Graduate School of Frontier Sciences, The University of Tokyo

Department of Computational Biology

2010-2011

Master's Thesis

Methylome analysis of *C. intestinalis* and the origin of tissue specific methylation

Submitted February, 2012

Adviser: Professor Takashi Ito

47-106907 Takao Yokoyama

Contents

Introduction

Methylation of cytosine at the 5-position, which I call "DNA methylation" in this thesis, plays various roles in the expression of genomic functions. Many phenomena, including repression of selfish genes like transposons, formation of heterochromatin, development, differentiation, cancer, X-chromosome inactivation and genomic imprinting in mammals, are known to be related to DNA methylation [1]. Although DNA methylation is observed in many organisms in various branches of the eukaryotic phylogenetic tree [2, 3], it is differently used in these organisms. Fungi such as *Neurospora crassa* [4] are known to selectively methylate repeat elements in the genome. On the other hand, it is known that gene body methylation is the sole DNA methylation observed in honey bee. For plants and vertebrates, methylation is found in both repeat elements and gene bodies.

Because transposon methylation is observed in all the four organisms in which DNA methylation is extensively studied, namely human [5], mouse [6], *Arabidopsis thaliana* [7, 8]and *N. crassa*, repression of selfish genes had been assumed as the ancient function of DNA methylation. However, recent studies revealed many exceptional cases to this notion. For instance, most, if not all, repeat sequences escape methylation in the tunicate *Ciona intestinalis* [9] and the silk moth *Bombyx mori* [10].

Gene body methylation is a recently recognized type of methylation, which was first described in *A. thaliana* [8]. While DNA methylation has been thought to play repressive roles in gene expression, gene body methylation levels were found to be positively correlated with the expression level [11]. These observations suggest that DNA methylation is related to not only repression but also induction or promotion of gene expression. After the first discovery of gene body methylation in *A. thaliana*, this relation has been confirmed in other organisms. Especially, whole genome bisulfite sequencing (WGBS) analysis conducted on 17 species belonging to diverse phyla [11] revealed that gene body methylation is conserved throughout plants and animals, leading to an assumption that this phenomenon is related to ancient properties of DNA methylation. While some studies propose potential roles of gene body methylation in the suppression of fortuitous transcriptional initiation from the inside of genes and the regulation

of RNA splicing, molecular mechanisms and biological significance of gene body methylation remain to be elucidated.

Regulation of transcription by tissue or cell-type specific differentially methylated region (tDMR) is one of the most intensively studied aspects of DNA methylation in mammals [12, 13, 14]. However, it is rather an exceptional phenomenon from the evolutionary point of view on DNA methylation. Because this phenomenon is observed only in vertebrate at present, it is thought to be a relatively new function that DNA methylation acquired after the separation of lineages leading to vertebrates and invertebrates. I am interested in the evolutionary origin of tDMRs that regulate gene expression in vertebrates and intended to know how distantly diverged organisms from the vertebrate have tDMRs. Vertebrate are thought to be divided from invertebrate 550 million years ago [15]. To investigate the origin of tDMR-mediated regulation of gene expression, comparative methylome analysis between various tissues of urochordata, or the closest invertebrate relative of the vertebrates, would be informative.

C. intestinalis belongs to urochordata and serves as one of the major experimental model organisms in developmental biology. Draft genome sequence of *C. intestinalis* was published in 2002 [15], which accelerates molecular genetic analysis with genome-scaled measurement. This organism has a much smaller genome (∼120Mb) [16] than vertebrates, which experienced two rounds of whole genome duplication [17], but shares almost all the fundamental gene sets with vertebrates.

There are some early studies on DNA methylation of *C. intestinalis*. Methylation in this organism is exclusively targeted to CpG dinucleotides and shows typical mosaic methylation pattern as frequently seen in invertebrates. Simmen et al. (1999) [18] showed that methylated genomic regions of *C. intestinalis* possess less CpG dinucleotides than unmethylated regions; methylation rate and CpG frequencies are inversely correlated. Suzuki et al. (2007) [9] used the local CpG frequencies to predict local methylation status and found that many expression units lie in the methylated regions. Suzuki et al. also performed a comparative methylation analysis among sperm, gastrula and adult muscle on randomly selected six genes and showed no difference in their methylation status. WGBS analysis conducted by Zemach et al. (2010) [**?**] on muscle tissue of *C. intestinalis* showed that methylation of repeat element was rarely

seen in this organism but gene body methylation was frequent, providing the first genome-scale methylome data on this organism. However, they used only a single tissue to grasp the outline of methylome of *C. intestinalis* and did not intend to detect tDMRs.

In this study, I tried to examine the presence of tDMRs in *C. intestinalis* through comparison of methylomes among four tissues using a recently emerged WGBS technology. Comparison of methylome data on the first two tissues analyzed so far, namely sperm and body-wall, revealed the presence of tDMRs in this organism for the first time. These tDMRs may represent a primordial form of tDMR-mediated regulation of gene expression.

Results

Whole genome bisulfite sequencing of *C. intestinalis*

In this study, four types of tissues, namely sperm, body wall, heart and gill, were selected for comparative methylome analysis. Since *C. intestinalis* was assumed to have no tDMRs, previous studies used sperm as a highly homogenous and easy-to-handle source for genomic DNA extraction. Thus, I selected sperm as a main analytical material for this study. Two types of muscle tissues were selected, one of which was smooth muscle from body wall and the other was striped muscle from heart. Both tissues are histologically different and expected to show differential gene expression. The forth tissue selected was gill, which is a characteristic structure that filtrates water and secretes mucus to capture foods in the water. [19] Among these four tissues, I report on the results on sperm and body wall in this thesis, because the amounts of genomic DNA extracted from these two tissues were much higher than those from the other two tissues and because I encountered a problem in template preparation.

For whole genome bisulfite sequencing (WGBS), I used Post-bisulfite adaptor tagging (PBAT) recently developed in our laboratory, because it requires much smaller amounts of DNA than conventional methods such as Shotgun Bisuflite Sequencing and MethylC-Seq. I first prepared WGBS templates from sperm and body wall because plenty of genomic DNA materials were available. On the other hand, since the amount of genomic DNA was quite limited for heart and gill, I postponed template preparation for these tissues, until the technical problems described below was settled.

During the preparation of WGBS templates, I encountered an unexpected problem in bisulfite treatment; an extremely low yield was specifically observed in bisulfite treatment of *C. intestinalis* genomic DNA. Similar problems were also observed in bisulfite treatment of genomic DNA from medaka (*Oryzias latipes*) and bluefish (*Trachurus japonicus*) for unknown reasons. Fortunately, I could circumvent the problem by replacing Imprint bisulfite DNA modification kit from Sigma with MethylCode bisulfite conversion kit from Invitrogen, again, for unknown reasons.

Deep sequencing was performed with Illumina Genome Analyzer IIx. Two lanes were

assigned for each sample to perform single-end sequencing with a read length of 120 nt. In total, 9.4G and 13.2G bases of reads were obtained for sperm and body wall, respectively. These reads were mapped to the *C. intestinalis* reference genome sequence with an averaged read depth of 17 for both samples (Table 1). About 90% of the genomic region was covered by more than 5-fold (Figure 1 and Table 1). In the following analysis, I used the data on these genomic regions.

Overview of DNA methylation in *C. intestinalis*

At first, I investigated how many cytosines were methylated in the *C. intestinalis* genome. As shown in Figure 2, more than 95% of cytosines showed methylation levels lower than 10%, indicating that most cytosines are unmethylated in this organism. On the other hand, about 4.1% and 3.6% of cytosines showed methylation levels higher than 50% in sperm and body wall, respectively. Also, the differences in highly methylated fractions between the two tissues suggested the presence of tDMRs. The WGBS analysis of muscle tissues reported by Zemach et al. mentioned that about 4.7% of cytosines are methylated, and our results are consistent with theirs.

Eukaryotic DNA methyltransferases (DNMTs) are classified into three subtypes based on their target sequences [3]. DNMTs targeting CpG are represented by MET1 from *A. thaliana* and DNMT1 from mammals. Those targeting CHG, where H is A, C or T, include CMT3 from *A. thaliana*, whereas those targeting CHH include DRM2 from *A. thaliana*. Thus, I investigated the sequence context of methylated cytosines found in *C. intestinalis*. I found that both tissues, sperm and body wall, showed very low methylation level for CHG and CHH (Figure 3). In contrast, a substantial population of CpG showed high methylation level (Figure 3). As sperm had more highly-methylated CpG's than body wall, it is likely that the origin of differences observed in Figure 2 is attributed to the differences of methylation rate in these CpG dinucleotides.

Next, from a converse point of view, I investigated the sequence context of highly-methylated cytosines. As shown in Figure 4, more than 97% of cytosines with >50% methylation levels were CpG dinucleotides. These results not only extended the classical observation that CpG dinucleotides are the sole target of DNA methylation in *C. intestinalis* to a genome-wide

scale but also suggested the presence of tDMRs between the two tissues.

Relationship between DNA methylation and CpG score

Deamination of cytosines is thought to occur frequently in living cells. Cytosines are deaminated to produce uracil, whereas 5-methyl-cytosines are converted to thymine with the same reaction. Products of the former reaction or uracils are readily recognized and removed by uracil DNA glycosylase, whereas those of the latter reaction are repaired by thymine DNA glycosylase which recognizes GT mismatches [20]. For unknown reasons, the latter process seems to be less effective than the former process, leading to preferential loss of CpG dinucleotides during evolution. Accordingly, methylated genomic regions tend to show lower CpG dinucleotide frequencies than unmethylated regions. The "CpG score" is defined as a ratio between the observed CpG frequencies and those expected from local GC content and has been used to predict methylated regions. Previous studies reported that the CpG score shows a bimodal distribution in the *C. intestinalis* genome (Figure 5). Simmen et al. revealed that CpG score is inversely correlated with the methylation level; highly methylated genomic regions showed low CpG scores, whereas unmethylated regions showed high CpG scores. Based on these results, Suzuki et al. predicted methylated regions from a 2-Mb genomic segment of *C. intestinalis*, and showed that those predictions were generally consistent with actual methylation status determined by a methylation-sensitive restriction enzyme-based assay and targeted bisulfite sequencing.

I tried to extend a similar analysis performed by Suzuki et al. to the whole-genome scale. First, I used the CpG score as a predictor to extract 25,763 "predicted methylated domains". In total, those regions account for 43.7% (50.4 Mb) of the *C. intestinalis* genome (112.5 Mb). I examined the distribution of the length of predicted methylated domain but failed to find any characteristic patterns (Figure 6).

Next I performed a genome-wide comparison between the CpG score and the methylation level determined by the WGBS data. As shown in Figure 8, those two values showed an apparent strong correlation. However, at the same time, I found a small but sizable fraction of the genome fails to show the expected correlation between the CpG score and the actual methylation level

(Figure 7). For example, predicted methylated domains, which are defined as the regions with CpG score smaller than 0.8, contained unmethylated regions showing methylation levels smaller than 10%. Such unmethylated regions comprised 3.23% and 3.18% of the genome, or 7.39% and 7.28% of the predicted methylation domain, for sperm and body wall, respectively. Conversely, 5.9% and 6.3% of the genome, or 10.5% and 11.2% of the predicted "unmethylated" regions turned out to be highly methylated (>50%) in sperm and body wall, respectively (Figure 9). These results underscored the importance of experimental determination of methylation levels.

Identification of tissue-specific methylated region

Patterns of DNA methylation in *C. intestinalis* are assumed to be the same among all tissues. This assumption is based on the observations that several genomic regions individually investigated showed no differential methylation among the tissues examined [9]. However, since no data are available for comparative methylome analysis among tissues in this organism, I wondered whether the *C. intestinalis* genome indeed lacks tDMRs. To answer this question, I compared the two WGBS data on sperm and body wall. As shown in Figure 8, almost all the genomic regions did show similar methylation levels between the two tissues. However, as was expected from the difference of overall methylation pattern between sperm and tissues (Figure 2 and 3), a sizable fraction of cytosine residues showed different methylation levels between the two tissues. Next I tried to calculate the number of cytosines with statistically significant difference in methylation level between the two tissues. Using Fisher exact test to examine the statistical significance, I extracted cytosines from 14 chromosomes that were differentially methylated between the two tissues with a p-value equal or smaller than 0.05. I found that 151,526 out of 2,067,759 cytosine residues in CpG dinucleotides (7.3%) showed statistically significant differences in methylation levels. These residues were defined as differentially methylated CpG (DmCpG).

Next, I tried to extract tDMRs between the two tissues. Previously, Lister et al. used WGBS data on two human cells to extract tDMRs as a region that has 4 or more DmCpGs in 1000-bp window. Since the frequency of CpG dinucleotides differs drastically from one window to another window in *C. intestinalis*, I introduced a filtering scheme for CpG frequency. I

extracted tDMRs as 500-bp windows in which 1) more than 60% of the CpG dinucleotides were judged as DmCpG's and 2) the CpG frequency is larger than 0.6. As shown in Figure 10, we found numerous tDMRs are found in each *C. intestinalis* chromosomes. These tDMRs include promoters showing sperm-specific methylation (Figure 14 and 13) as well as sperm-specific demethylation of the body of the gene encoding adult-specific subtype of myosin (Figure 11 and 12). These observations demonstrated the presence tDMRs as well as their potential involvement/association with transcriptional activities in *C. intestinalis*.

Discussion

I conceived a whole-genome comparative methylome analysis among four tissues of sea squirt *C. intestinalis*. Sperm, body wall, heart and gill, were selected for the analysis, and the data on the first two tissues were presented in this thesis. Thus, the goal of my research has not been attained at present, but even the comparison between the two methylome data brought new observations. As the methylome data on the other two tissues were recently obtained, I expect to extend the comparative methylome analysis to the four tissues in the near future.

Prediction of methylated region by CpG score is not always correct

Prediction of methylated region by CpG score is based on a model that 5-methyl-cytosines in CpG dinucleotides are gradually converted to thymines during the course of evolution, leading to low CpG frequencies in highly methylated genomic regions. This explanation is persuasive and CpG score can indeed predict most methylated regions (Figure 8). However, we found that CpG score fails to predict the methylation status of a sizable fraction of genomic regions (Figure 7). These regions include those actually methylated but predicted to be unmethylated as well as those unmethylated but predicted to be methylated (Figure 9).

Genomic regions predicted to be unmethylated but actually methylated can be explained by the evolutionary model on CpG depletion by taking the evolutionary time into account. In order for a genomic region to show a low CpG score, the region has to be methylated in the germ line for a long evolutionary time. It is conceivable that a gene or genomic region that acquires its "methylatability" in a relatively recent past has not spent sufficient time to lose its CpGs, thereby retaining high CpG score. To evaluate this scenario, comparative genomic analysis to investigate evolutionary age of such regions would be required.

On the other hand, I have no idea to explain the presence of genomic regions that were predicted to be methylated but actually unmethylated at present. Of course, it is conceivable that these regions are tDMRs that are unmethylated in both sperm and body wall but methylated in some other tissues. Further investigation of the mechanisms to maintain methylated and unmethylated regions has to be conducted to explain differences of these regions.

Identification tDMRs in urochordate

It is well known that methylation of cytosines, if occurred at regulatory region of gene expression, causes gene repression, and in this case, methylation rates and expression levels correlate inversely. Methylation of promoter region like this is well documented but limited only in vertebrates. With evolutionary comparison of DNA methylation among 17 species, Zemach et al. revealed that repression of repeat elements and gene body methylation are the most fundamental functions of eukaryotic DNA methylation. On the contrary, it is accepted that regulation of gene expression by differentially methylated promoters is an evolutionarily new role of DNA methylation. Zemach et al. described WGBS data on muscle tissues of *C. intestinalis* and concluded that DNA methylation is not directed to repeat elements and promoter regions but that gene body methylation is the main target of DNA methylation in this organism. Based on the previous studies that compared DNA methylation status of a small number of genes among several tissues, DNA methylation pattern of *C. intestinalis* has been assumed to be identical among the tissues, retaining the primitive mode of eukaryotic DNA methylation.

In this thesis, I presented comparative methylome analysis between sperm and body wall and revealed, for the first time, the presence of many tDMRs in *C. intestinalis*. These tDMRs include differential gene body methylation of adult-type myosin gene as well as sperm-specific promoter methylation. Since the levels of gene body methylation were shown to positively correlate with those of gene expression, the former tDMR or gene body tDMR could be explained with tissue-specific differential expression of the gene. On the other hand, the latter tDMR resembles promoter-specific methylation frequently found in vertebrates. This may lead to the first discovery of differential methylation-mediated regulation of gene expression in invertebrates. To pursue the possibility, I have to examine not only the promoter structure but the expression of the downstream gene. It is quite important for me to integrate the methylome data with other omics data such as TSS-Seq, RNA-Seq and ChIP-Seq data for further investigation of tDMRs.

Methods

Genomic DNA

Genomic DNAs from sperm, body wall, heart and gill of *C. intestinalis* were generous gifts from Takehiro Kusakabe (Konan University).

Quantitation of DNA

The concentration of sample DNA was determined by Quant-iT dsDNA BR Assay Kit with Qubit fluorometer (Invitrogen) as described in manufacturer's instructions.

Preparation of sequencing template

Bisulfite conversion

Bisulfite treatment of genomic DNA was performed with Imprint DNA modification kit (Sigma) according to manufacturer's instruction except for column purification steps. PureLink PCR purification kit (Invitrogen) was used for the purification, because it allows higher yields of DNA than the column in the Imprint kit. Ten μl solution containing sample DNA (100 ng) was combined with 110 μl of Imprint DNA modification buffer, and incubated at 99°C for 5 min and 65°C for 90 min. Then, 480 μl of PureLink Binding Buffer and 1 μg of yeast DNA-free carrier RNA were supplemented, and the mixture was loaded on PureLink column to capture DNA. The column was washed with Wash Buffer W1, and then 90% ethanol supplemented with balance solution was loaded on the column. After incubation at room temperature for 15 min, the column was washed with Wash Buffer W1 twice, and the DNA was eluted with 20 μl of 10 mM Tris-HCl, pH 8.5.

First strand synthesis

To the 20 μl of bisulfite treated DNA, 5 μl of 10x NEBuffer 2 (NEB), 5 μl of 2.5 mM dNTPs and 4 μl of 100 μM first strand primer (Table 2) were supplemented and the total volume was adjusted to 50 μl with water. Then, the solution was incubationed at 95°C for 5min and 4°C for 5min. Polymerization was started by adding 50 units of Klenow fragment exo(-). The

reaction was incubated at 4°C for 15 min and then subjected to gradual increment of reaction temperature to 37°C at a rate of 1°C/min, incubation at 37°C for 90 min and 70°C for 10 min. To the reaction, 50 μl of AMPure XP was added and incubated at room temperature for 10 min. After washing the AMPure beads with 75% ethanol, DNA was eluted with 50 μl of 10 mM Tris-acetate (pH8.0).

Second strand synthesis

Twenty μl of Dynabeads M280 streptavidin were washed twice with 200 μl of 2x BW buffer (10 mM Tris-HCl, pH7.5, 1 mM EDTA, 3 M LiCl) and dissolved in 50 μl of 2x BW buffer. Then, the beads suspension was mixed with purified product of 1st strand synthesis. Biotinylated DNA was captured on the magnetic beads with rotating at room temperature for 30 min. The magnetic beads were washed twice with 2x BW buffer, twice with 0.1 N NaOH, once with 2x BW buffer and once with 10 mM Tris-HCl (pH7.5). For the first washing step with 0.1 N NaOH solution, incubation at room temperature for 2 min was included. Fifty μl solution containing of 5 μl of 10x NEBuffer 2, 5 μl of 2.5 mM dNTPs and 4 μl of 100 μM 2nd strand primer (Table 2) were prepared and added to the washed beads, and then incubation at 95°C for 5 min and 4°C for 5min were performed. Polymerization was started by adding 50 units of Klenow fragment exo(-). The reaction was incubated at 4°C for 15 min and then subjected to gradual increment of reaction temperature to 37°C at a rate of 1°C/min, incubation at 37°C for 30 min and 70°C for 10 min. Supernatant was removed from the beads, and the beads were again suspended in 50 μl solution containing 1x ThermoPol buffer (NEB) and 250 μM dNTPs. Polymerization was started with adding 8 units of Bst DNA polymerase large fragment (NEB) and executed with incubation at 50°C for 30 min.

Elution of template and size fractionation

Supernatant of 2nd strand synthesis was removed and the beads were resuspended in 50 μl solution containing 1x Phusion HS Buffer (Finzyme), 250 μM dNTPs and 200 nM elution primer (Table 2). After 2 unit of Phusion Hot Start High-Fidelity DNA Polymerase (Finzyme) was supplemented, elution was performed with incubating at 94°C for 5 min, 55°C for 10 min

and 72°C for 30 min. The supernatant was moved to a new tube, supplemented with 1 unit of exonuclease I and incubated at 37°C for 15 min. After killing the enzyme by incubating at 70°C for 10 min, 50 μl of AMPure XP was added and incubated at room temperature for 10 min. After washing the AMPure beads with 75% ethanol, DNA was eluted with 20 μl of 10 mM Tris-acetate (pH8.0).

Quantitation of template concentration

To calculate effective concentration of template molecules, quantitation with real-time PCR were performed. PhiX control (Illumina) was used as a quantitation standard. Real-time PCR assay was executed with SYBR Premix ExTaq kit (Takara) and Applied Biosystems 7000 real-time PCR system.

Sequencing

Sequencing was conducted by Kyushu University Genome Analysis Consortium. Each sample was single-end sequenced in two lanes (read length is 120 nt).

Mapping and visualization

Mapping and visualization of WGBS data were performed with a pipeline developed in our laboratory. Genomic sequence (Joined-scaffold, KH) and gene annotations (KH gene models, ver.2012) for *C. intestinalis* were downloaded from Ghost Database at Kyoto University (http://ghost.zool.kyoto-u.ac.jp/indexr1_kh.html).

Identification of differentially methylated CpGs

Fisher's exact text was employed to examine statistical significance of difference in methylation level for individual CpG dinucleotides. The test was performed on R platform using 2x2 contingency table filled with the numbers of reads that support methylated and unmethylated status of each cytosine from the two tissues to be compared.

References

[1] Bird, A. (January, 2002) DNA methylation patterns and epigenetic memory. *Genes & Development,* **16**(1), 6–21.

[2] Suzuki, M. M. and Bird, A. (June, 2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics,* **9**(6), 465–476.

[3] Zemach, A. and Zilberman, D. (September, 2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Current biology : CB,* **20**(17), R780–5.

[4] Rountree, M. R. and Selker, E. U. (April, 2010) DNA methylation and the formation of heterochromatin in Neurospora crassa. *Heredity,* **105**(1), 38–44.

[5] Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (October, 2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* **462**(7271), 315–322.

[6] Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., and Lander, E. S. (July, 2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature,*.

[7] Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (May, 2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell,* **133**(3), 523–536.

[8] Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (February, 2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature,* **452**(7184), 215–219.

[9] Suzuki, M. M., Kerr, A. R. W., De Sousa, D., and Bird, A. (May, 2007) CpG methylation is targeted to transcription units in an invertebrate genome.. *Genome Research,* **17**(5), 625–631.

[10] Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., Zhang, H., Zhang, G., Li, D., Dong, Y., Zhao, L., Lin, Y., Cheng, D., Yu, J., Sun, J., Zhou, X., Ma, K., He, Y., Zhao, Y., Guo, S., Ye, M., Guo, G., Li, Y., Li, R., Zhang, X., Ma, L., Kristiansen, K., Guo, Q., Jiang, J., Beck, S., Xia, Q., Wang, W., and Wang, J. (May, 2010) Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology,* **28**(5), 516–520.

[11] Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (May, 2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation.. *Science (New York, NY),* **328**(5980), 916–919.

[12] Song, F., Smith, J. F., Kimura, M. T., Morrow, A. D., Matsuyama, T., Nagase, H., and Held, W. A. (March, 2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proceedings of the National Academy of Sciences of the United States of America,* **102**(9), 3336–3341.

[13] Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., and Beck, S. (2004) DNA Methylation Profiling of the Human Major Histocompatibility Complex: A Pilot Study for the Human Epigenome Project. *PLoS Biology,* **2**(12), e405.

[14] Rakyan, V. K., Down, T. A., Thorne, N. P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E. M., Backdahl, L., Johnson, N., Herberth, M., Howe, K. L., Jackson, D. K., Miretti, M. M., Fiegler, H., Marioni, J. C., Birney, E., Hubbard, T. J. P., Carter, N. P., Tavare, S., and Beck, S. (July, 2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Research,* **18**(9), 1518–1529.

[15] Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., de Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E. M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lemaire, P., Martinez, D., Meinertzhagen, I. A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Saiga, H., Satake, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-Bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D. N., Haga, S., Hayashi, H., Hino, K., Imai, K. S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K. W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M. M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P. D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., and Rokhsar, D. S. (December, 2002) The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.. *Science (New York, NY),* **298**(5601), 2157–2167.

[16] Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J., Wasserscheid, J., Dewar, K., Wiley, G. B., Macmil, S. L., Roe, B. A., Zeller, R. W., Hastings, K. E. M., Lemaire, P., Lindquist, E., Endo, T., Hotta, K., and Inaba, K. (2008) Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations.. *Genome Biology,* **9**(10), R152.

[17] Ohno, S. (1970) Evolution by Gene Duplication, Allen Unwin, City.

[18] Simmen, M. W. (February, 1999) Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome. *Science (New York, NY),* **283**(5405), 1164–1167.

[19] Corbo, J. C., Di Gregorio, A., and Levine, M. (September, 2001) The ascidian as a model organism in developmental and evolutionary biology.. *Cell,* **106**(5), 535–538.

[20] Wu, S. C. and Zhang, Y. (August, 2010) Active DNA demethylation: many roads lead to Rome. *Nature Publishing Group,* **11**(9), 607–620.

Samples	Base	total reads	uniquely mapped	unmapped	depth	coverage	(strand specific)
Sperm	9.4 Gb	7.8×10^7 (78 M)	51.8%	34.8%	×17	89.2%	80.8%
BodyWall	13.2 Gb	10.9×10^7 (109 M)	35.9%	54.7%	×17	90.8%	83.2%

Table 1: Mapping results

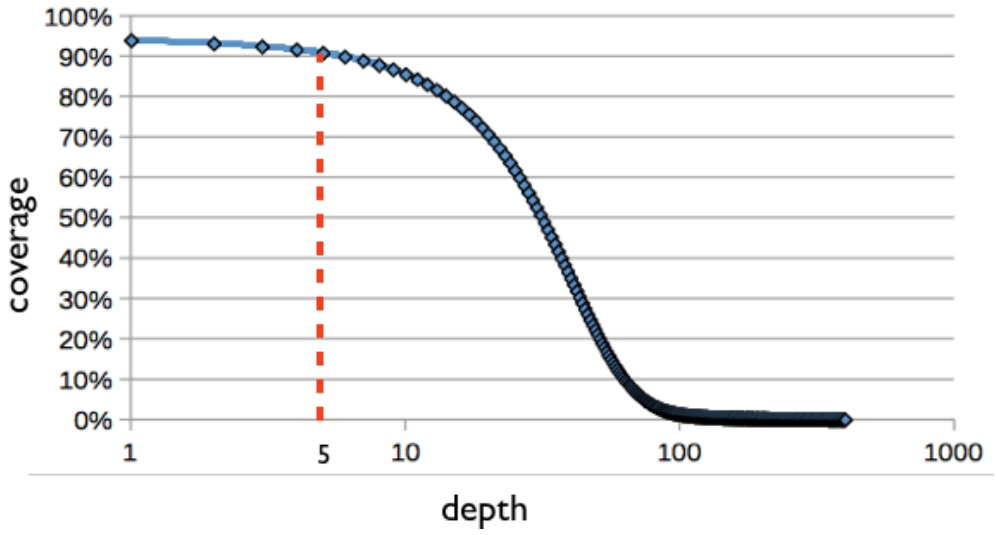

Figure 1: **Read depth and coverage**
Fraction of the reference genome covered by sperm WGBS data (vertical axis) was plotted against the minimal depth of the reads (horizontal axis). About 90% of genomic regions were covered by 5 or more reads (indicated red dashed line).

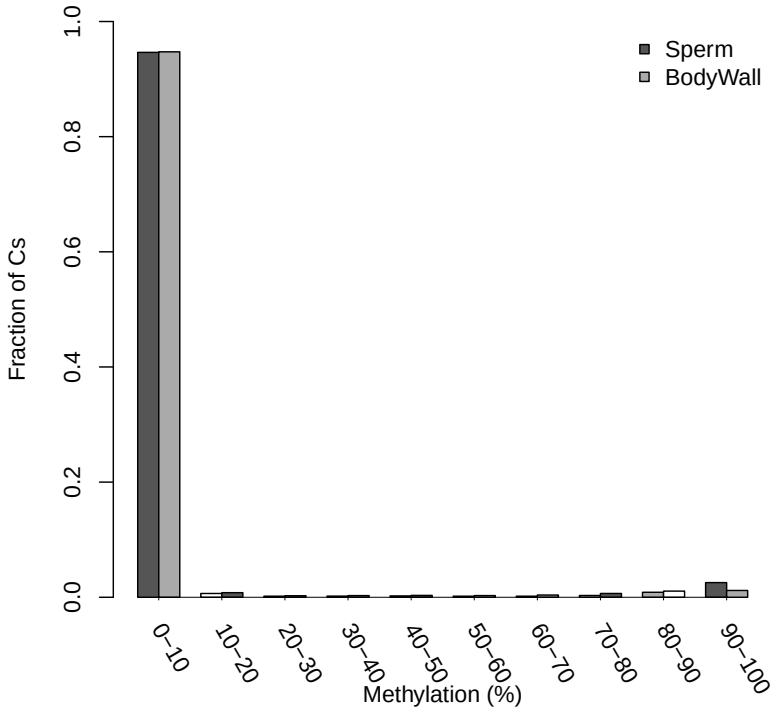

Figure 2: **About 4% of cytosines are methylated in C. intestinalis**
Distribution of methylation levels was shown for sperm and body wall. More than 95% of cytosine residues showed lower methylation rate (<10%). Highly methylated cytosines (>50%) in sperm and body wall were accounted for 4.1% and 3.6% of all cytosines, respectively.

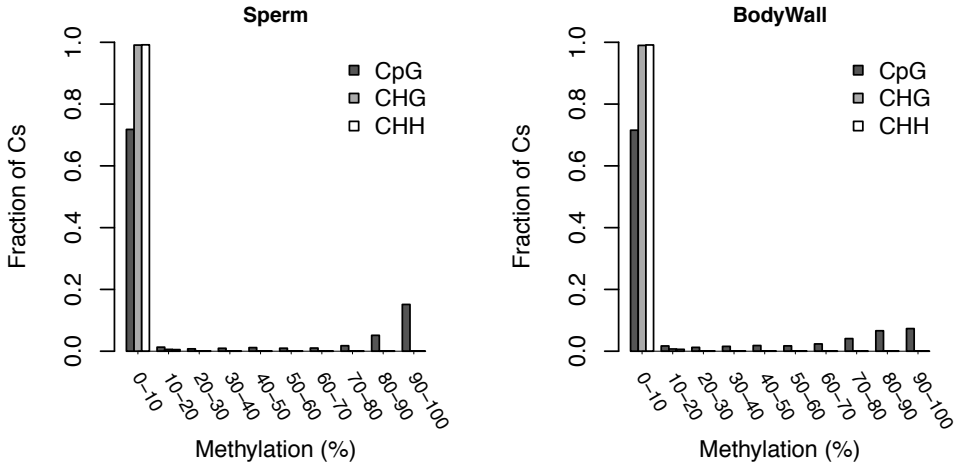

Figure 3: **Heavily methylated cytosines are observed in CpGs**
Distribution of methylation rates according to the sequence context of the cytosines, namely CG, CHG, and CHH (H = A, C, or T), are shown for sperm (left) and body wall (right). While almost all cytosines showed lower (<10%) methylation rate for CHG and CHH, highly methylated (>50%) cytosines were observed in CpGs. Highly methylated cytosines were accounted for 24.0% and 22.1% of cytosines in the context of CpGs in sperm and body wall, respectively.

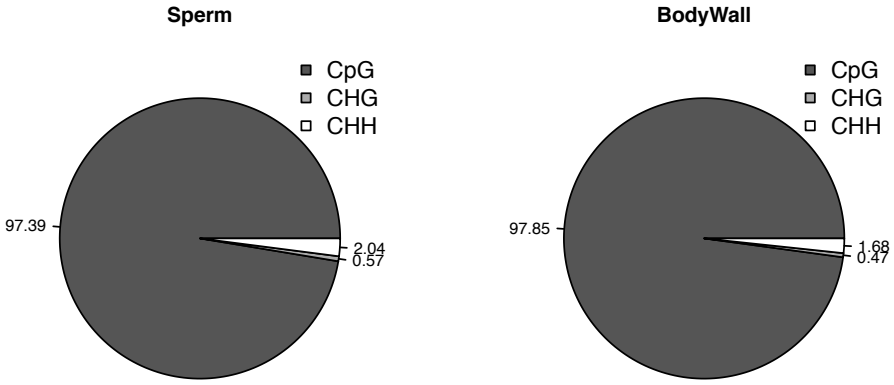

Figure 4: **Almost of all methylated cytosines are found in CpG dinucleotides**
Highly methylated (>50%) cytosines were classified with its sequence context, namely CG, CHG and CHH. More than 97% of methylated cytosines are classified into CpGs.

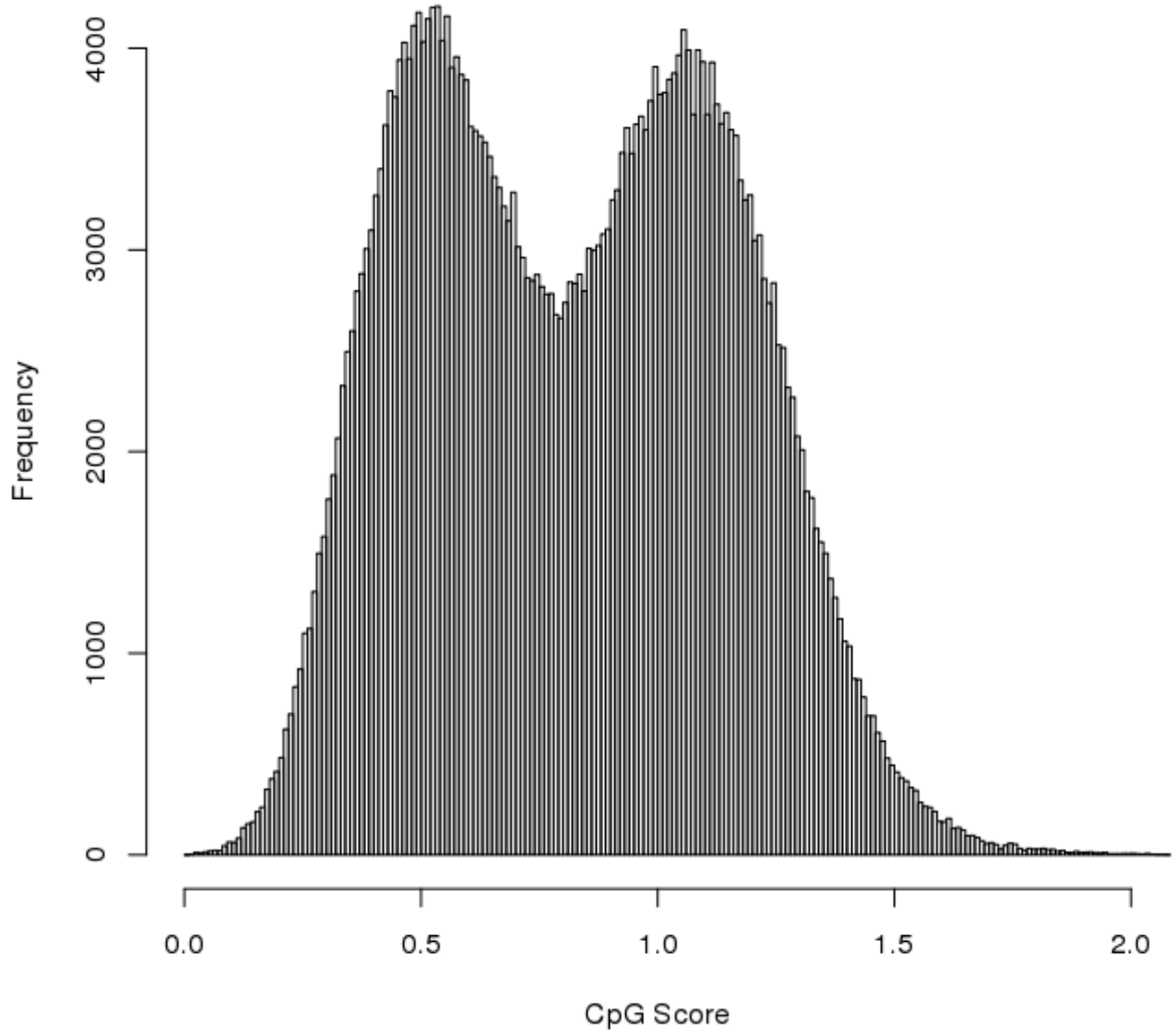

Figure 5: **Bimodal distribution of CpG score for *C. intestinalis* genome**
Distribution of CpG scores (horizontal), a ratio of CpG frequency observed per expected from local GC content, for the 1,000-bp sliding windows with 200-bp interval were shown.

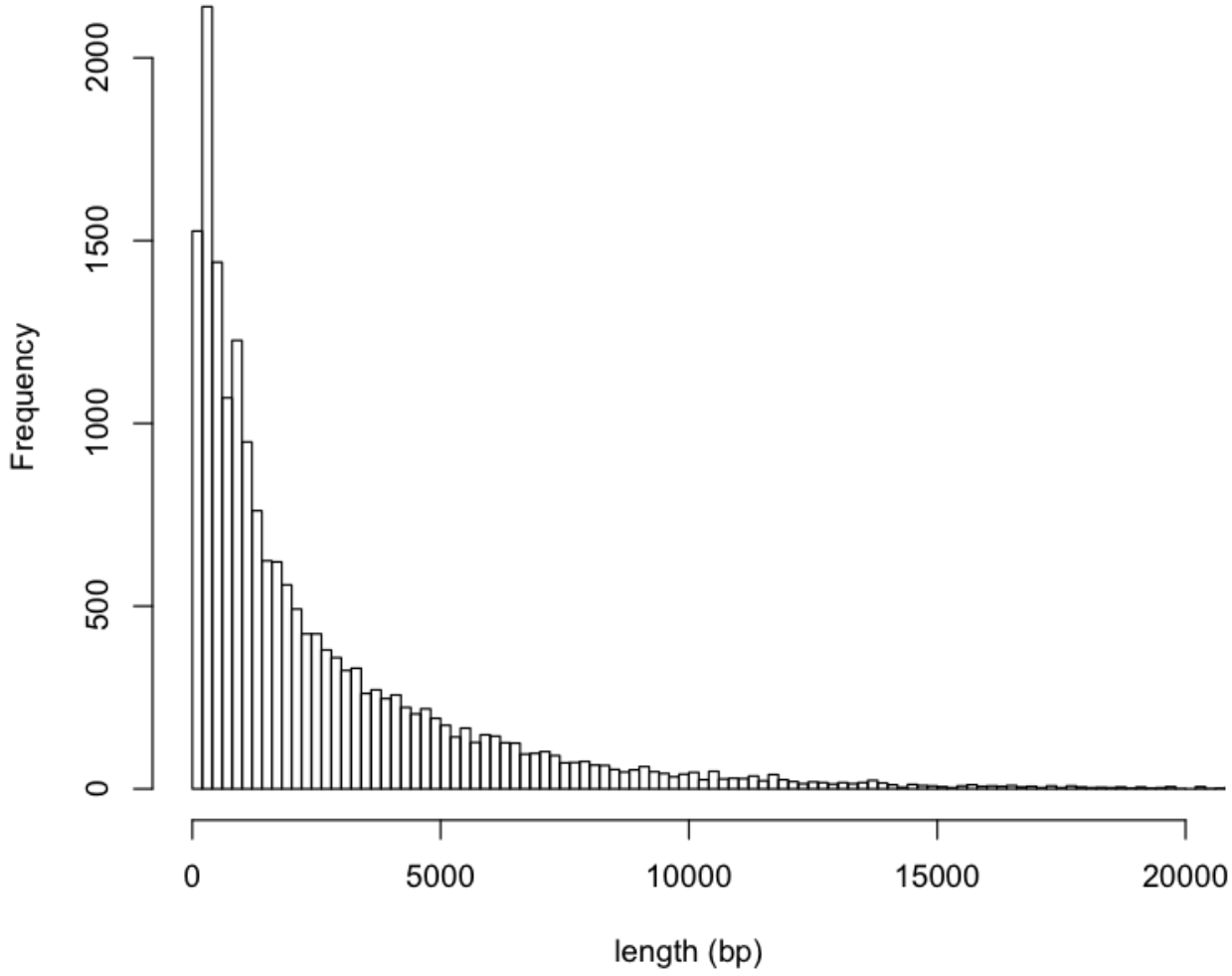

Figure 6: **Size distribution of "predicted methylated domain"**
CpG scores for the 1,000-bp sliding windows with 200-bp interval were calculated. Windows with CpG score less than 0.8 were defined as "predicted methylated domain" and successive windows were fused to obtain a single domain.

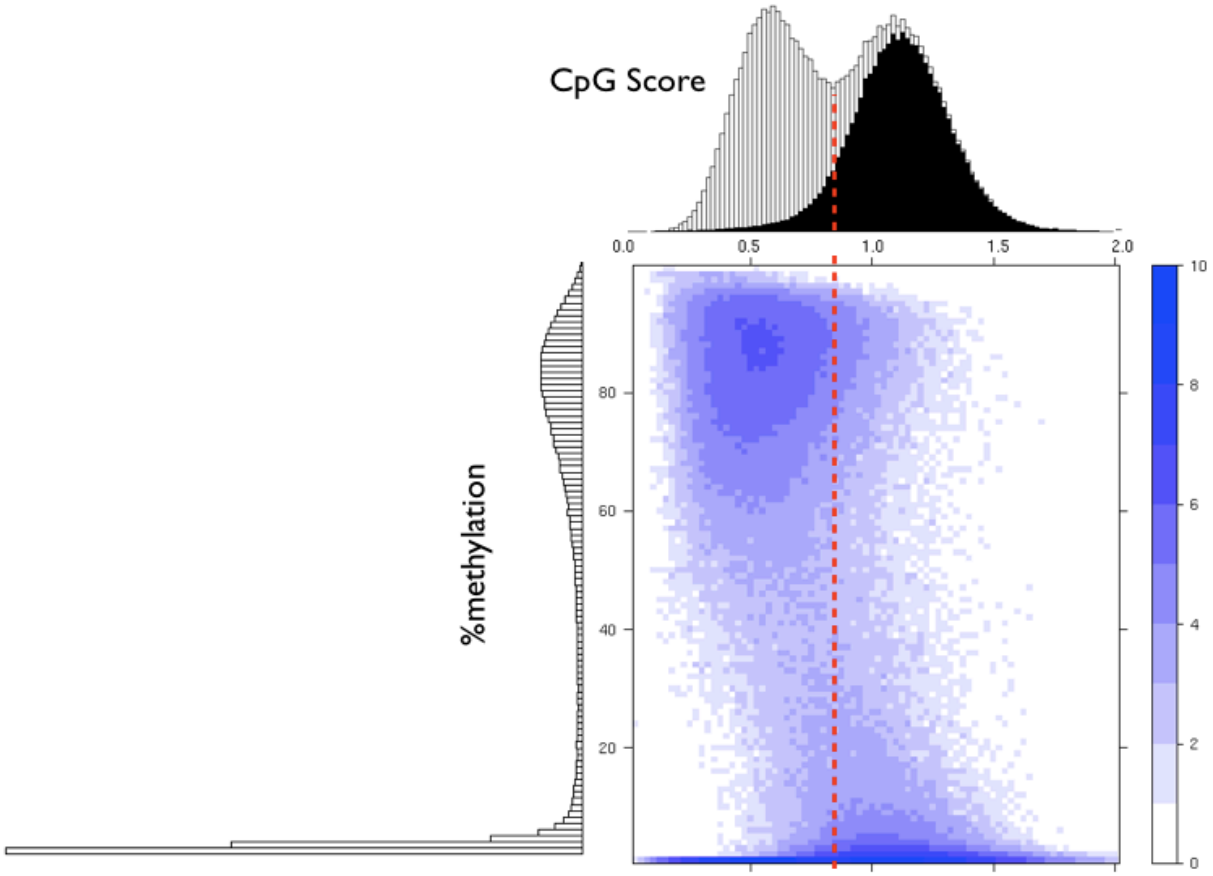

Figure 7: **Comparison between CpG score and observed methylation rate.** CpG score (horizontal axis) and observed methylation level (vertical axis) are plotted for 1,000-bp sliding windows with 200-bp interval. Frequencies of windows are expressed with a 10-color gradient from white to blue. Distribution of CpG score (up) and observed methylation rate (left) are also shown. Red dashed line; CpG score = 0.8.

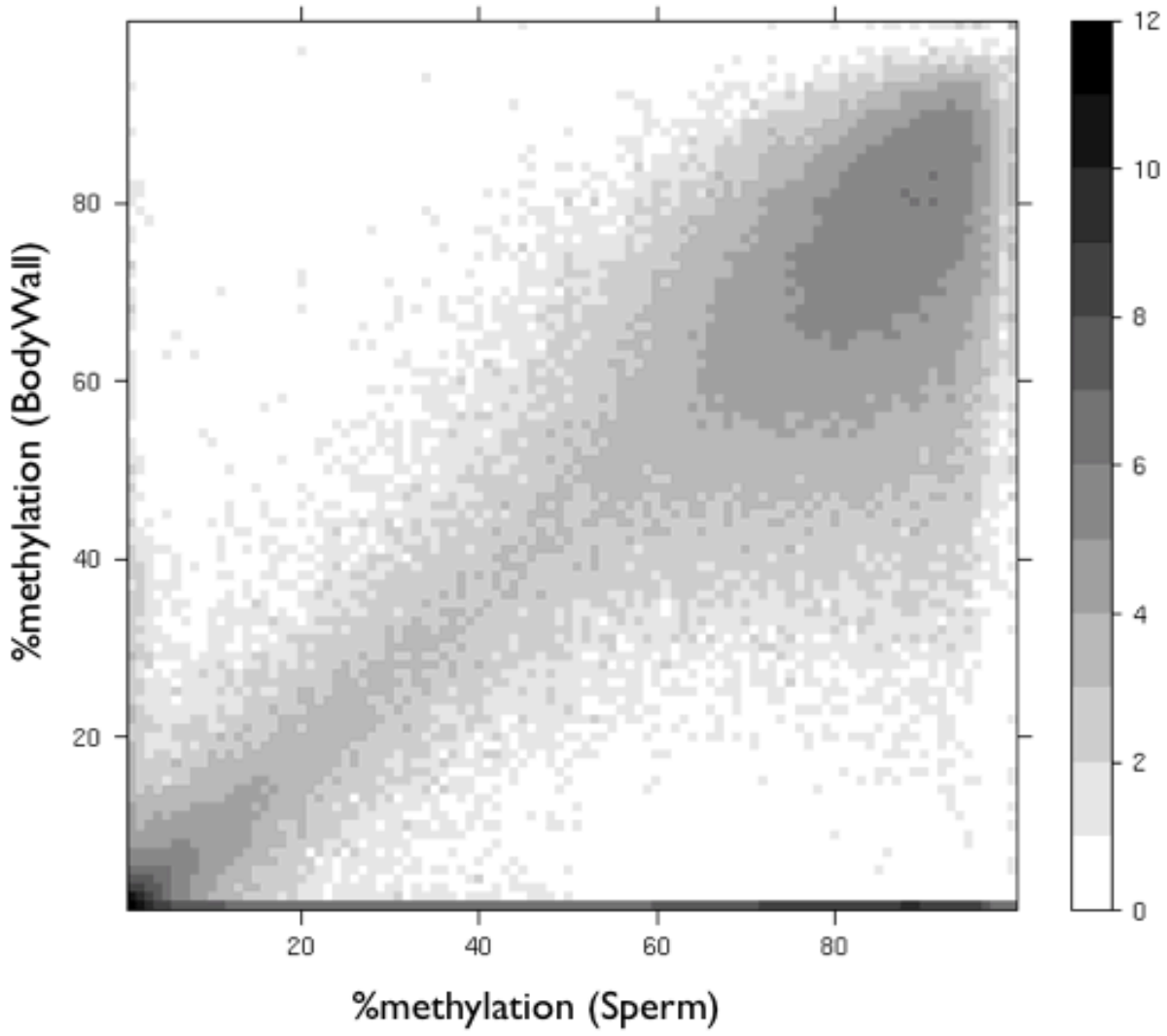

Figure 8: **Comparative methylome analysis between sperm and body wall**
Observed methylation level for sperm (horizontal axis) and body wall (vertical axis) were plotted for 1,000-bp sliding windows with 200-bp interval. While most residues showed consistent methylation rate between two tissues, sizable amount of residues showed different methylation rate.

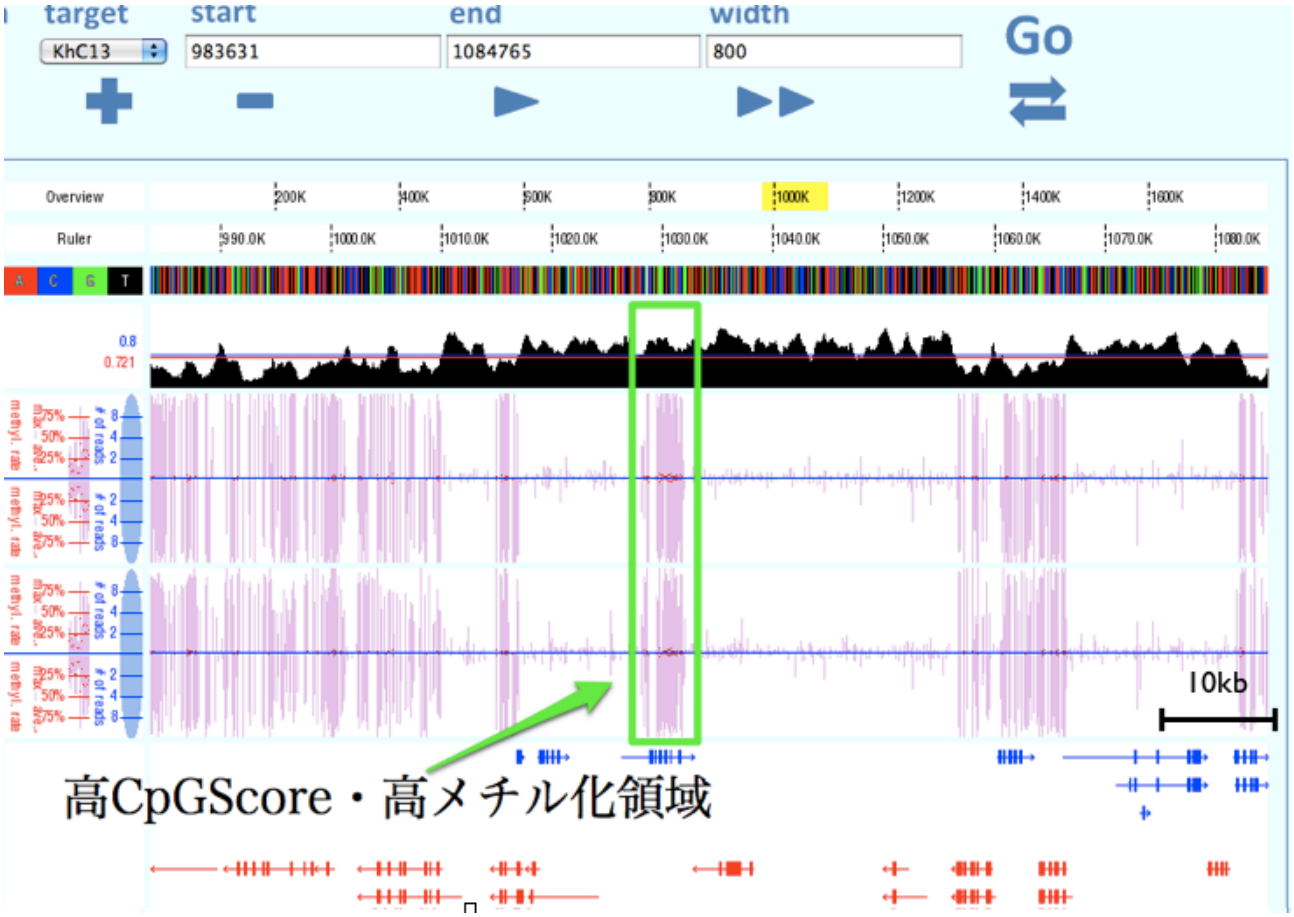

Figure 9: **An example of highly methylated "predicted unmethylated region"** Screen shot of genome browser showing a region predicted not to be methylated but heavily methylated actually (indicated in green box).

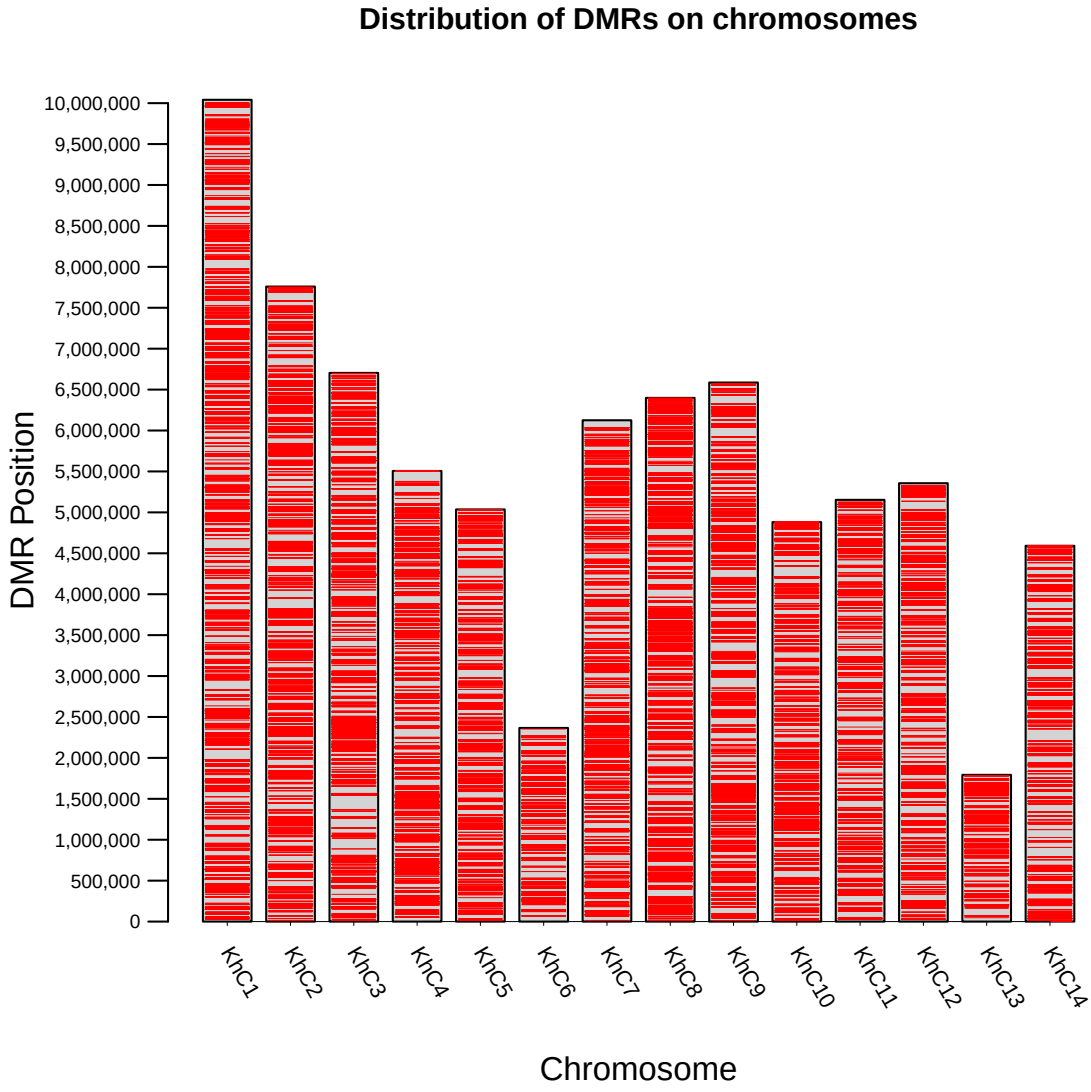

Figure 10: **Distribution of tDMRs on chromosomes**Red bar indicates the tDMRs.

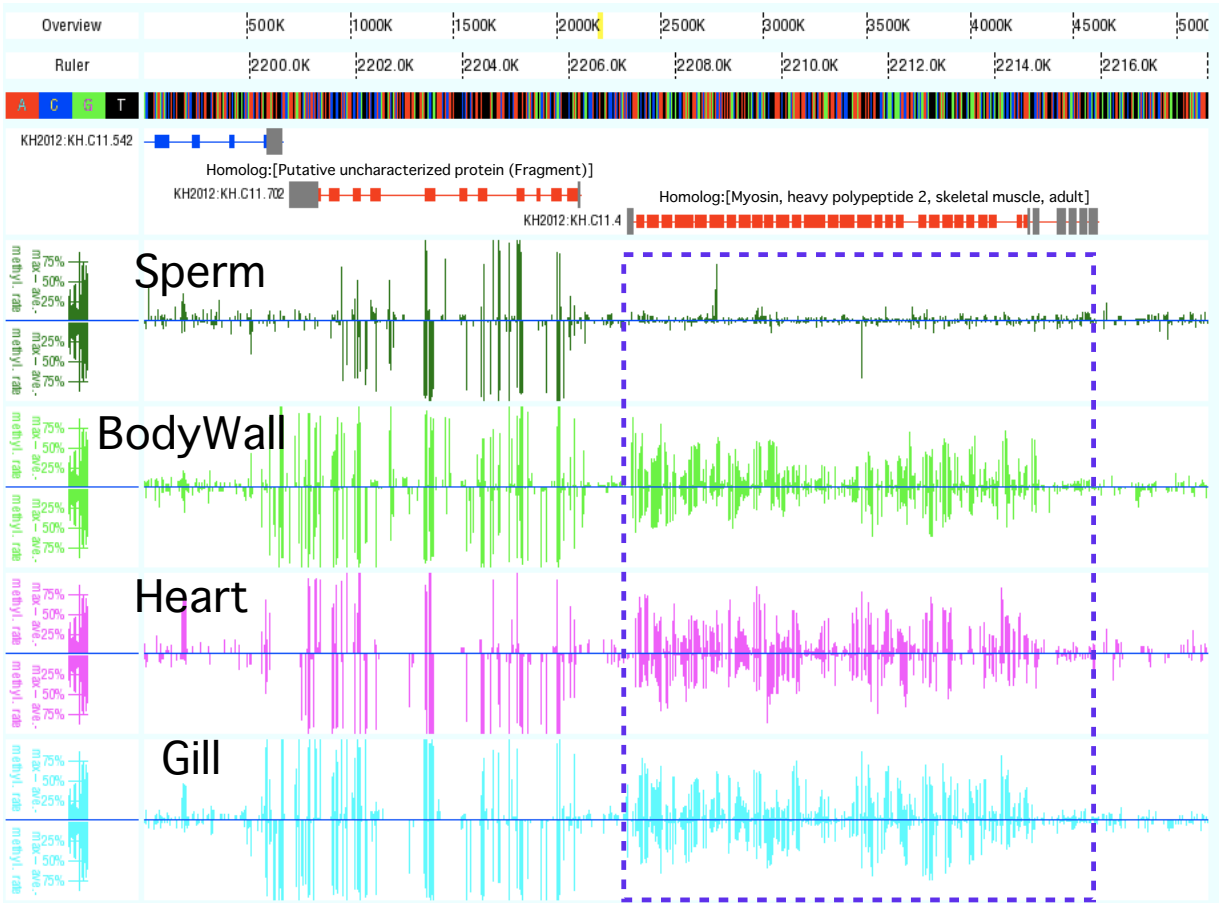

Figure 11: **Tissue specific gene body methylation** Differential methylation, or sperm-sepcific unmethylation, was observed gene body of adult specific myosin subtype (indicated blue dashed line).

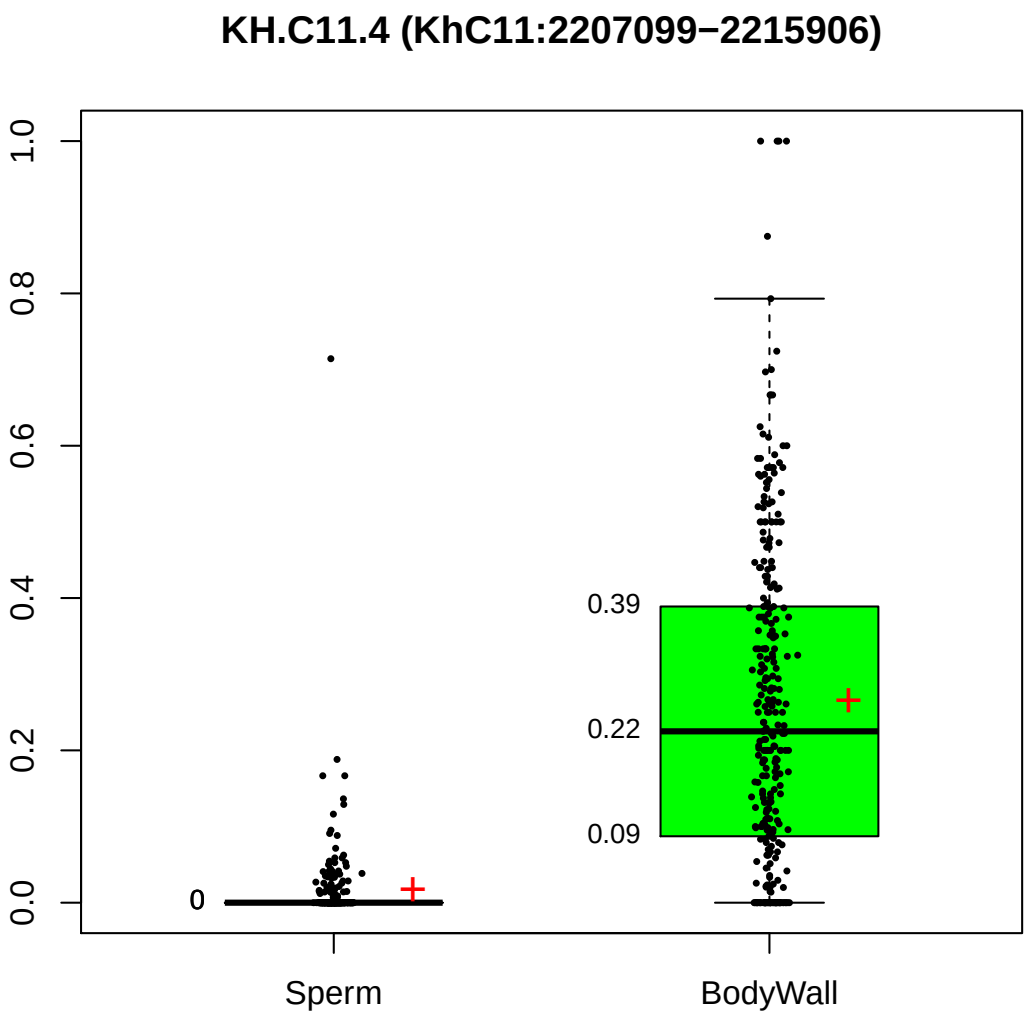

Figure 12: **DMR on KH.C11.4 (Myosin Adult Subtype)**

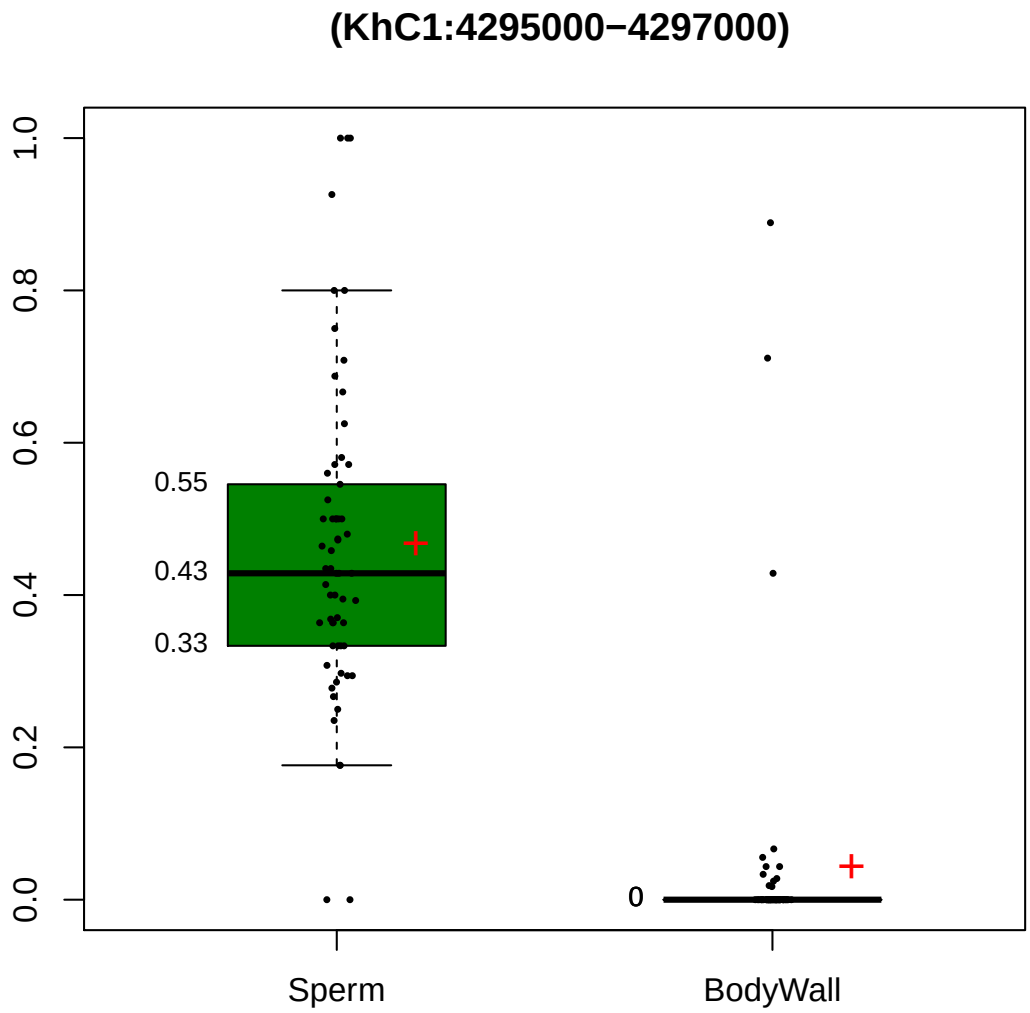

Figure 13: **DMR on KhC1:4295000-4297000 (Sperm promoter methylation)**

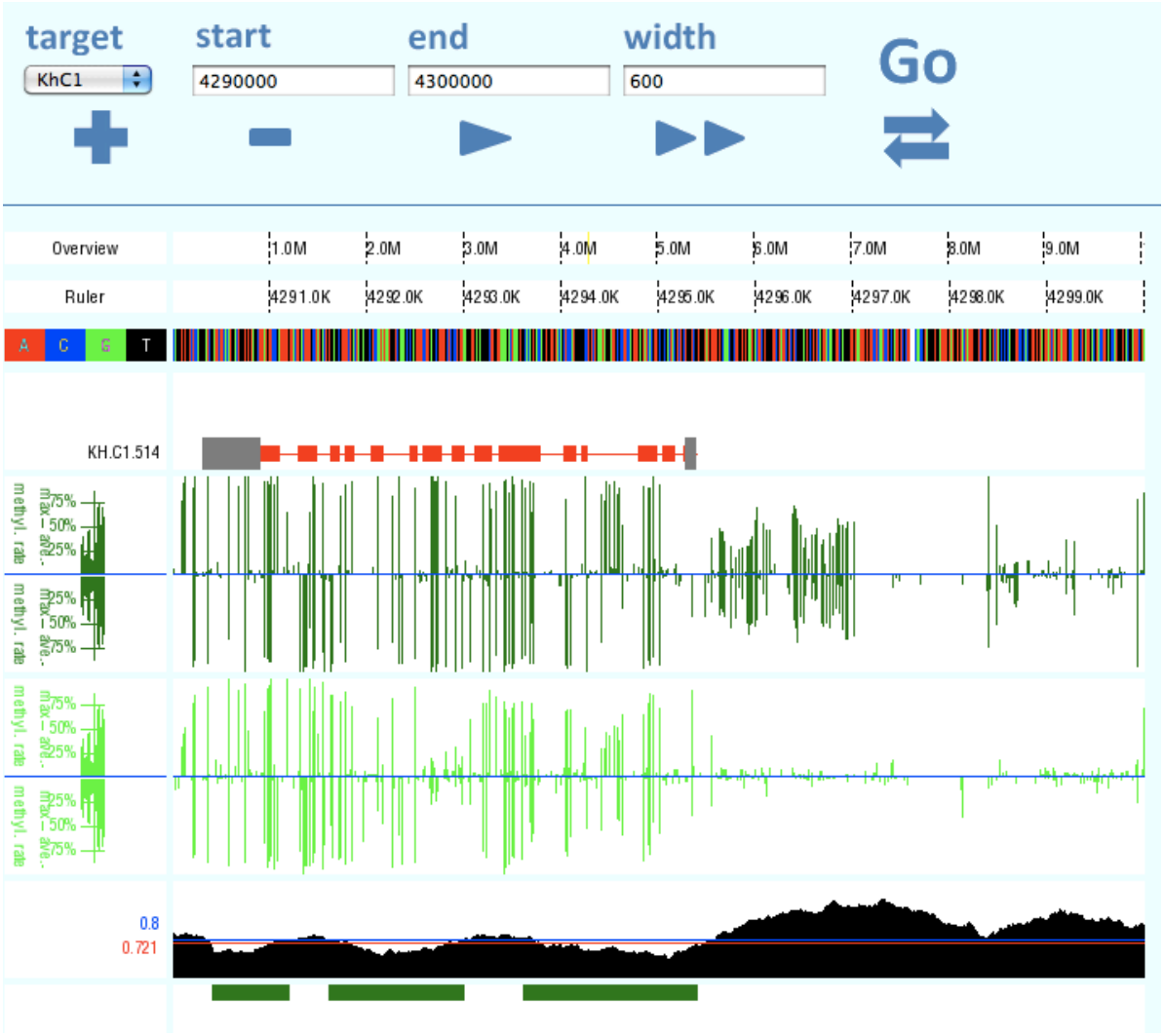

Figure 14: **An example of tissue-specific promoter methylation**
Putative promoter region of this gene is methylated in sperm but escaped methylation in body wall.