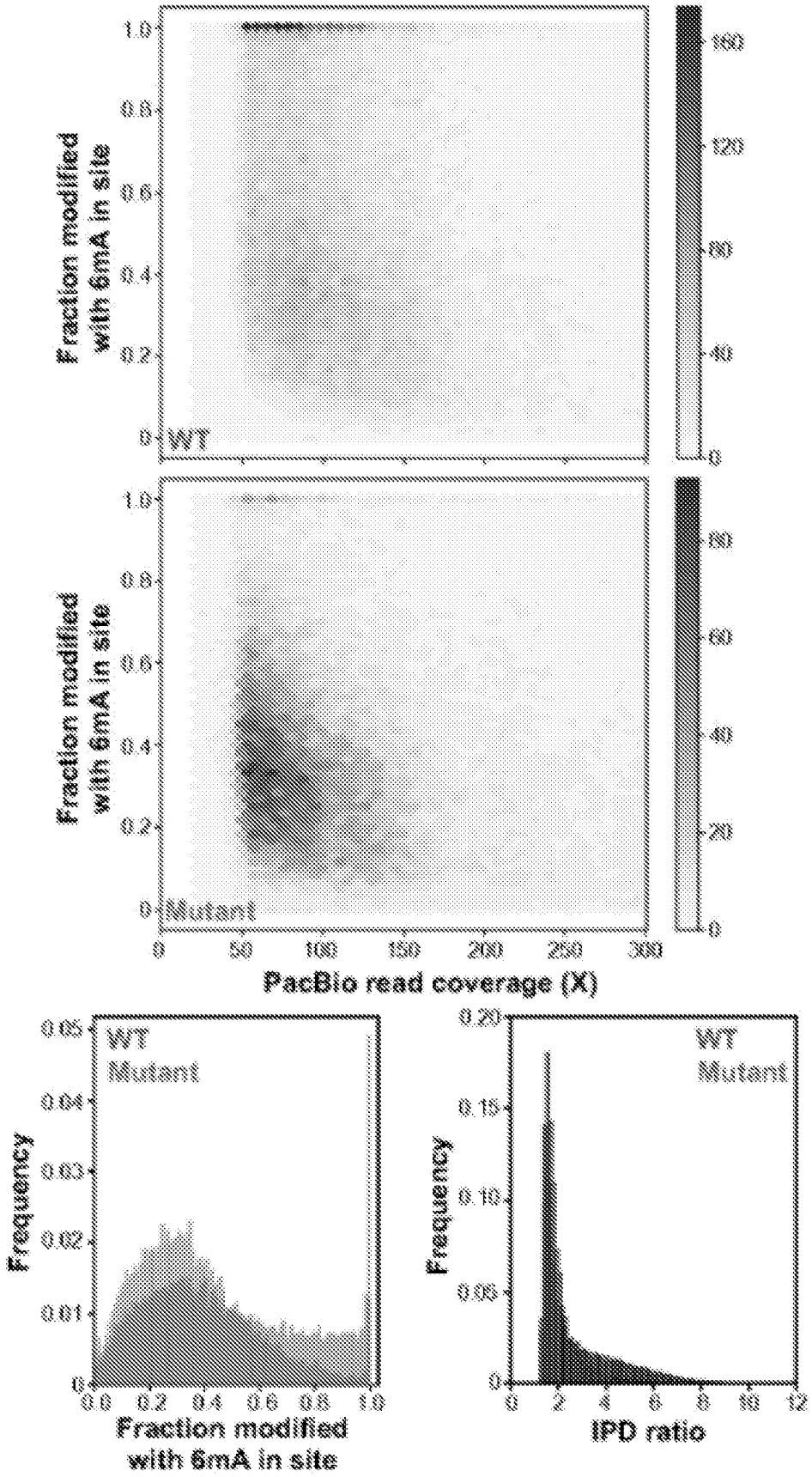

Fraction modified with 6mA in site
WT
Mutant
PacBio read coverage (X)
Frequency
Fraction modified with 6mA in site
IPD ratio

Fig. 11B

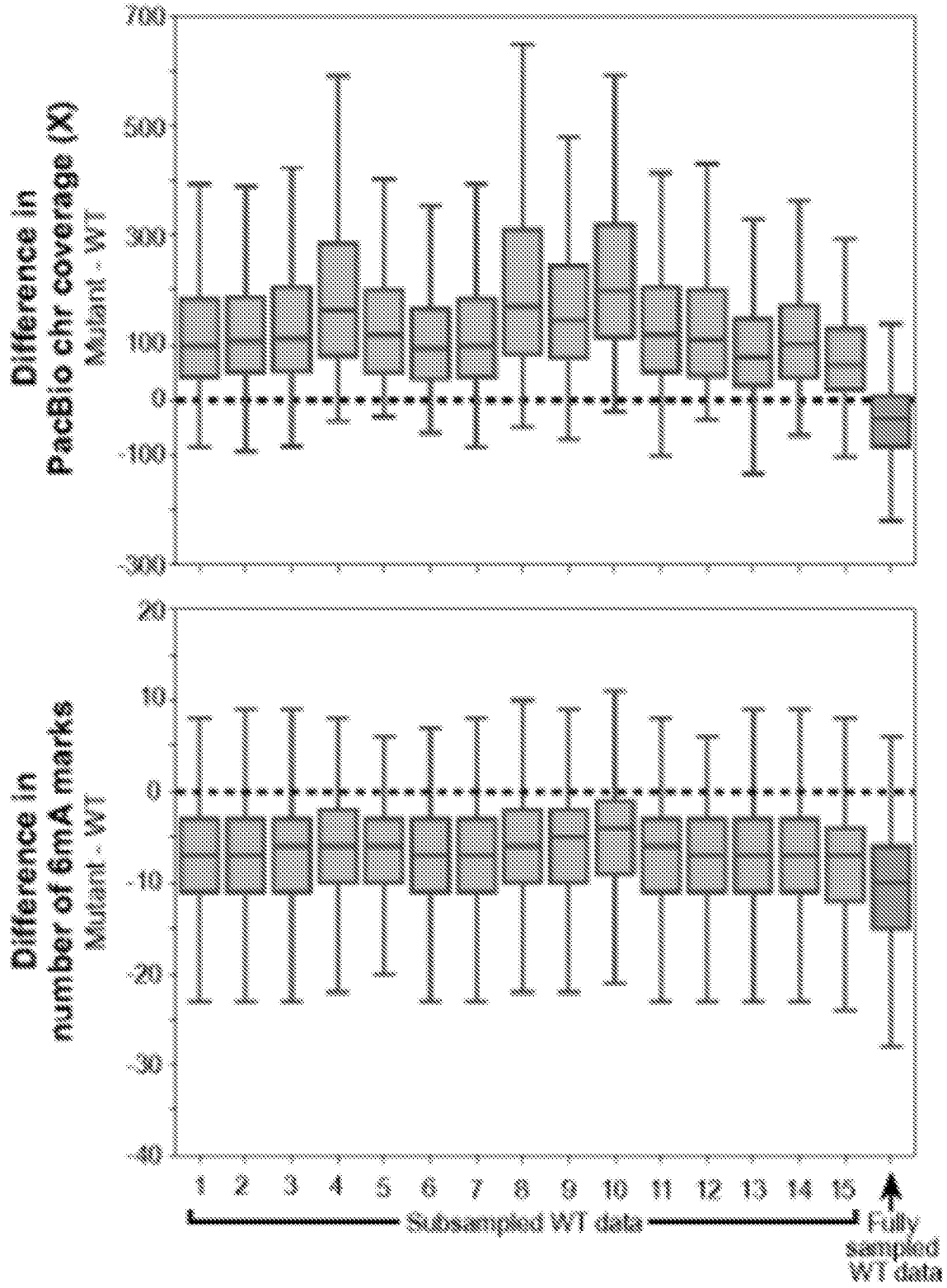
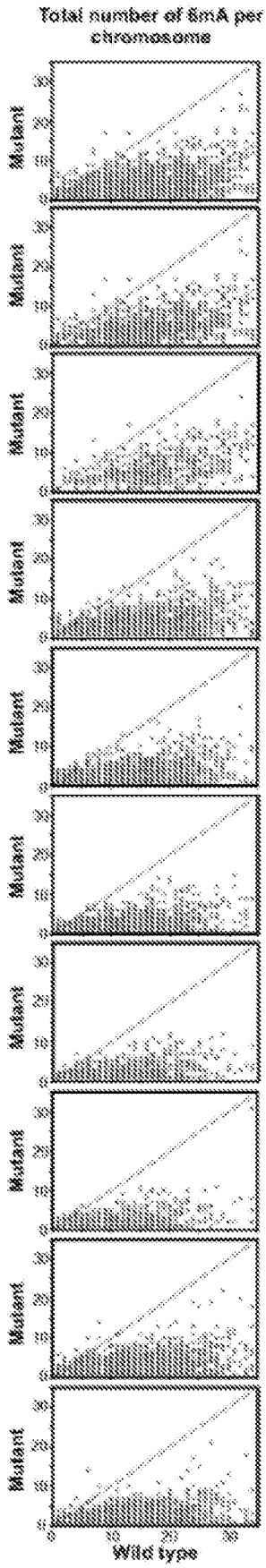

Difference in PacBio chr coverage (X)
Mutant - WT
700
500
300
100
0
-100
-300
Difference in number of 6mA marks
Mutant - WT
20
10
0
-10
-20
-30
-40
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Subsampled WT data
Fully sampled WT data

Fig. 11C



Min. PacBio chromosome coverage	Min. Confidence score	Min. Q-value
100x	30	20
150x	30	20
200x	30	20
100x	30	30
100x	30	40
100x	30	50
100x	30	60
100x	30	70
100x	40	20
100x	50	20

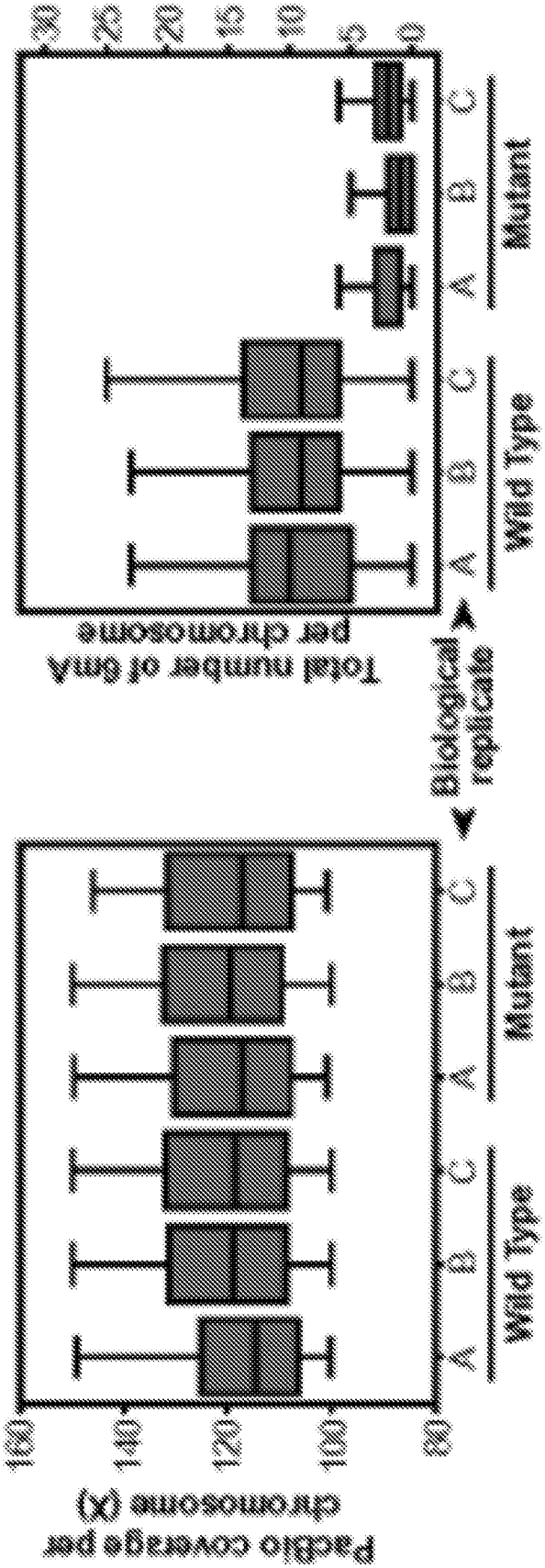

PacBio coverage per chromosome (X)
160
140
120
100
80
Wild Type
Mutant
A
B
C
Biological replicate
Total number of bma per chromosome
0
5
10
15
20
25
30
Wild Type
Mutant
A
B
C

Fig. 11D

Fig. 12A

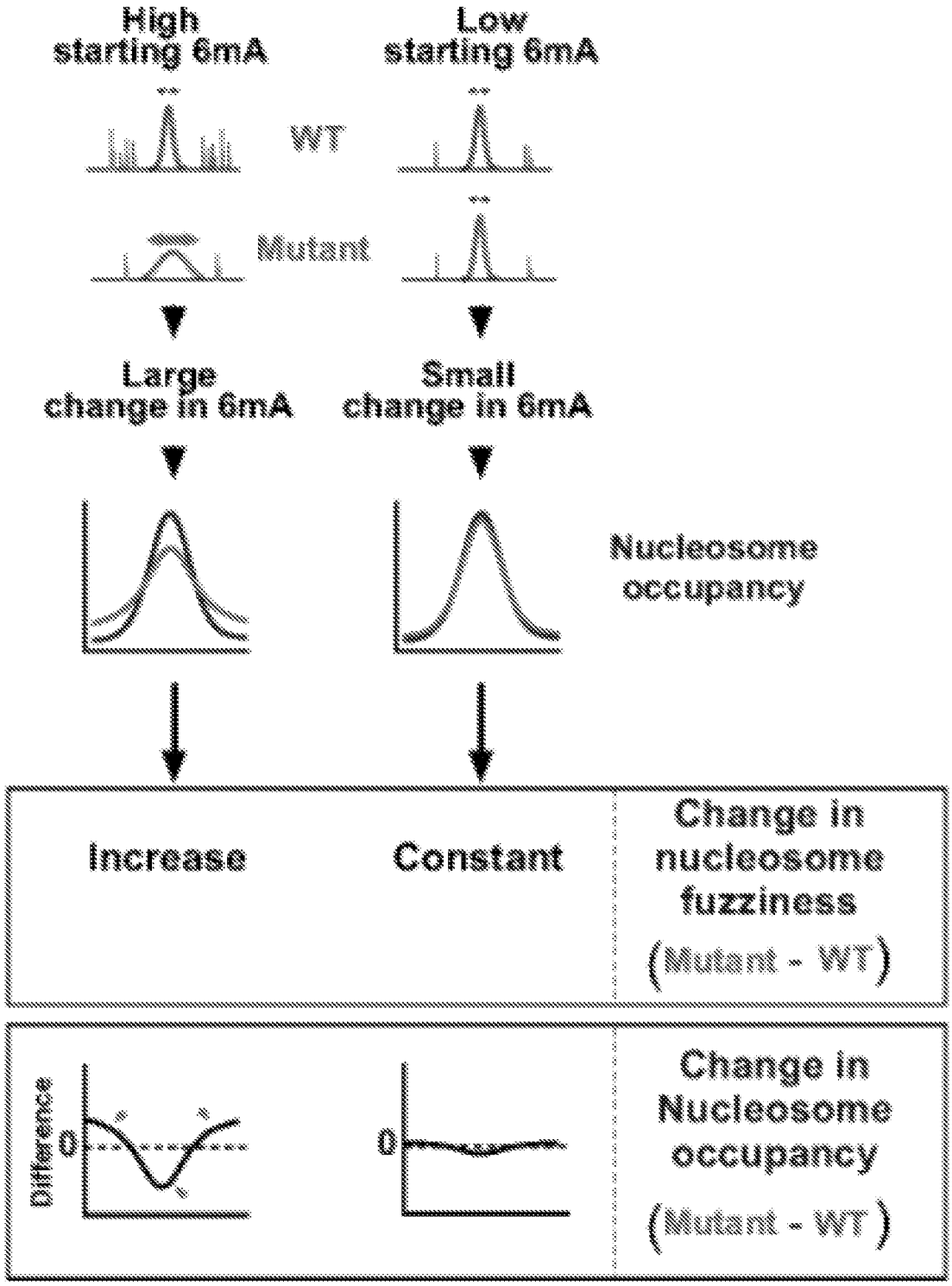

High starting 6mA
Low starting 6mA
WT
Mutant
Large change in 6mA
Small change in 6mA
Nucleosome occupancy
Increase
Constant
Change in nucleosome fuzziness (Mutant - WT)
Difference
0
0
Change in Nucleosome occupancy (Mutant - WT)

Fig. 12B

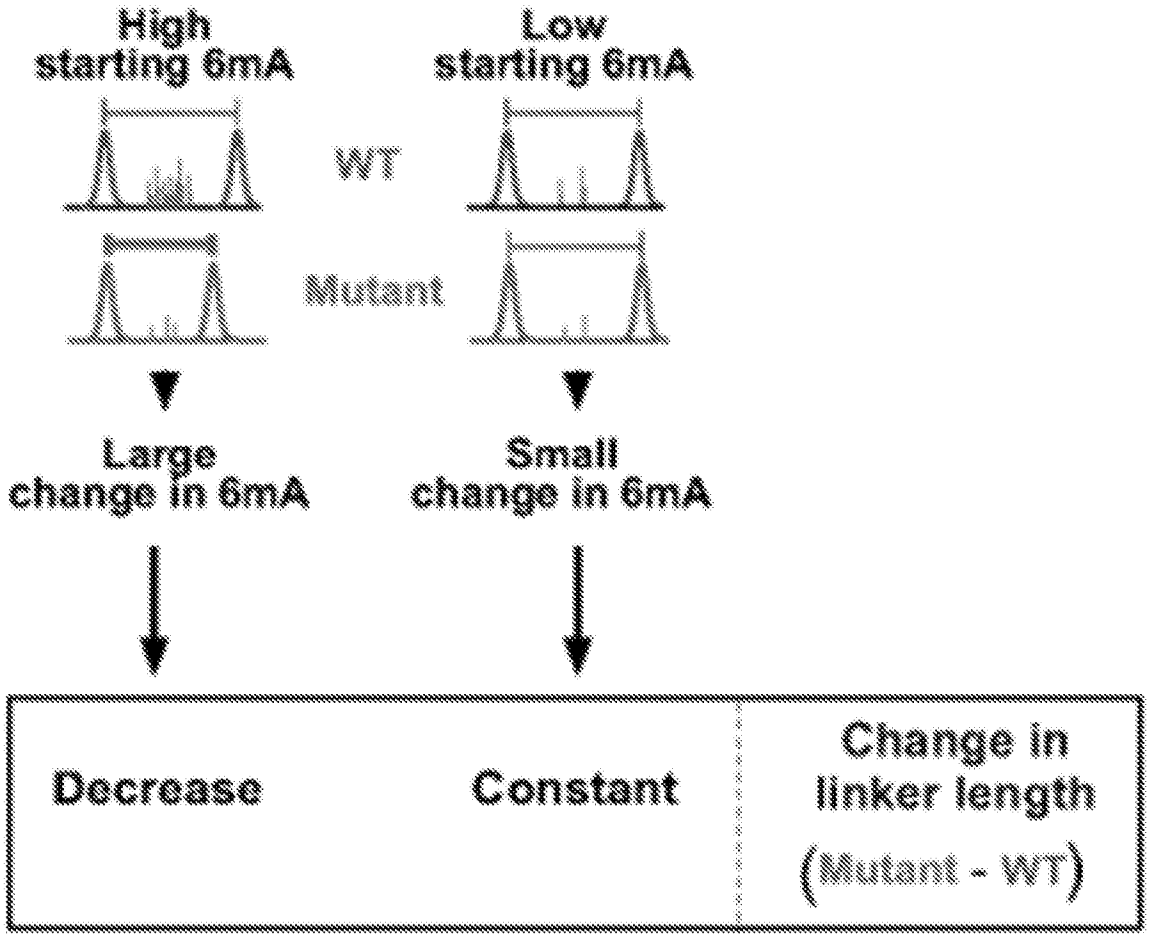

High starting 6mA
Low starting 6mA
WT
Mutant
Large change in 6mA
Small change in 6mA
Decrease
Constant
Change in linker length (Mutant - WT)

Fig. 12C

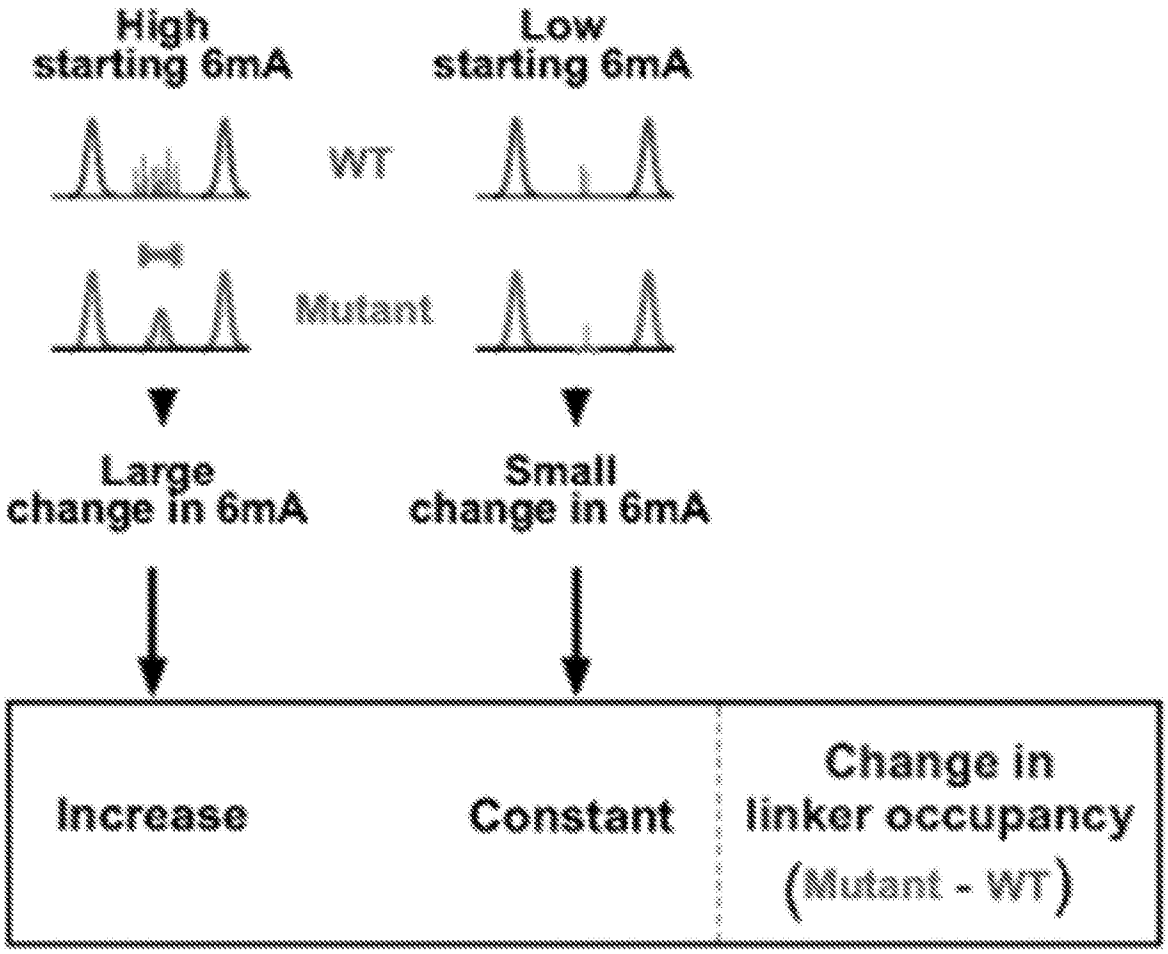


Fig. 12D

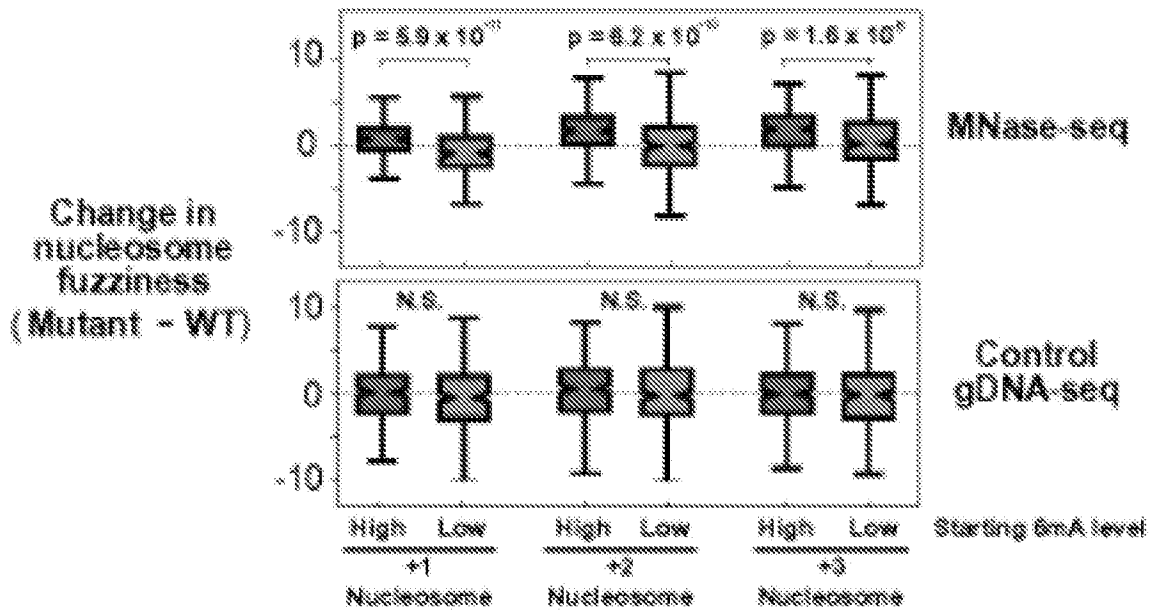


Fig. 12E

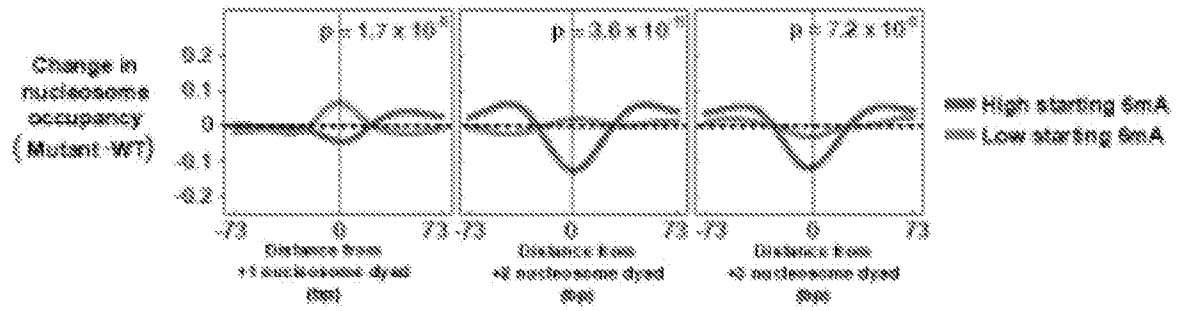


Fig. 12F

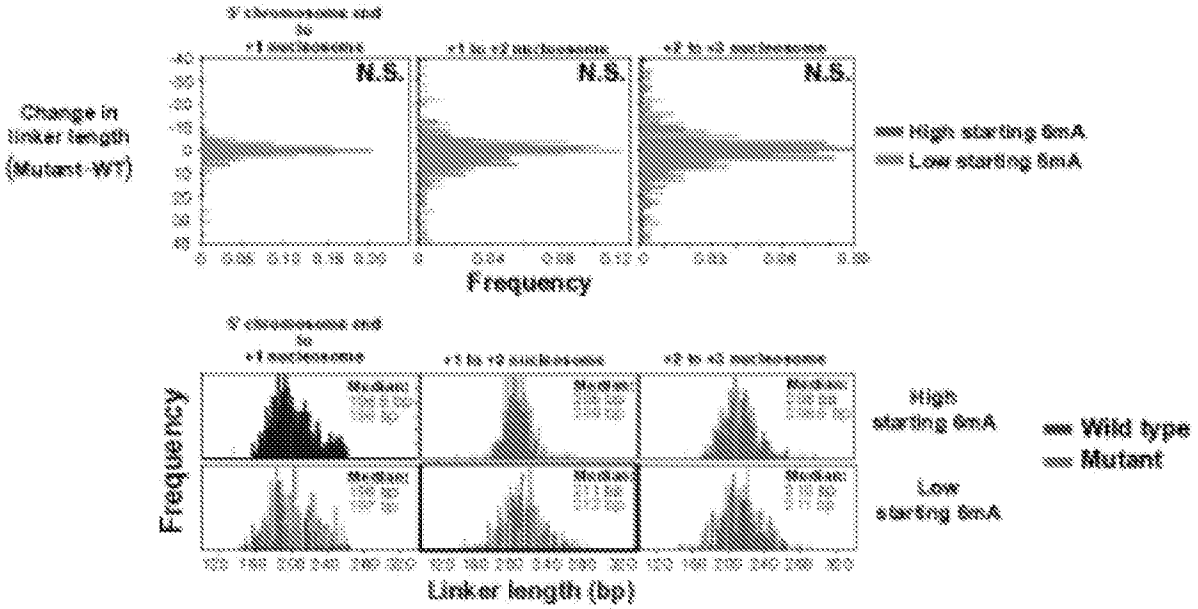


Fig. 12G

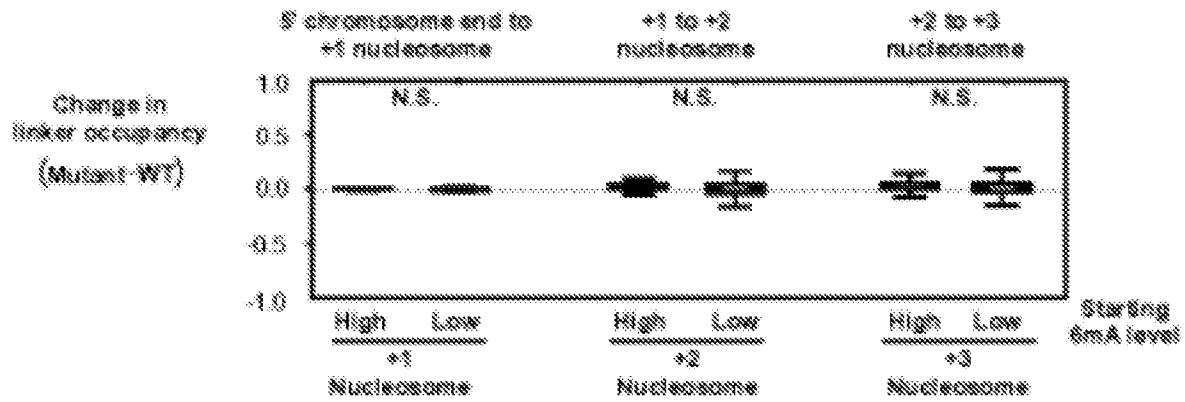


Fig. 12H

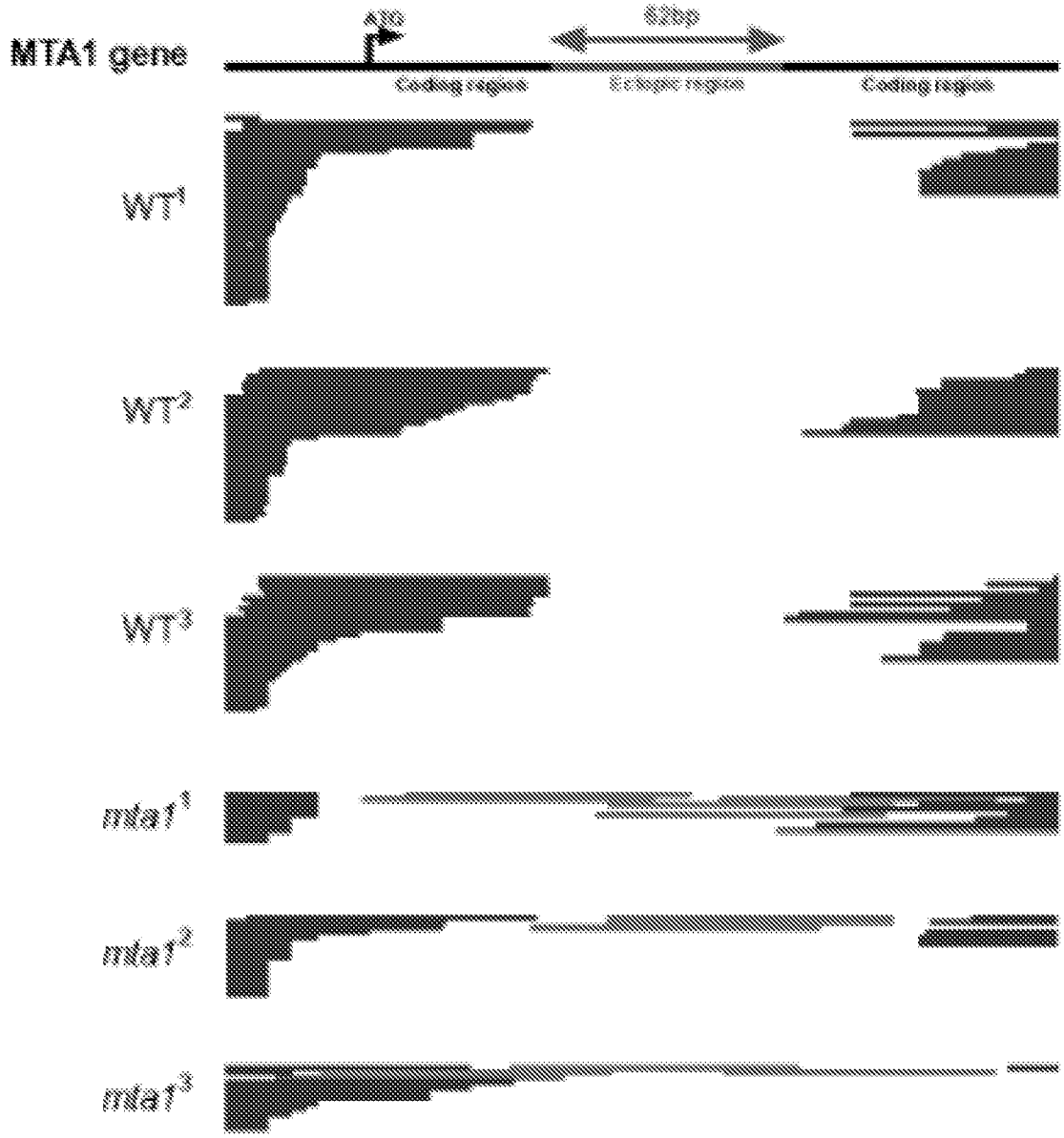

MTA1 gene
Coding region
Coding region
WT1
WT2
WT3
mta1-1
mta1-2
mta1-3

Fig. 13A

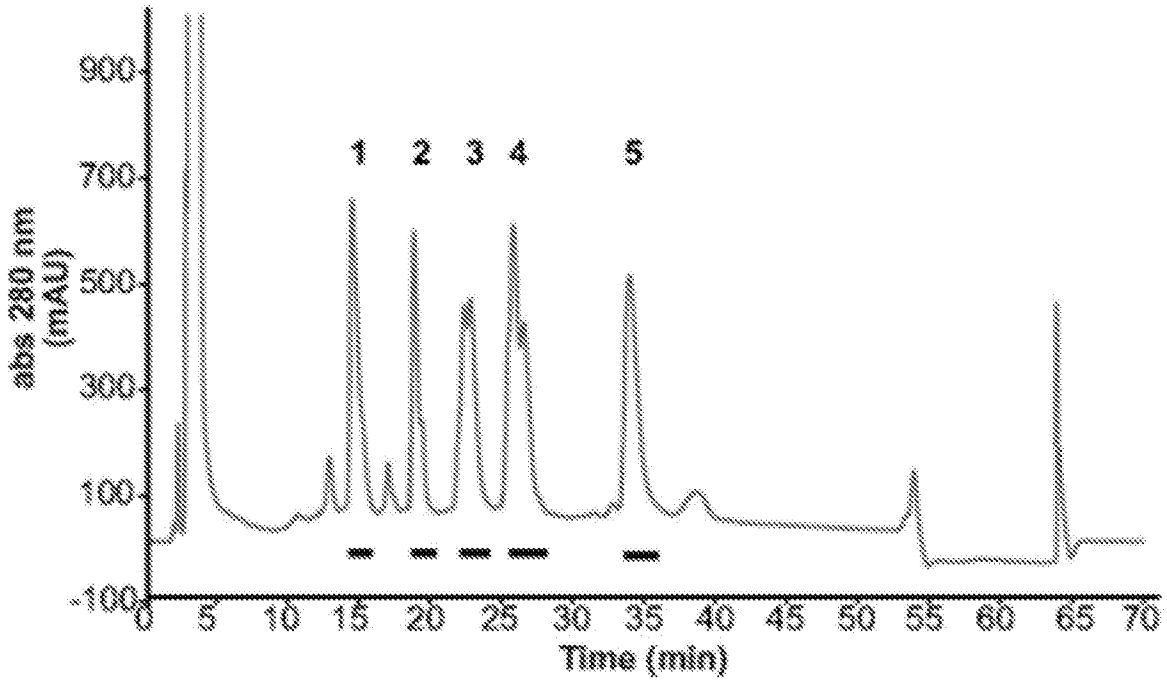


Fig. 13B

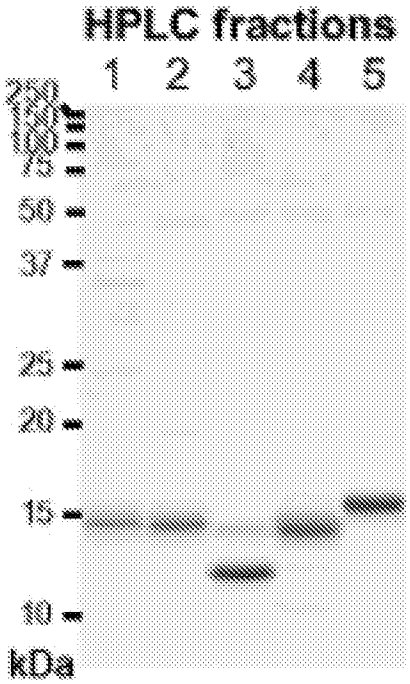


Fig. 13C

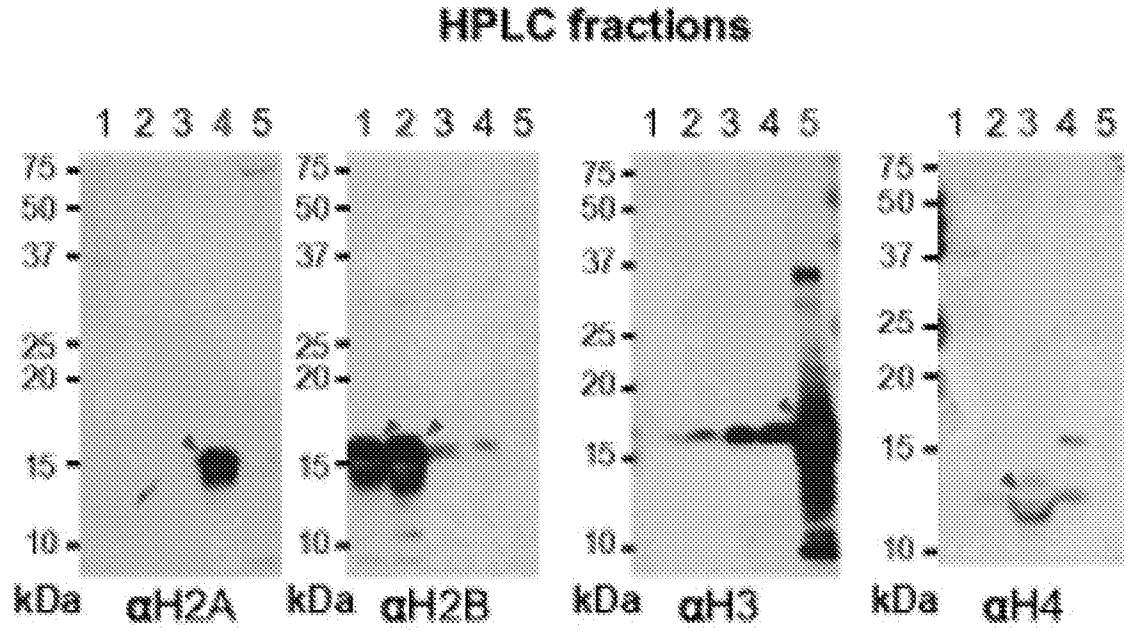


Fig. 13D

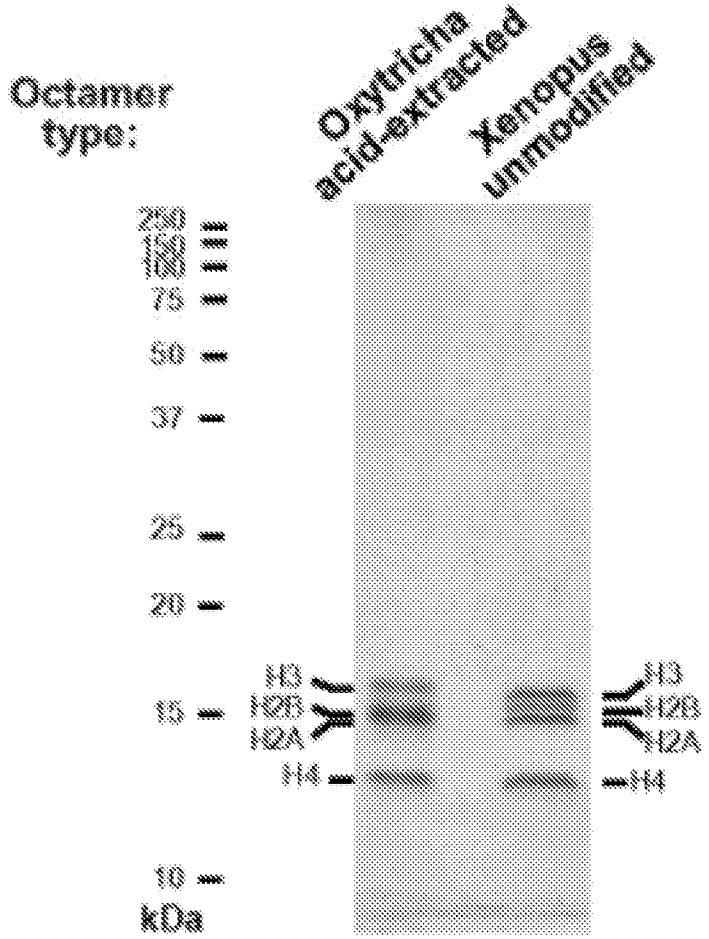


Fig. 13E

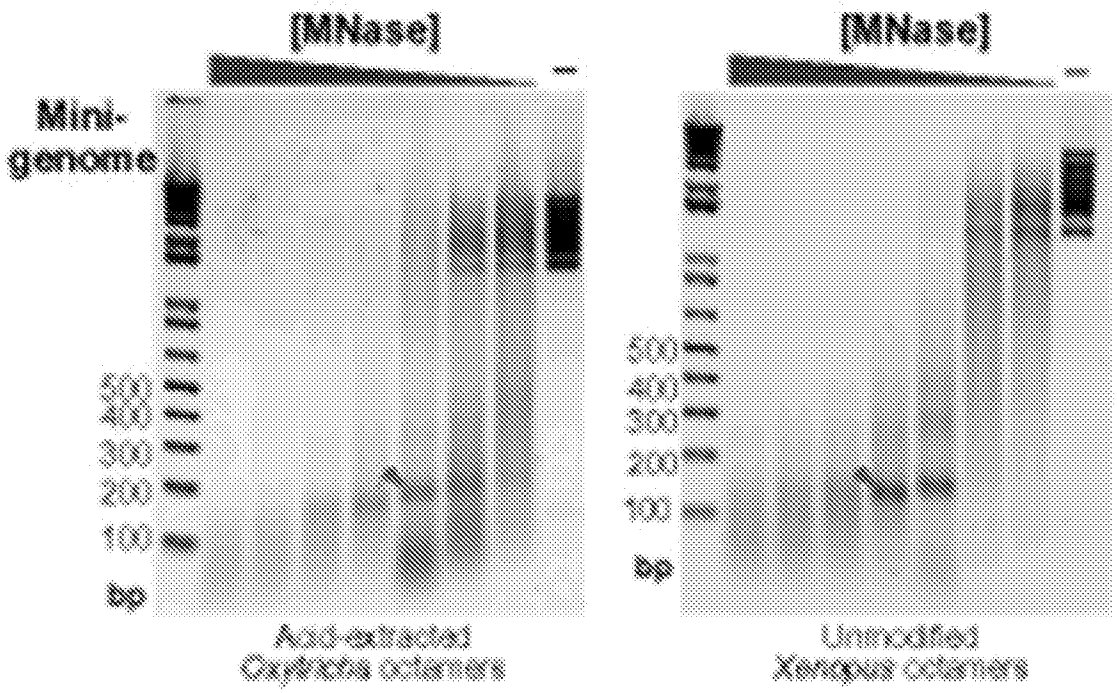


Fig. 13F

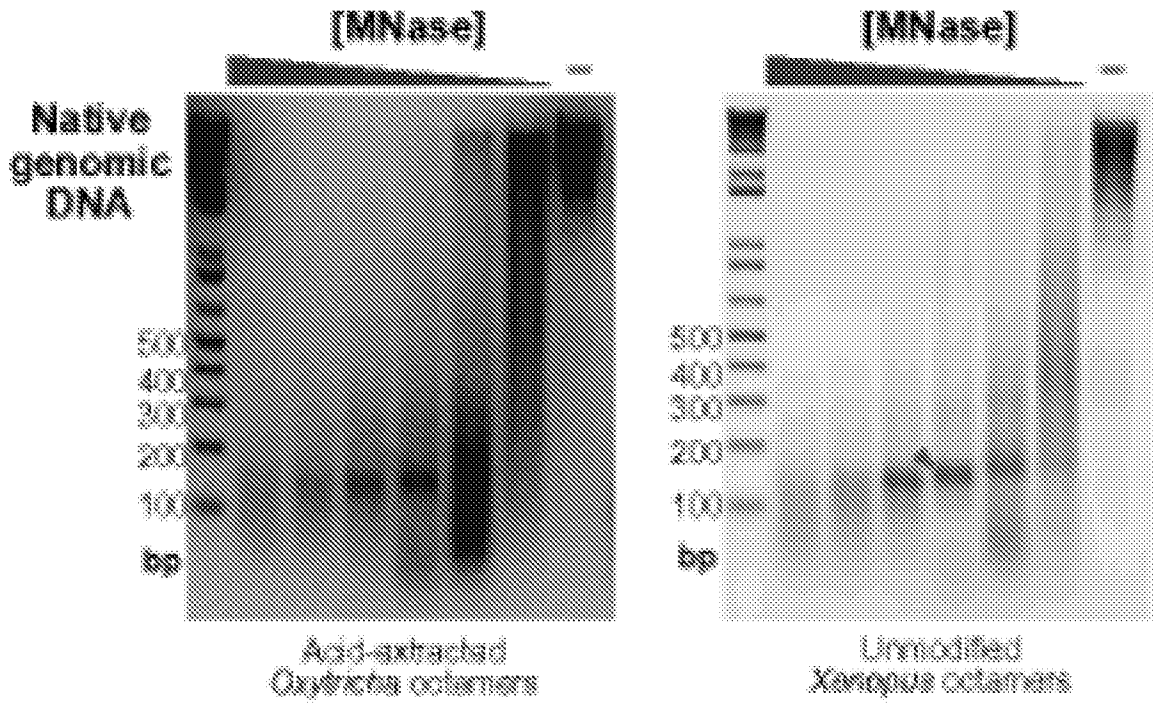


Fig. 13G

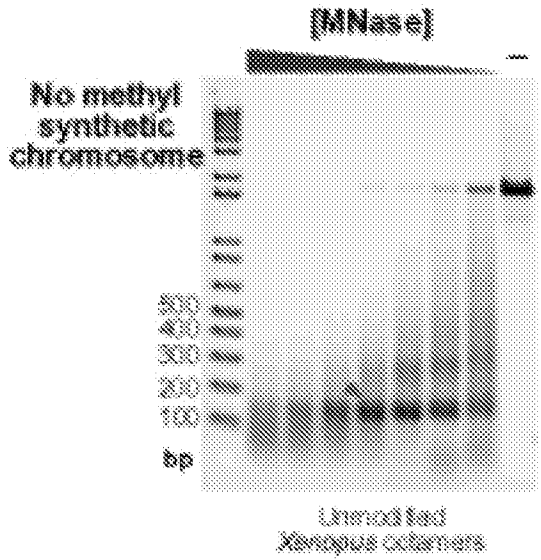


Fig. 13H

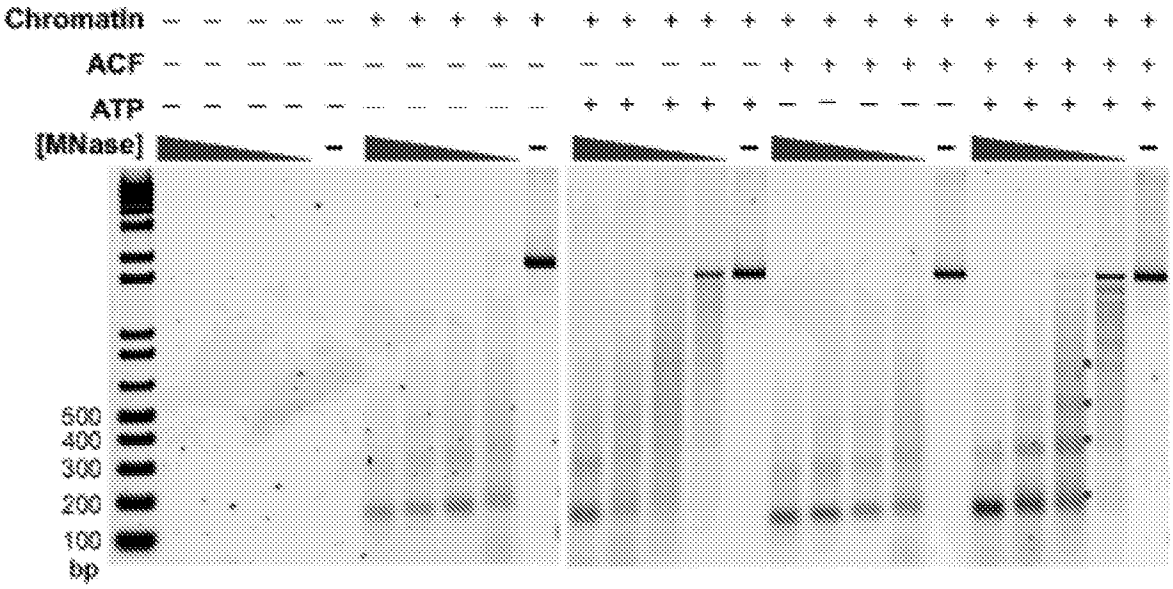


Fig. 13l

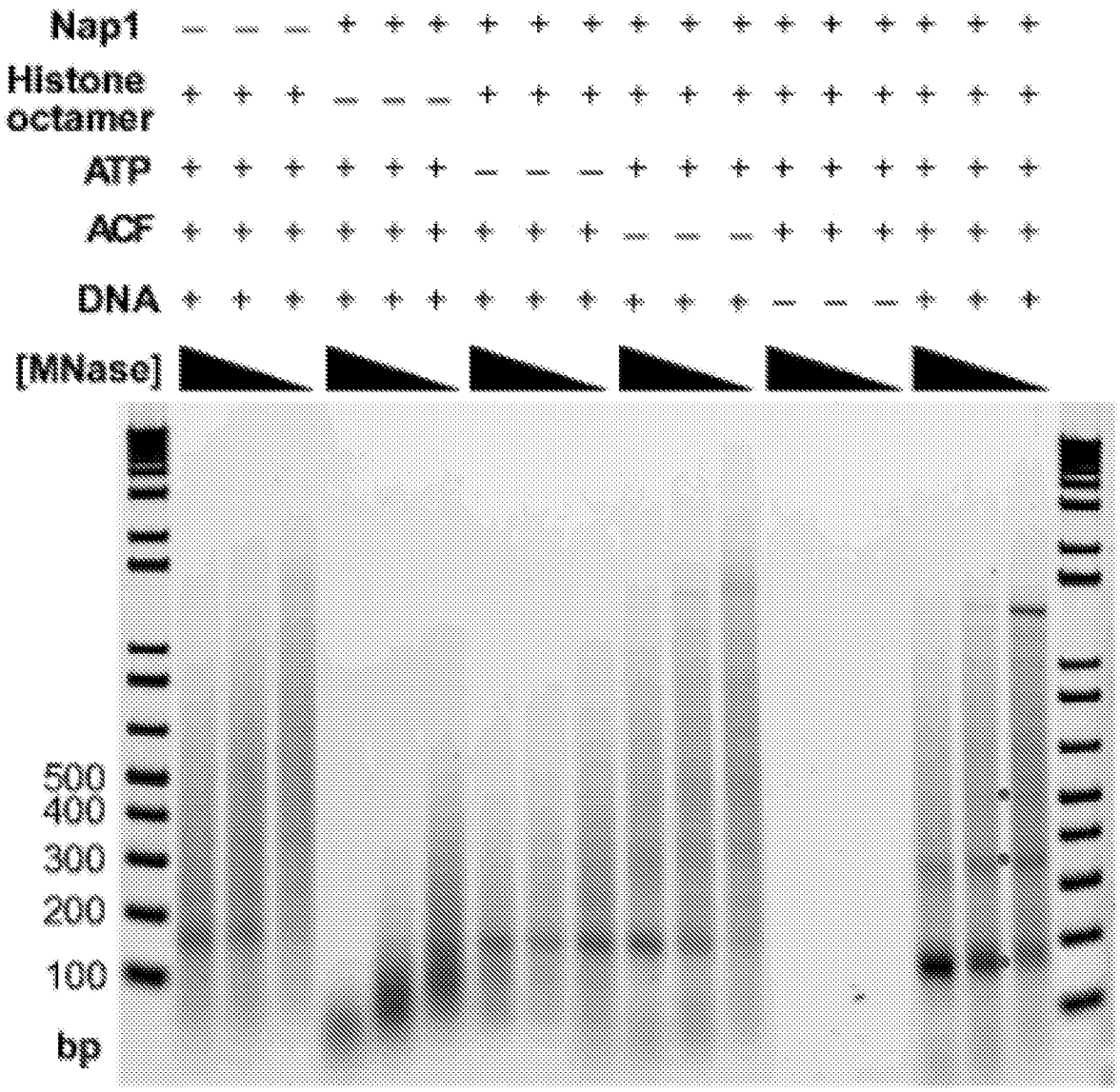


Fig. 14A

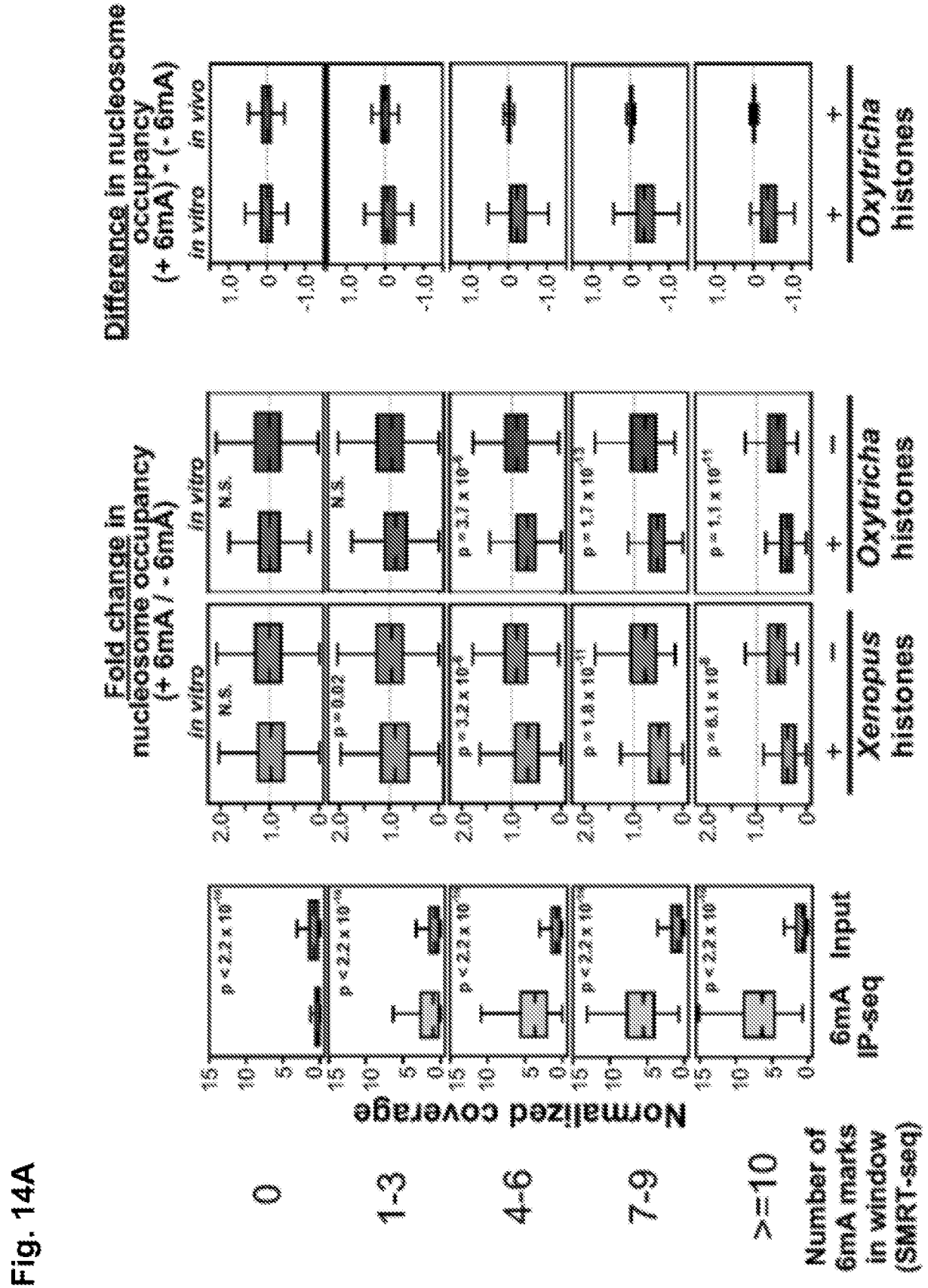

Number of 6mA marks in window (SMRT-seq)
0
1-3
4-6
7-9
>=10
Normalized coverage
6mA IP-seq
Input
Fold change in nucleosome occupancy (+ 6mA / - 6mA)
in vitro
N.S.
p = 0.02
p = 3.2 x 10^-2
p = 1.8 x 10^-11
p = 6.1 x 10^-6
p = 3.7 x 10^-2
p = 1.7 x 10^-13
p = 1.1 x 10^-13
+ + — —
Xenopus histones
+ + — —
Oxytricha histones
Difference in nucleosome occupancy (+ 6mA) - (- 6mA)
in vitro
in vivo
+ + + +
Oxytricha histones

Fig. 14B

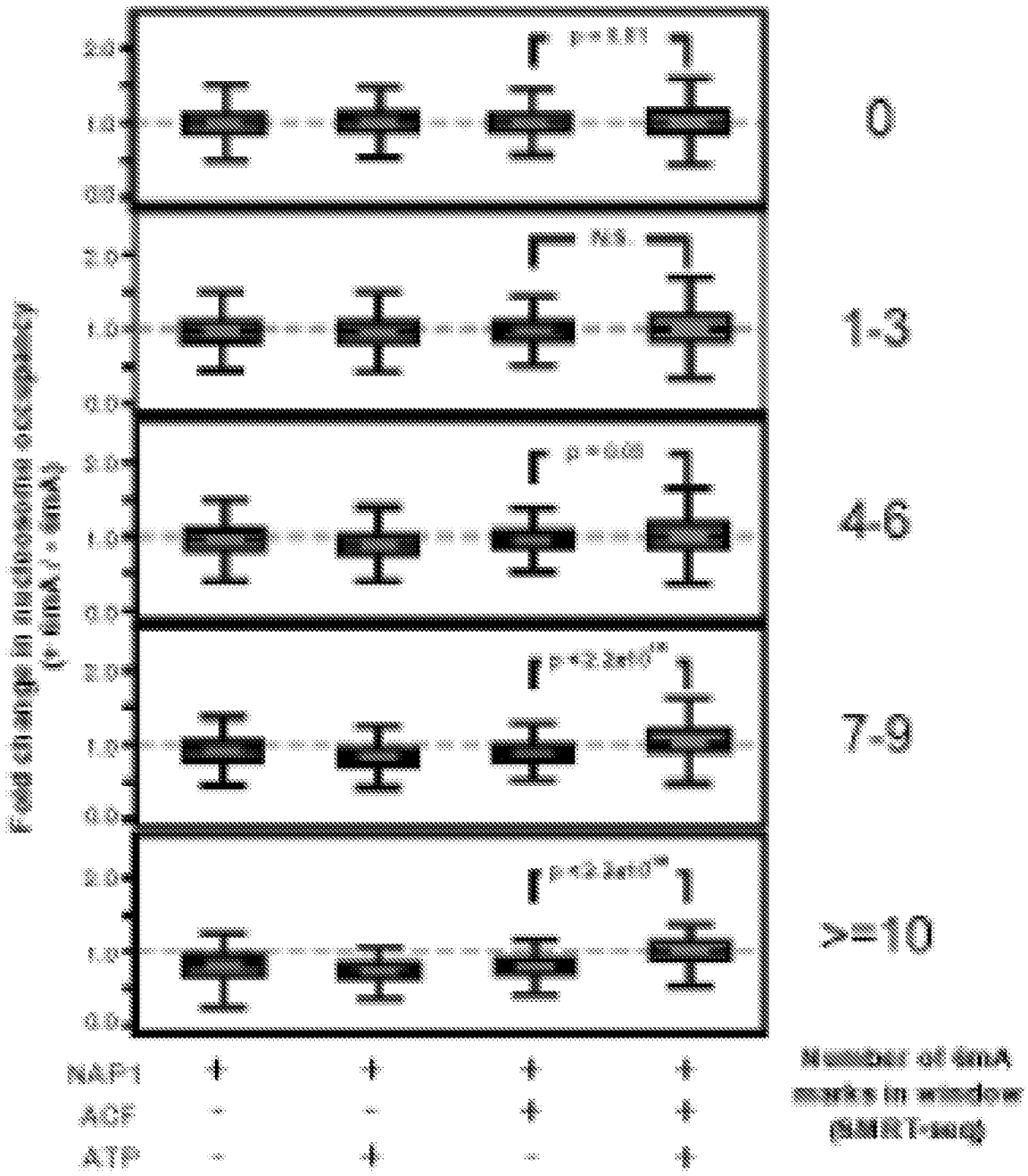


Fig. 14C

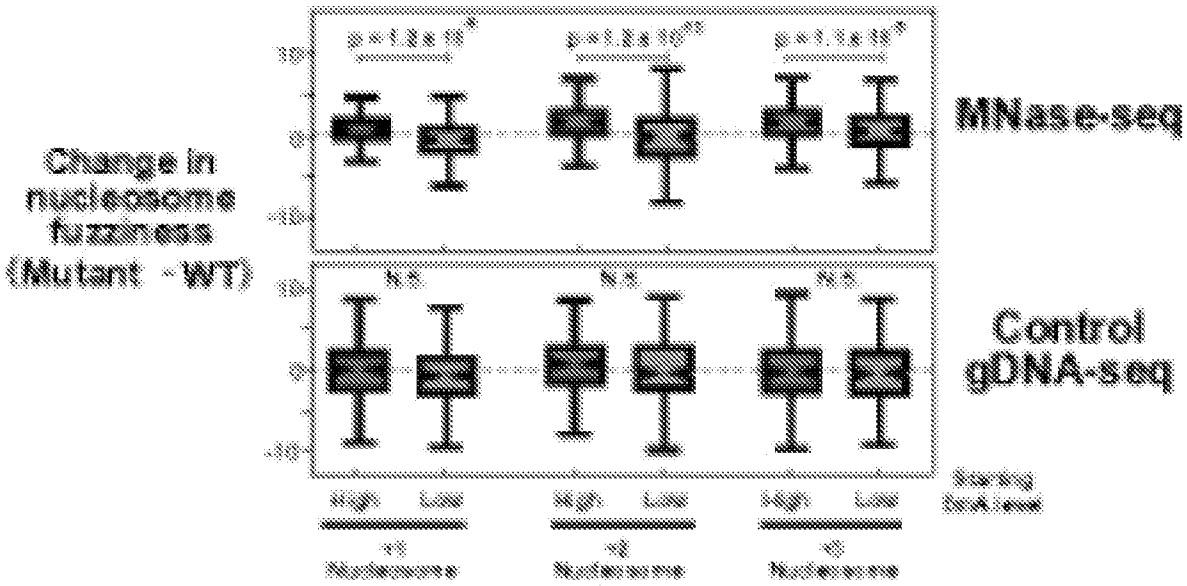
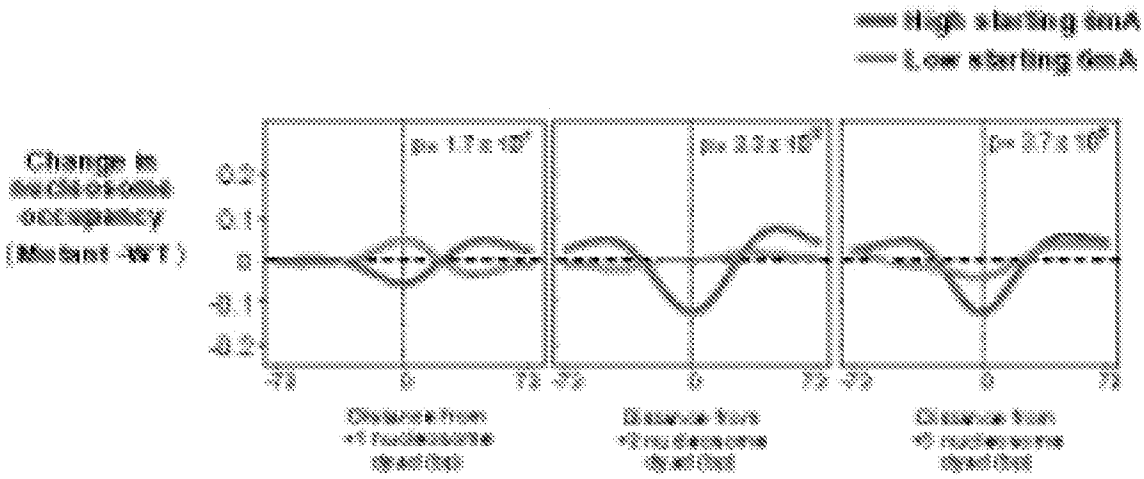


Fig. 14D



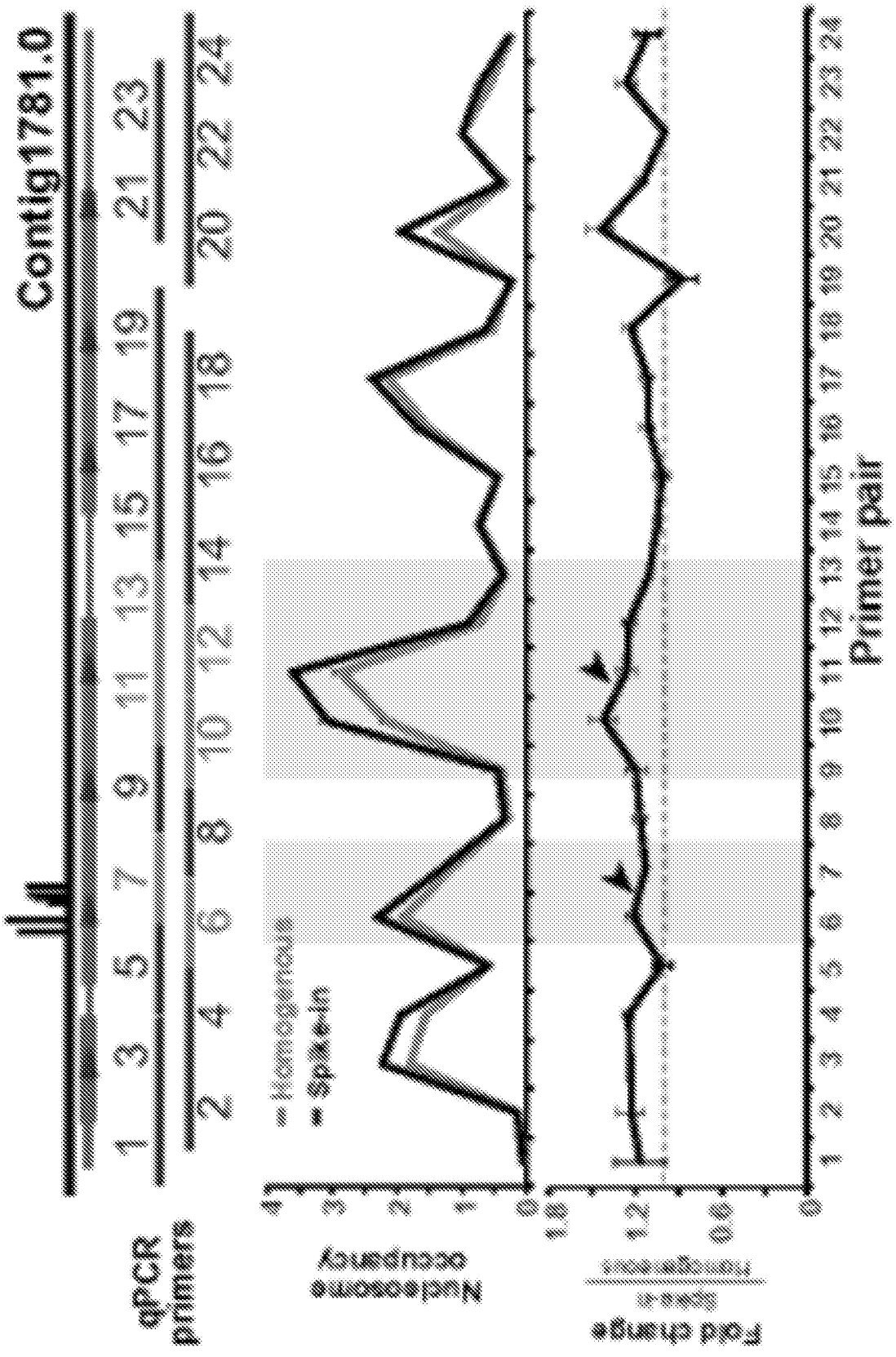

Contig1781.0
qPCR primers
1 3 5 7 9 11 13 15 17 19 21 23
2 4 6 8 10 12 14 16 18 20 22 24
= Homeologous
= Spike-in
Nucleosome occupancy
Fold change
Primer pair

Fig. 14E

Fig. 14F

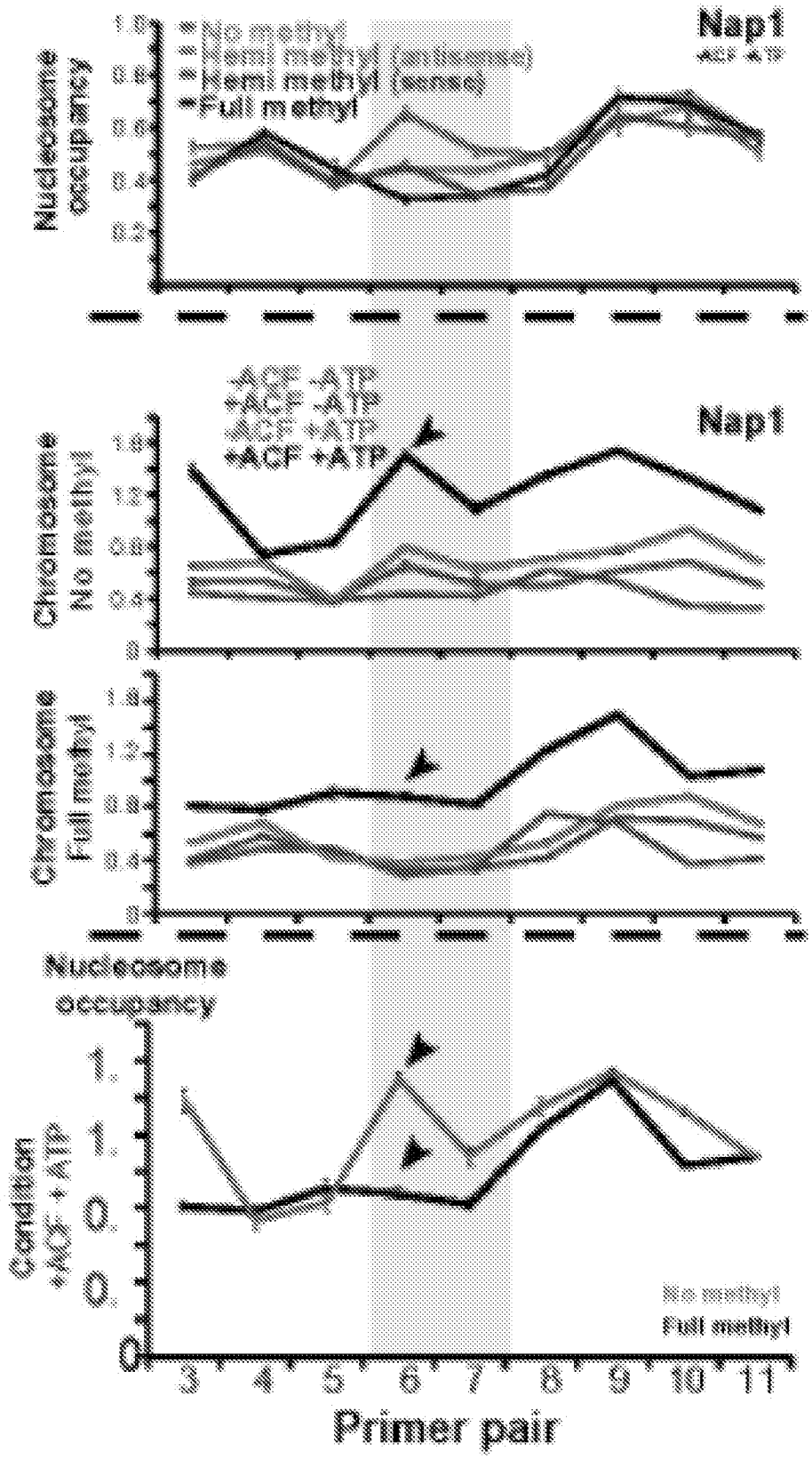
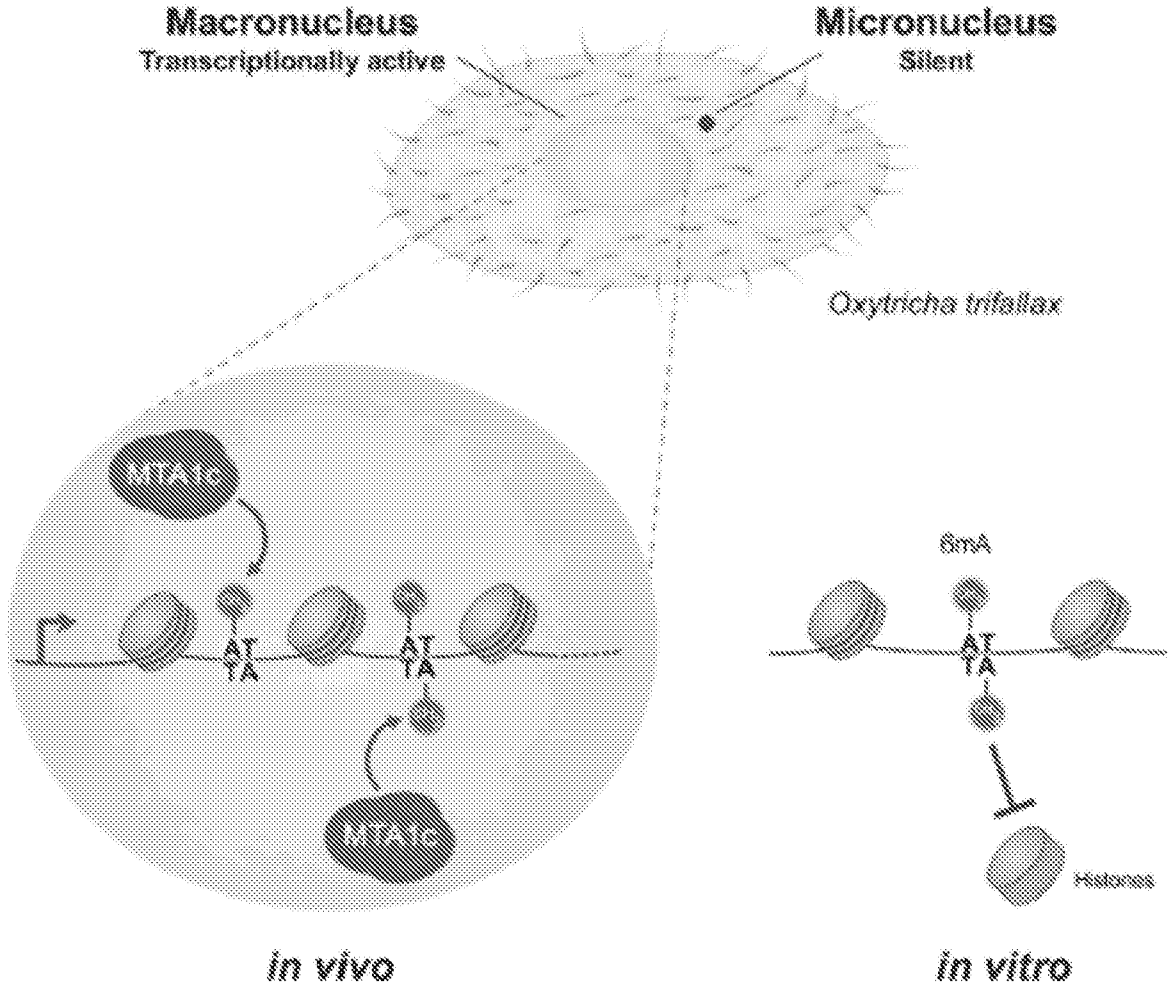

Nap1
Nucleosome occupancy
No methyl
Hemi methyl (antisense)
Hemi methyl (sense)
Full methyl
-ACF -ATP
+ACF -ATP
-ACF +ATP
+ACF +ATP
Chromosome No methyl
Chromosome Full methyl
Nucleosome occupancy
Condition +ACF +ATP
Full methyl
Primer pair

Fig. 15


Macronucleus
Transcriptionally active
Micronucleus
Silent
Oxytricha trifallax
MTA1c
AT
TA
MTA1c
SmA
AT
TA
Histones
in vivo
in vitro

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/42625

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G01N 33/68 (2019.01)

CPC - G01N 33/6848, G01N 33/6854, G01N 33/6857

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Tetrahymena Genome Database Wiki, "MTSA1 (MethylTransferase targeting position N-Six of Adenine (MTSA) 1)", 19 October 2017 (19.10.2017), retrieved on 19 September 2019 from https://www.google.com/search?q=inurl%3Ahttp%3A%2F%2Fciliate.org%2Findex.php%2Ffeature%2Fdetails%2FMTSA1&rlz=1C1SQJL_enUS861US861&oq=inurl%3Ahttp%3A%2F%2Fciliate.org%2Findex.php%2Ffeature%2Fdetails%2FMTSA1&aqs=chrome..69i57j69i58.8054j0j7&sourceid=chrome&ie=UTF-8&as_qdr=y15; entire document especially pg 1 para 1, pg 1 para 4	1-8
A	US 2008/0003209 A1 (Delack) 03 January 2008 (03.01.2008); entire document, especially abstract, [0055]	1-8
A	Xiao et al. ,"N6-Methyladenine DNA Modification in Human Genome", 16 August 2017 (16.08.2017), bioRxiv preprint doi: http://dx.doi.org/10.1101/176958; entire document, especially pg 3 summary, pg 5 para 2 pg 18 para 2, pg 31 last para - pg 32 para 1	1-8
A	Parashar, "N6-adenine DNA methylation demystified in eukaryotic genome: From biology to pathology" 24 October 2017 (24.10.2017); Biochimie 144 (2018) 56-62; entire document	1-8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 September 2019	27 NOV 2019
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/42625

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I: Claims 1-8, directed to a composition and method for modulating m6dA levels.

Group II: Claims 9-16, directed to a method of modifying a nucleic acid from a cell.

Group III: Claims 17-18, directed to a cell line obtained comprising a nucleic acid encoding MTA1c or any components thereof.

Group IV: Claims 19-20, directed to a method of identifying protein binding sites on DNA.

*****Continued in Supplemental Box*****

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-8

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US 19/42625

Continuation of Box III Observations where unity of invention is lacking

The inventions listed as Groups I, II, III, and IV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Group I requires using MTA1c or components thereof for modulating m6dA in a subject; not required by groups II, III, and IV.

Group II requires a method of modifying a nucleic acid from a cell, the cell derived from a multicellular eukaryote, comprising the steps of: (a) obtaining the nucleic acid from the cell; and (b) contacting the nucleic acid with MTA1c or any components thereof under conditions effective to methylate the nucleic acid; not required by groups I, III, and IV.

Group III requires a cell line obtained from a multicellular eukaryote comprising a nucleic acid encoding MTA1c or any components thereof and/or an MTA1c protein complex or any components thereof; not required by groups I, II, and IV.

Group IV requires a method of identifying protein binding sites on DNA comprising the steps of: (a) providing DNA; (b) contacting the DNA with MTA1c or any components thereof under conditions effective to methylate the DNA; (c) contacting the DNA with one or more proteins; (d) contacting the DNA with an enzyme effective to hydrolyze the DNA in positions where no protein binding occurs; (e) removing the DNA bound protein; and (f) isolating and sequencing the DNA fragments; not required by groups I, II, and III.

Common Technical Features:

Groups I, II, III, and IV share the technical feature of a component of MTA1c. However, these shared technical features do not represent a contribution over prior art, because the shared technical feature is being anticipated by document titled "MTSA1 (MethylTransferase targeting position N-Six of Adenine (MTSA) 1)" to Tetrahymena Genome Database Wiki (hereinafter "Tetrahymena-Database"). Tetrahymena-Database teaches a component of MTA1c (pg 1 para 1, "Standard Name... MTSA1 (MethylTransferase targeting position N-Six of Adenine (MTSA) 1)... Aliases... MTA1"; pg 2 table, "Source... Tetrahymena borealis"; see instant specification para [0007]-[0009]).

As the shared technical features were known in the art at the time of the invention, they cannot be considered common technical features that would otherwise unify the groups. Therefore, Groups I, II, III, and IV lack unity under PCT Rule 13.
Note:
Claims 1-20: "MTA1c" is unclear as presented. For the purpose of completing this ISR, "MTA1c" and components thereof are interpreted as defined in instant specification para [0007]-[0009].