Comprehensive characterization of differentially methylated regions associated with mouse imprinted genes

Hisato Kobayashi

Doctor of Philosophy

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies

2005

Abstract

Imprinted genes in mammals show monoallelic expression dependent on parental origin and are often associated with differentially methylated regions (DMRs). There are two classes of DMR: germline DMRs acquire gamete-specific methylation in either spermatogenesis or oogenesis and maintain the allelic methylation differences throughout development; secondary DMRs establish differential methylation patterns after fertilization. Targeted disruption of some germline DMRs showed that they dictate the allelic expression of nearby imprinted genes and the establishment of the allelic methylation of secondary DMRs. However, how the imprinting machinery recognizes germline DMRs is unknown. As a step toward elucidating the sequence features of the germline DMRs, I have determined the extents and boundaries of 15 germline mouse DMRs (including 12 maternally methylated and 3 paternally methylated ones) in 12.5-dpc embryos and sperm by bisulfite sequencing. I found that the average size of the DMRs was 2.7 kb and that their average G+C content was 54.2%. Oligonucleotide content analysis of the DMR sequences revealed that, although they are generally CpG rich, the paternally methylated DMRs contain less CpGs than the maternally methylated ones. Furthermore, based on the SOM analysis, I found that most germline DMRs have features distinct from typical mouse sequences. My findings provide a basis for the further characterization of DMRs.

Contents

Abbeviations

bp: base pair

CAT: chloramphenicol acetyltransferase

DDT: dithiothreitol

DDW: distilled deionaized water

DMR: differentially methylated region

EDTA: ethylenediaminetetraacetic acid

ICR: imprint control region

kb: kilo base pair

PCR: polymerase chain reaction

SDS: sodium dodecyl sulfate

SNPs: single nucleotide polymorphisms

SOM: self-organizing map

Gene symbols

Dnmt: DNA methyltransferase

Dlk1: Delta like 1

Gnas: G protein alpha-subunit

Gnasxl: extra large G protein alpha-subunit

Gtl2: gene trap locus 2

Igf2r: insulin-like growth factor 2 receptor

Impact: the gene is imprinted and encodes a protein with an ancient conversed region

Kcnq1/KvLQT1: potassium voltage-gated channel, KQT-like subfamily, member 1

Lit1/ Kcnq1ot1: long QT intronic transcript 1/*Kcnq1* overlapping transcript 1

Meg1/*Grb10*: maternally expressed gene 1/growth factor receptor bound protein 10

Mkrn1-p1: markorin1 pseudogene

Mkrn3: markorin 3

Nap1l5: nucleosome assembly protein 1-like 5

Nesp: neuroendocrine secretory protein

Nespas: *Nesp* antisense

Peg1/Mest: paternally expressed gene 1/Mesoderm-specific transcript

Peg3: paternally expressed gene 3

Peg10: paternally expressed gene 10

Peg12: paternally expressed gene 12

Rasgrf1: Ras protein-specific guanine nucleotide-releasing factor 1

Rtl1: retrotransposon-like gene 1

Sgce: epsilon-sarcoglycan

Snrpn: small nuclear ribonucleoprotein N

TCEB3C; transcription elongation factor B polypeptide 3C

U2af1-rs1: U2 small nuclear ribonucleoprotein auxiliary factor 1 related sequence 1

Zac1: zinc-finger protein that regulates both apoptosis and cell cycle arrest

1. Introduction

In early 1980s, McGrath and Solter (1984) and Surani et al. (1984) performed pronuclear transfer experiments with fertilized mouse eggs and showed that diploid conceptuses with genomes from only one parent never develop to term. The results indicated that both the maternal and paternal genomes are required for normal development (McGrath and Solter 1984; Surani et al. 1984). In humans, it is known that parthenogenetic activation of unfertilized oocytes results in ovarian teratomas, while conceptuses with no maternal genome (androgenones) results in complete hydatidiform moles. Uniparental disomies, which inherit a pair of particular chromosomes from only one parent, show abnormal phenotypes in mice (Cattanach et al. 2004) and develop disease conditions in humans. The diseases include Prader-Willi syndrome, Angelman syndrome and Beckwith-Wiedemann syndrome (Reik and Walter 2001). The functional non-equivalance between the parental genomes or parental chromosomes is caused by parent-of-origin-dependent monoallelic expression of certain genes. These genes are called “imprinted” genes. Over 80 imprinted genes have thus far been identified in humans and mice, and they play important roles in fetal growth, development of particular somatic lineages, and maternal behavior. Most of them form clusters in the particular genomic regions, and these regions are responsible for the abnormal phenotypes of uniparental disomies. It is estimated that a few to several hundred imprinted genes exist in the placental mammals (Hayashizaki et al. 1994).

The imprinted genes are marked or imprinted in a parent-of-origin dependent way in the parental germline. This marking is called genomic imprinting. Because the parental genomes are essentially identical in sequence, epigenetic mechanisms should be involved in imprinting. One candidate for such epigenetic mark is DNA methylation. In mammals and other vertebrates, DNA methylation occurs exclusively at the cytosine of symmetrical CpG dinucleotides, and it plays important roles in normal development (Li et al. 1992). In fact, within or near the imprinted genes, differentially methylated regions (DMRs), which show parent-of-origin-dependent DNA methylation patterns, are identified (Reik and Walter 2001). Subsequent studies revealed that there are two classes of DMR; the germline DMR (primary DMRs) acquire methylation during gametogenesis and show differential methylation between sperm and oocytes; the secondary DMRs establish their allelic methylation patterns after fertilization, most probably through the influence of the germline DMRs (Lopes et al. 2003).

Direct evidence that DNA methylation is involved in genomic imprinting came from gene knockout experiments in mice. The maintenance methyltransferase Dnmt1 is required to maintain the differential methylation patterns at the DMRs and also the monoallelic expression patterns of the imprinted genes in the embryo proper, at least after implantation (Li et al. 1993). Furthermore, targeted deletion of some germline DMRs in mice have demonstrated that they result in aberrant expression of the associated imprinted genes and loss of allelic methylation of the secondary DMRs. Such germline

DMRs are called imprint control regions (ICRs)(Fitzpatrick et al. 2002; Lin et al. 2003; Liu et al. 2005; Thorvaldsen et al. 1998; Williamson et al. 2004; Wutz et al. 1997; Yoon et al. 2002). Recent studies demonstrated that a *de novo* DNA methyltransferase related protein, Dnmt3L, and a *de novo* DNA methyltransferase Dnmt3a, are required for the establishment of the primary methylation imprints in both the paternal and maternal germlines (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). However, features of the germline DMRs that attract Dnmt3L and Dnmt3a in the germline remain an enigma (see below).

A number of studies have suggested that imprinted genes may have characteristic structural features. For example, it was reported that imprinted genes tend to have few and small introns (Hurst et al. 1996). Other reports described that human and mouse imprinted gene regions contain a relatively low density of short interspersed transposable elements (SINEs) compared with non-imprinted regions (Greally 2002; Ke et al. 2002). Regarding the germline DMRs, other than their relative CpG richness, the only potential feature identified so far is the presence of direct repeats near or within the DMRs (Neumann et al. 1995). However, these features do not hold true for all imprinted genes or regions, and their functional relevance is controversial.

In order to elucidate the structural characteristics of the DMRs, I wanted to obtain their nucleotide sequences. However, I realized that the precise extents of most DMRs are unknown. The lack of accurate information on the extents of

the DMRs is a barrier to the studies on their structural features. I therefore decided to analyze the methylation patterns of all 15 known germline DMRs in their entirety by bisulfite sequencing. In this thesis, I report the results of such studies performed with embryo DNA and sperm DNA. Based on the data obtained, I determined the extents and boundaries of these germline DMRs, although the final decision awaits data from oocyte DNA. As an initial attempt to characterize the DMR sequences, I also carried out oligonucleotide content analysis. The results revealed that the paternally methylated DMRs contain less CpGs than the maternally methylated DMRs. Furthermore, to explore the features of the DMRs concerning the oligonucleotide contents, I used self-organizing maps (SOMs), which were constructed by an unsupervised neural network algorithm. The SOM analysis can classify DNA sequences from various sources into subgroups that generally correspond to species or biological categories based on oligonucleotide frequencies (Abe et al. 2003). In this study, I found that most germline DMRs have features distinct from typical mouse sequences. Based on all these results, I discuss the possible features of the germline DMRs that are recognized by the imprinting machinery containing both Dnmt3L and Dnmt3a.

2. Bisulfite sequencing analysis

2-1. Introduction

Many imprinted genes are associated with DMRs, which show parent-of-origin-dependent DNA methylation patterns. Based on the difference in timing of *de novo* methylation, the DMRs are classified into two classes: the germline DMRs (primary DMRs) acquire methylation during gametogenesis and show differential methylation between sperm and oocytes; the secondary DMRs establish their allelic methylation patterns after fertilization, most probably through the influence of the germline DMRs (Lopes et al. 2003). Recent studies showed that Dnmt3L and Dnmt3a are necessary for the establishment of the methylation imprints at the germline DMRs (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004).

So far, 15 germline DMRs have been identified in mice (Tables 1 and 2). They include 12 maternally methylated DMRs (the *Nespas-Gnasxl*, *Gnas* 1A, *Peg10*, *Peg1*, *Peg3* DMR, *Snrpn*, *Lit1/Kcnq1ot1* (KvDMR), *Zac1*, *Meg1/Grb10*, *U2af1-rs1*, *Igf2r*, and *Impact* DMRs), and 3 paternally methylated DMRs (the *H19*, *Rasgrf1* and *Dlk1-Gtl2* DMRs (IG-DMR)). Deletion mutations in mice have demonstrated that at least six of the germline DMRs act as ICRs and are essential for imprinting (Fitzpatrick et al. 2002; Lin et al. 2003; Liu et al. 2005; Thorvaldsen et al. 1998; Williamson et al. 2004; Wutz et al. 1997; Yoon et al. 2002). Although previous studies identified the location of each DMR (Table 2), the accurate extent of the DMR has not been determined in many cases.

To elucidate the structural characteristics of the germline DMRs, I decided to analyze the extent of these DMRs by bisulfite sequencing. The principle of bisulfite sequencing is depicted in Figure 1. Treatment of the genomic DNA with sodium bisulfite converts unmethylated cytosine to uracil residues (Frommer et al. 1992). As a consequence, the converted DNA is no longer self-complementary, and amplification of either the top or bottom is possible using strand-specific PCR primers. PCR amplification results in conversion of uracil (previously cytosine) to thymine, and of methyl-cytosine to cytosine, and these change can be easily detected by sequencing the PCR products. The main advantage of this method is that the methylation status of individual CpGs is detectable in the short genomic region amplified by PCR.

In this chapter, I report the results of my bisulfite sequencing analysis on the allelic methylation patterns of all 15 germline DMRs in whole embryos (12.5 day postcoitum) DNA and in sperm DNA. Based on these results, I determined the DMR boundaries and the extents of the germline DMRs.

2-2. Materials and methods

Animals

C57BL/6 (Mus musculus domesticus) mice were purchased from Clea Japan, Inc., and JF1 (Mus musculus molossinus) mice (Koide et al. 1998) were obtained from the Genetic Strains Research Center of our institute. To isolate F_1 hybrid embryos at 12.5 dpc and sperm from F_1 hybrid adult (10-week-old) mouse, C57BL/6 females were crossed with JF1 males.

Genome sequences

Mouse (C57BL/6) genome sequences (containing the DMRs) were obtained from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/). I checked the positions of the DMRs based on previous reports (all these papers were listed in Table 1.). These sequences were used for designing PCR primers for bisulfite sequencing. Sequernce data were analyzed by Genetyx version 7 (Genetyx, Tokyo, Japan). Student's two-tailed *t*-test was used to show significant differences between values.

DNA isolation

Genomic DNA was prepared from whole 12.5-dpc F_1 hybrid embryos and spermatozoon. The embryos were homogenized using an all-glass Dounce homogenizer (20 strokes) in 5 ml of extraction buffer (20 mM Tris-HCl [pH 7.5], 10 mM EDTA, 150 nM NaCl, 10 mM KCl). The homogenate was digested by 200

μg/ml Proteinase K with 1% sodium dodecyl sulfate (SDS) overnight at 50□. Genomic DNA from embryos was extracted with phenol, and phenol/chloroform (1:1), and precipitated with sodium acetate and isopropanol. DNA pellets were washed in 70% ethanol, and dissolved in distilled deionized water. Sperm was released from ductus deferens of F_1 hybrid adult mouse and lysed in 500 μl of extraction buffer with 10 mM dithiothreitol (DDT). Sperm DNA was isolated as descrived above.

Bisulfite sequencing

Genomic DNA (5 μg) isolated from mouse embryos or sperm was treated with sodium bisulfite (Frommer et al. 1992) using an EZ DNA Methylation Kit (Zymo Research, Orange). The bisulfite-treated DNA was dissolved in 100 μl of distilled deionaized water (DDW). Aliquots of bisulfite-treated DNA solution, equivalent of 5 ng genomic DNA, were used for PCR. One hundred and three oligonucleotide primer pairs were used to study the 15 DMRs. All primer sequences and PCR conditions are shown in Table 3. The PCR products (amplicons) were subcloned into pGEM-T vector (Promega, Medison) and transformed into JM109. Colonies were picked up (about 24 colonies for each PCR product) and transferred into 96-well plates, and DNA was amplified by rolling circle amplification (Neumann et al. 1995) using a TempliPhi DNA Amplification Kit (GE Healthcare Bio-Sciences, Little Chalfont, UK). DNA sequencing was done using a BigDye Terminator version 3.1 Cycle Sequencing

Kit (Applied Biosystems, Foster city) with standard primers (M13 For; 5'-GTTTTCCCAGTCACG-3' and M13 Rev; 5'-CAGGAAACAGCTATG-3'). Sequences were analyzed on an ABI Prism 3700 and 3130xl Genetic Analyzer (Applied Biosystems, Foster city). Sequence data analysis was done using ATGC version 4 (Genetyx, Tokyo, Japan).

2-3. Results

Bisulfite sequencing analysis

To accurately determine the extents of the 15 mouse germline DMRs, I retrieved the genomic sequences containing the DMRs from public databases. I then determined which portions of the respective sequences had been examined for allelic methylation and which portions had been judged as DMRs in previous reports (references listed in Table 1). Bisulfite PCR primer pairs were designed to close the gaps and also to extend the analysis to the DMR boundaries. To distinguish between the parental alleles using SNPs, a cross was made between C57BL/6 females and JF1 males. Genomic DNA was isolated from F_1 embryos at 12.5 dpc and F_1 epididymal sperm subjected to sodium bisulfite treatment. I simultaneously identified SNPs and determined the allelic methylation status by cloning and sequencing of the PCR products (amplicons).

Primer design was done according to the following rules. (1) Primers should be approximately 30-40 bp in length to increase primer specificity and complexity because bisulfite-treated DNA is very AT-rich and basically devoid of cytosines. (2) Primers should not contain any CpG sites within their sequences. (3) Primers should contain thymines converted from cytosines of non-CpG sites, to amplify only bisulfite-treated DNA.

In this study, I prepared 127 primer pairs, of which 103 worked fine and 24 failed (most of the failed primer pairs were designed within repetitive elements). Among the 103 amplicons that I obtained, 89 contained one or more

SNPs and thus could be analyzed in an allele-specific way. The average size of the amplicons was 551 bp (minimum size, 227 bp; maximum size, 785 bp). The total length of DNA covered by these amplicons was 56,769 bp. Twenty-four clones were picked up for each amplicon and sequenced. Because I had to handle a large number of samples, a large-scale bisulfite sequencing protocol was devised. A rolling circle amplification of plasmid DNA (Nelson et al. 2002) was introduced, and this greatly simplified the protocol by avoiding bacterial culture and plasmid preparation. About 6200 sequencing runs (including those for both 12.5-dpc embryo and sperm) were carried out, and total sequencing output in this study was about 2.8 MB. The scale of the experiment and some useful parameters obtained from the bisulfite sequencing results are summarized in Table 4.

Extents and boundaries of the DMRs

I adopted the following criterion to determine the extent of a DMR. As for 12.5-dpc embryos, if none or only one of four consecutive CpGs shows a methylation level below 70% on the more-methylated allele or over 30% on the less-methylated allele, these four CpGs are judged to be a part of the DMR. Any four consecutive CpGs that do not meet the criterion are at least partially excluded from the DMR: the CpGs that do not meet the criterion and CpGs located between them are disqualified as part of the DMR. As for sperm DNA, if none or only one of four consecutive CpGs associated with paternally

methylated DMRs shows a methylation level below 70%, these four CpGs are judged to be a part of the DMR. If none or only one of four consecutive CpGs associated with maternally methylated DMRs shows a methylation level over 30%, these four CpGs are judged to be a part of the DMR. Any four consecutive CpGs that do not meet the criterion are at least partially excluded from the DMR: the CpGs that do not meet the criterion and CpGs located between them are disqualified as part of the DMR.

The results obtained by the bisulfite sequencing studies are shown in Figure 2 (a-o). For example, I prepared 4 primer pairs to study the maternally methylated *Gnas* 1A DMR (Figure 2 (b)). In the region A, no SNP was available for allele discrimination, but all CpGs were highly methylated, most likely on both parental alleles in 12.5-dpc embryo. In the region B, most CpGs were maternally methylated, but 5 CpGs near the 5' end were also partially methylated on the paternal allele. Thus, the 5' boundary in 12.5-dpc embryo was identified between the 5^{th} and the 6^{th} CpG in the region B. All CpGs in the region C were maternally methylated. In the region D, no SNP was available, but we presumed that 10 clones were from the maternal allele and the remaining 12 from the paternal allele, based on the known allelic methylation pattern. Interestingly, the maternal allele started to lose methylation at the 35^{th} CpG and became completely unmethylated after the 47^{th} CpG, consistent with the previous report by Lin et al. (2000). According to the described criterion, I judged that these 20 (13 partially methylated and 7 unmethylated) CpGs near the 3' end of the region H were

outside of the DMR, and the 3' boundary in 12.5-dpc embryo was identified between the 34th and the 35th CpG in the region D. The bisulfite analysis of sperm DNA showed hypomethylation in all regions and thus did not change the boundaries. Based on these results, the size of the *Gnas* 1A DMR was determined to be 2.2 kb.

Another example is shown in Figure 2 (n). I prepared 14 primer pairs to study the paternally methylated *Rasgrf1* DMR. The regions A and B was highly methylated on the paternal allele and partially methylated on the maternal allele in 12.5-dpc embryos. The regions D-H were methylated only on the paternal allele. In the region C, no SNP was available for allele discrimination, but I presumed that this region is paternally methylated based on the allelic methylation pattern of the region D. The 1st CpG in the region C showed a methylation level below 70 % on the paternal allele, so I judged that the 5' boundary is located between the 1st and 2nd CpG in the region C. The regions A-H were highly methylated on both parental alleles in sperm. I was unable to prepare primers specific to the tandem repeat region (1.7 kb) between the regions H and I. The regions I-L were highly methylated only on the paternal allele in 12.5-dpc embryos, however, they showed hypomethylation on both parental alleles in sperm. The region M was partially methylated only on the paternal allele, and the region N was partially methylated on both parental alleles. Thus, I judged that the 3' boundary was located between the regions H and I. Based on these results, the size of the *Rasgrf1* DMR was determined to be 3.2

kb.

The boundaries and extents of all 15 germline DMRs were thus determined. I found that several methylation patterns are observed at the DMR boundaries. In the 12.5-dpc embryos, some DMRs had sharp boundaries (for example, *Impact* DMR; Figure 2 (l)) and others had transition zones (for example, *Gnas* 1A DMR; Figure 2 (b), and *Peg1/Mest* DMR; Figure 2 (d)). The largest transition zone that I identified so far had a size of over 1.4 kb (the 5' transition zone of the *Peg1/Mest* DMR), but in most cases the exact size of the transition zone is unknown. In most cases, the region beyond the boundary was highly methylated on both parental alleles, but there were exceptions such as the *Gnas* 1A DMR, whose 3' flanking region was unmethylated (Figure 2 (b)), and the *Lit1/Kcnq1ot1* DMR, whose 3' flanking region was partially methylated (Figure 2 (f)).

The gene organizations, CpG distributions, and extents of all 15 DMRs are shown schematically in Figure 3. The sizes and G+C contents of the DMRs determined from the data are summarized in Table 5. The results showed that the size of the germline-DMRs ranged from 1.5 kb (*Igf2r* DMR2) to 4.8 kb (*Peg1/Mest* DMR), with an average size of 2.7 kb. Interestingly, I found that the average G+C content of the paternally methylated DMRs (47.7%) was significantly smaller than that of the maternally methylated DMRs (55.8%). Interestingly, it also appeared that CpG is less densely distributed in the paternally methylated DMRs than in the maternally methylated DMRs (Figure 3).

2-4. Discussion

In order to collect the accurate sequence information on the germline DMRs, I determined the site-by-site allelic methylation patterns of 15 germline DMRs in 12.5-dpc mouse embryos and sperm by bisulfite sequencing. To do this, I developed a pipeline for large-scale bisulfite sequencing data production in Prof. Sasaki's Laboratory. The introduction of the rolling circle amplification of plasmid DNA (Nelson et al. 2002) greatly simplified the protocol and contributed to the high-throughput analysis. By integrating both the methylation data from embryos and those from sperm, I tried to determine the extent of the 15 germline DMRs.

The criteria that I devised to determine the extent of the DMRs are strict because I wanted to know the extent of the "core" regions of the germline DMRs rather than the entire regions showing both clear and biased allelic differences in methylation. Therefore, transitions zones were excluded from the DMRs (also see the discussions below). The size of the germline-DMRs determined according to the criteria ranged from 1.5 kb (*Igf2r* DMR2) to 4.8 kb (*Peg1/Mest* DMR), with an average size of 2.7 kb (Table 5). It was also revealed that the average G+C content of the germline DMRs (54.2%) is higher than that of the whole mouse genome (41.7%), which is consistent with the fact that almost all imprinted gene clusters are located in the G+C-rich chromosome regions (isochors) (Neumann et al. 1995). Interestingly, I found that the average G+C content of the paternally methylated DMRs (47.7%) was significantly smaller than that of the maternally methylated DMRs (55.8%). Furthermore, it also

appeared that CpG is less densely distributed in the paternally methylated DMRs than in the maternally methylated DMRs (Figure 3). Such sequence features of the DMRs are further pursued in the next chapters.

One thing one has to keep in mind is that the bisulfite methylation analysis was done in 12.5-dpc embryos and sperm, but not in oocytes. Because the primary methylation imprints and methylation of the surrounding regions can change during embryonic development (Yoder et al. 1997), the final decision on the precise extent of the germline DMRs awaits the methylation analysis in the oocytes. Although challenging, such a study is not impossible if one could collect enough number of oocytes. My next goal is to determine the methylation status of the germline-DMRs in oocytes.

During this work, I also found that the DMRs had several different methylation patterns at the boundaries. For example, some DMRs had sharp boundaries and others had transition zones. The largest transition zone that we identified so far had a size of over 1.4 kb (the 5' transition zone of the *Peg1* DMR), but in most cases the exact size of the transition zone is unknown. Usually, the region beyond the boundary was highly methylated on both parental alleles, but there were exceptions such as the *Gnas* 1A DMR, whose 3' flanking region was unmethylated, and the *Lit1/Kcnq1ot1* DMR, whose 3' flanking region was partially methylated. One interesting observation made in the study was that the 3' flanking region of the *Rasgrf1* DMR is initially unmethylated in sperm DNA but has become methylated on the paternal allele in 12.5-dpc embryos (Figure

2(n)). This is a clear example of an expansion of methylation form the germline DMR to the adjacent regions during embryonic development.

In summary, this is the first comprehensive description of the extents and boundaries of the germline DMRs and the information obtained should be useful to the studies on the mechanisms of genomic imprinting. I will describe my initial studies to identify the structural features of the DMRs in the following chapters.

3. Oligonucleotide frequencies

3-1. Introduction

The germline DMRs associated with the imprinted genes are recognized in the germline by the methylation machinery containing Dnmt3a and Dnmt3L (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). The oocyte- or sperm-specific methylation patterns established at the DMRs are then passed onto the zygote and maintained and propagated throughout embryonic development (Reik and Walter 2001). However, no consensus sequence has been identified for the germline DMRs. In the previous chapter, I determined the extents of 15 germline DMRs by bisulfite sequence analysis of the DNA from 12.5-dpc mouse embryos and sperm. Based on the data, I derived a collection of nucleotide sequences spanning the core region of all these DMRs. I therefore wanted to use these sequences to identify the structural features specific to the germline DMRs.

Oligonucleotide frequency is a fundamental characteristic of individual genomes and used for a long period as a basic phylogenetic parameter to characterize individual genomes and genomic portions (Gentles and Karlin 2001; Karlin et al. 1998; Nussinov 1984; Phillips et al. 1987; Rocha et al. 1998). Furthermore, specific oligonucletides contribute to higher order structure of DNA, such as bending (Bolshoy 1995), and local oligonucleotide biases appear to reflect the properties of DNA such as stacking energies, modification, replication and repair mechanisms (Karlin 1998). Profound flanking sequence preference of

Dnmt3a and Dnmt3b has also been noted (Handa and Jeltsch 2005) suggesting that CpGs contained in specific oligonucleotides may be prone to methylation. As described in the previous chapter, the germline DMR sequences that I determined are more G+C-rich than the whole genome, and the average G+C content of the paternally methylated DMRs was significantly smaller than that of the maternally methylated DMRs (Table 5). Furthermore, it appeared that CpG dinucleotide is less densely distributed in the paternally methylated DMRs than in the maternally methylated DMRs (Figure 3). I therefore performed oligonucleotide content analyses on the germline DMRs to identify their possible features.

In this chapter, I describe the results of di-, tri-, and tetranucleotide content analyses on the core region sequences of the 15 germline DMRs. This study indicated, for example, that the paternally methylated DMRs contain less CpGs than the maternally methylated DMRs. The significance and functional relationship of these findings with genomic imprinting and methylation will be discussed.

3-2. Materials and methods

Genome sequences

Mouse (C57BL/6) genome sequences (containing the DMRs) were obtained from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/). Non-imprinted mouse CpG islands were collected from Gardiner-Garden and Frommer (1987) (16 islands) or from GenBank by a keyword (CpG island) search (33 islands). The exact sequence of each island was determined by the criteria proposed by Gardiner-Garden and Frommer (1987). All collected CpG islands are shown in Table 6.

Data analysis

Oligonucleotide content calculation was done using Genome1 (UNTROD, Kyoto, Japan). Student's two-tailed *t*-test was used to show significant differences between values.

3-3. Results

Dinucleotide content analysis

As an initial attempt to elucidate the sequence features of the germline DMRs, I examined the contents of dinucleotides in the 15 DMRs according to the determined sequences (Table 7). The whole mouse genome and 49 randomly selected non-imprinted CpG islands (fulfilling the criteria by Gardiner-Garden and Frommer, 1987; see Materials and methods for details) were also analyzed as references. We found that the contents of some dinucleotides (such as ApC+GpT, ApG+CpA, and GpA+TpC; complementary dinucleotides are grouped) were not much different among the DMRs, CpG islands and whole genome (Table 7). By contrast, the contents of the remaining dinucleotides showed significant differences among the three groups. Among these dinucleotides, CpG showed the greatest difference: its content in the DMRs (4.92%) was 5.9-times higher than that in the whole mouse genome (0.84%) and 1.6-times lower than that in the 49 non-imprinted CpG islands (7.80%)(Table 7). Thus, the DMRs showed a CpG content intermediate between the whole genome and CpG islands. For dinucleotides whose contents showed differences among the three groups, the DMRs always showed a value intermediate between those of the whole genome and CpG islands (Table 7).

I also compared the dinucleotide contents between the paternally methylated DMRs and maternally methylated DMRs. Statistical analysis using Student's *t*-test revealed that the maternally methylated DMRs have significantly

lower CpA+TpG, ApT and TpA contents and significantly higher CpG and GpC contents than the paternally methylated DMRs. The fact that the smallest P value was obtained for the CpG content clearly indicates that the paternally methylated DMRs contain less CpGs than the maternally methylated DMRs (Table 7).

Trinucleotide content analysis

Next, I examined the contents of trinucleotides in the 15 DMRs (Table 8). I found that the contents of TpCpA+TpGpA, TpGpG+CpCpA, ApApG+CpTpT, CpApG+CpTpG, ApGpG+CpCpT, GpTpG+CpApC, GpGpT+ApCpC and GpApC+GpTpC, were not much different among the DMRs, CpG islands and whole genome. By contrast, the contents of the remaining trinucleotides showed differences among the three groups. For these remaining trinucleotides, the DMRs almost always showed a content value intermediate between that of the whole genome and CpG islands. I found that trinucleotides containing CpG showed large differences among the DMRs, whole mouse genome and CpG islands. The greatest difference was observed for CpGpC+GpCpG: its content in the DMRs (3.40%) was 16.2-times higher than that in the whole genome (0.21%) and 1.7-times lower than that in the 49 non-imprinted CpG islands (5.87%)(Table 8). Only the content of TpGpC+GpCpA was higher in the DMRs (4.33%) than in the CpG islands (3.53%) and whole genome (2.47%).

I also compared the trinucleotide contents between the paternally

methylated DMRs and maternally methylated DMRs. Statistical analysis using Student's *t*-test revealed that the maternally methylated DMRs have significantly lower TpCpA+TpGpA, TpApT+ApTpA, TpGpT+ApCpA and ApApT+ApTpT contents and significantly higher TpCpG+CpGpA, GpApG+CpTpC, CpCpG+CpGpG, GpCpG+CpGpC and GpGpC+GpCpC contents than the paternally methylated DMRs (Table 8).

Tetranucleotide content analysis

Finally, I examined the contents of tetranucleotides in the 15 germline DMRs (Table 9). I found that the contents of many tetranucleoide showed differences among the DMRs, CpG islands and whole genome. For these tetranucleotide, the DMRs almost always showed a content value intermediate between that of the whole genome and CpG islands. I found that almost all tetranucleotides containing CpG dinucleotide (for example, CpGpCpG, GpCpGpC, GpCpGpG+CpCpGpC, GpGpCpG+CpGpCpC, and GpCpCpG+CpGpGpC) show great difference among the DMRs, CpG islands and whole genome. The contents of GpCpApA+TpTpGpC and TpGpCpA were higher in the DMRs (0.93% and 0.54% respectively) than in the CpG islands (0.57% and 0.37% respectively) and whole genome (0.70% and 0.37% respectively).

I also compared the tetranucleotide contents between the paternally methylated DMRs and maternally methylated DMRs. Statistical analysis using

Student's *t*-test revealed that the maternally methylated DMRs have significantly lower contents for 11 tetranucleotides (for example, GpApApT+ApTpTpC, ApTpGpG+CpCpApT, TpGpApA+TpTpCpA, GpCpTpA+TpApGpC and ApApGpT+ApCpTpT) and significantly higher contents for 16 tetranucleotides (most of them involved CpG dinucleotide: for example, GpTpCpG+CpGpApC, CpGpGpG+CpCpCpG, CpGpCpA+TpGpCpG, GpCpGpG+CpCpGpC and GpGpCpG+CpGpCpC) than the paternally methylated DMRs (Table 9).

3-4. Discussion

My ultimate goal is to identify the structural features specific to the germline DMRs. As an initial step, I used the germline DMR sequences determined in the previous chapter to calculate their di-, tri-, and tetranucleotide contents. Regarding dinucleotides, I found that the DMRs have more CpGs than the whole mouse genome, but in general less CpGs than the non-imprinted CpG islands. Furthermore, we found that the paternally methylated DMRs contain less CpGs than the maternally methylated DMRs (Figure 3 and Table 7).

The notion that the germline DMRs contain less CpGs than non-imprinted CpG islands was previously suggested by the observations that some DMRs (such as the *H19* DMR) do not have a CpG island. Indeed, based on the sequences collected here, and according to the criteria proposed by Gardiner-Garden and Frommer (1987), three (*Peg3*, *Snrpn* and *H19*) out of the 15 DMRs had no CpG islands. Also, three others (*Nespas-Gnasxl*, *Rasgrf1* and *Dlk1-Gtl2* IG-DMR) had only weak and small (<400 bp) CpG islands.

Why are the DMRs less CpG rich than non-imprinted CpG islands? One possible explanation is that the DMRs were initially more CpG rich but, because CpG dinucleotides are mutation hotspots, they gradually lost CpGs during evolution. It is well known that methylated cytosine is mutable to thymine (C/T transition) by spontaneous deamination (Holliday and Grigg 1993). The germline DMRs inevitably accumulate C/T transitions and lose CpGs in successive generations. This idea is supported by the fact that the DMRs have a

higher CpA+TpG content than the non-imprinted CpG islands (Table 7). Furthermore, the mutability may also explain the difference in CpG content between the paternally and maternally methylated DMRs. It has been shown that the paternal methylation imprints are established in gonocytes (or prospermatogonia) in the fetal testis and persist in the germline throughout the reproductive life of the male (Davis et al. 1999; Li et al. 2004; Ueda et al. 2000). By contrast, the maternal methylation imprints are imposed in growing oocytes after birth (Li et al. 2004; Lucifero et al. 2004). Thus, methylation imprints persist longer in the male germline than in the female germline. In addition, male germ cells divide many times after methylation imprints are established, but female germ cells do not. Therefore, the paternally methylated DMRs have more chance to accumulate C/T mutations than the maternally methylated DMRs.

It is also possible, however, that the CpG content is one of the features recognized by the *de novo* methylation machinery. CpG islands are generally free of methylation, but weak or small CpG islands may lose protection from methylation in one of the germlines and could behave as DMRs. At present, we do not know why the DMRs with higher CpG contents (maternally methylated DMRs) can be recognized preferentially by the *de novo* methylation machinery in the female germline compared with those having lower CpG contents (paternally methylated DMRs).

In the tri- and tetranucleotide content analyses, I found that oligonucleotides containing CpG shows very different content values among the

DMRs, CpG islands and whole mouse genome. For the CpG-containing oligonucleotides, the DMRs almost always showed a content value intermediate between that of the CpG islands and whole mouse genome. However, trinucleotides TpGpC+GpCpA and tetranucleotides containing these trinucleotides (GpCpApA+TpTpGpC and TpGpCpA) showed higher contents in the DMRs than in the CpG islands and whole genome. The significance of these ologonucleotide is currently unknown.

Recently, methylation kinetic experiments revealed that there is a clear relationship between the tendency of a CpG site to undergo methylation and its flanking sequence (Handa and Jeltsch 2005). It was observed that CpG with purine bases at the 5' end and pyrimidine bases at the 3' end (YpCpGpR) were preferred by the activity of Dnmt3a and Dnmt3b. I therefore examined whether YpCpGpR tetranucleotides are preferred in the DMRs, CpG islands, or the mouse genome. However, no discrete positive or negative correlation was observed in any of the three categories. Further studies are needed identify the features and significance in oligonucleotide frequencies of the germline DMRs.

4. SOM analysis

4-1. Introduction

In the previous chapter, the oligonucleotide content analysis revealed that the 15 germline DMRs have some unique oligonucleotide frequency patterns among the mouse genome sequences. It is known that oligonucleotide frequency is a fundamental characteristic of individual genomes and used for a long period as a basic phylogenetic parameter to characterize individual genomes and genomic portions. Many groups have reported that oligonucleotide frequency, which is an example of high-dimensional data, varies significantly among genomes and can be used to study genome diversity (Gentles and Karlin 2001; Karlin 1998; Karlin et al. 1998; Nussinov 1984; Phillips et al. 1987; Rocha et al. 1998).

Unsupervised neural network algorithm, Kohonen's Self-Organizing Map (SOM), is a powerful tool for clustering and visualizing high-dimensional complex data on a two-dimensional map. On the basis of batch learning SOM, Abe et al. have developed a modification of the conventional SOM for genome sequence analyses, which makes the learning process and resulting map independent of the order of data input (Abe et al. 2002; Abe et al. 2003; Kanaya et al. 2001). They previously constructed the SOMs for di-, tri-, and tetranucleotide frequencies in genomic sequences (fragments) from 65 bacteria and 6 eukaryotes. In the resulting SOMs, the sequences were clustered (i.e., self-organized) according to species without any information regarding the

species, and increasing the length of the oligonucleotides from di- to tetranucleotides increased the clustering power (Abe et al. 2003). They also used the SOM analysis to identify horizontally transferred “alien” sequences in a prokaryotic genome (Kanaya et al. 2001).

A possible link between epigenetic modifications of parasitic DNA and imprinting has been postulated (Barlow 1993; Yoder et al. 1997). Several lines of evidence support this hypothesis. For example, at least 8 of the imprinted genes appear to have arisen by retrotransposition events (Walter and Paulsen 2003). Among these, 6 genes (*Nap1l5*, *Mkrn3*, *Mkrn1-ps1*, *Peg12, TCEB3C* and *U2af1-rs1*) are derived by retrotransposition events from multi-exonic precursor genes and 2 (*Peg10* and *Rtl1*) are retrotransposon-derived genes. Furthermore, it has been suggested that transgenes possessing prokaryotic sequences, such as plasmid vector DNA, are prone to methylation imprinting (Barlow 1993).

In this study, I used di-, tri- and tetranucleotide SOMs, constructed using 2-kb genomic sequences derived from 10 eukaryotes, including the mouse, and 143 prokaryotes, to investigate the genomic characteristics of the germline DMRs. Mapping of the germline DMRs on the SOMs revealed that they have features distinct from typical mouse sequences. Implications of the findings will be discussed.

4-2. Materials and methods

Genome sequences

The genome sequences of 10 eukaryotes (including the mouse) and 143 prokaryotes (Table 10) were obtained from the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu/). When the number of undetermined nucleotides (Ns) in a sequence exceeded 10% of the window size (2-kb), the sequence was omitted from the analysis. When the number of Ns was less than 10%, the oligonucleotide frequencies were normalized to the length without Ns and included in the analysis. The 15 germline DMR sequences were obtained as described in Chapter 2 and are listed in Table 5. Fifty-nine non-imprinted mouse CpG islands were collected as described in Chapter 3 and are shown in Table 6.

SOM analysis

The di-, tri- and tetanucleotide SOMs were constructed by Dr. Takashi Abe (Laboratory of Research and Development of Biological Databases, NIG) from the genome sequences of 10 eukaryotes and 143 prokaryotes (Table 10) as following. SOM implements nonlinear projection of multi-dimensional data onto a two-dimensional array of weight vectors, and this effectively preserves the topology of the high-dimensional data space (Kohonen, 1982 and 1990; Kohonen et al., 1996). On the basis of batch learning SOM, conventional Kohonen's SOM was modified for genome informatics to make the learning

process and resulting map independent of the order of data input (Kanaya et al., 1998 and 2001; Abe et al., 2002 and 2003). The initial weight vectors were defined by Principal Component Analysis (PCA) instead of random values. Weight vectors (w_{ij}) were arranged in the two-dimensional lattice denoted by i (= 0, 1, .., I-1) and j (= 0, 1, .., J-1). The j was defined by the nearest integer greater than (σ_2/σ_1) x I. σ_1 and σ_2 were the standard deviations of the first and second principal components, respectively. Weight vectors (w_{ij}) were set and updated as described previously (Abe et al. 2003). Di-, tri- and tetranucleotide frequencies in each sequence was calculated, and the frequency for the complimentary di-, tri, and tetranucleotide pairs (degenerated oligonucleotide pairs) was mapped to the lattice point with the shortest distance in the multidimensional frequency space represented by the 2-kb DegeDi-, DegeTri- and DegeTetra- SOMs, after normalization for the sequence. All germline DMRs and non-imprinted CpG islands were mapped on the SOMs based on their oligoncleotide frequencies.

4-3. Results

The DegeDi-, DegeTri- and DegeTetra-SOM were constructed by Dr. Takashi Abe using 2,538,037 2-kb sequences derived from the genomes of 10 eukaryotes containing the mouse and 143 prokaryotes. In the DegeTri- and DegeTetra-SOM, most of the 2-kb segments from each organism are distributed to form species-specific territories (Figure 4(a)). In contast, the classification efficiency of the DegeDi-SOM is a little lower than the others. In the DegeTri-SOM, a cluster of mouse genome sequences was located in a small area near the right edge of the SOM, away from the main mouse territory. Forty-nine non-imprinted CpG islands were mapped in the rightmost quarter of the SOMs, mainly in this small area in the DegeTri-SOM (Figure 4 (c)).

To assess the characteristics of the 15 germline DMRs, they were mapped on these SOMs (Figure 4 (b)). Coordinates of the lattice point of each DMR segment and origins of the comapped genome segments are shown in Table 11. The results demonstrated that all DMRs are located in the right half, notably near the right edge, of the SOMs. In the DegeDi-SOM and DegeTri-SOM, all DMR were mapped to lattice points that include sequences from more than one species. In fact, many of the DMRs were located on the borders of the mouse territories, and a few of them were even in other species' territories. Interestingly, the 12 maternally methylated DMRs were located in the rightmost quarter of the SOMs, specifically in the CpG island area (Figure 4 (c)), while the paternally methylated DMRs were more centrally located. In the DegeTetra-SOM,

most of the DMRs were again mapped to lattice points that include sequences from more than one species, but the *H19* DMR was mapped to a lattice point that includes only mouse sequences. Surprisingly, the *Meg1/Grb10* and *U2af1-rs1* DMRs were mapped to lattice points to which only prokaryote sequences were clustered.

To reveal which oligonucleotide frequencies contribute to the sequence distribution on the SOMs, the frequency of each oligonucleotide obtained from the weight vector for each lattice point in the SOMs was calculated and normalized with the level expected from the mononucleotide composition for each lattice point. This normalization allowed oligonucleotide frequencies in each lattice point to be studied independently of mononucleotide compositions. I found that some oligonucleotide frequencies for each weight vector were reflected in the horizontal or vertical axis. For example, oligonucleotide frequencies containing CpG dinucleotide were reflected in vertical axis and increased from top to bottom, and ApT, TpA, TpApT+ApTpA and GpCpApA+TpTpGpC frequencies were reflected in horizontal axis and decreased from left to right in each SOM (Figure 5).

Based on these data, it was estimated that the distribution of the DMRs in the right side on each SOM is caused by the frequencies of ApT, TpA, TpApT+ApTpA and GpCpApA+TpTpGpC, which were reflected in horizontal axis on the SOMs. However, it was unidentified which oligonucleotide frequency contributes the difference in distribution of paternally methylated

DMRs, maternally methylated DMRs and non-imprinted CpG islands.

4-4. Discussion

The SOM was recently used to analyze oligonucleotide frequencies in a wide variety of prokaryotic and eukaryotic genomes. The SOM analysis can classify DNA sequences from various sources into subgroups that generally correspond to species or biological categories (Abe et al. 2003). In this study, the 15 germline DMR sequences of mouse imprinted genes were mapped on the SOMs, which were constructed using 2-kb genome sequences from 10 eukaryotes and 143 prokaryotes, to reveal characteristics of the germline DMRs.

I found that many DMRs are located at lattice points that are on the borders of the mouse territories and include sequences from more than one species. This indicates that these DMRs have features distinct from typical mouse sequences. In fact, *Meg1/Grb10* and *U2af1-rs1* DMRs were mapped to prokaryote territory in the DegeTetra-SOM. I also found that the distribution of the 3 paternally methylated DMRs (located rather centrally in the main mouse territory) differs from that of the 12 maternally methylated DMRs (clustered in the rightmost quarter of the SOM). The results indicate the SOM analysis may have the power to separate the maternally methylated DMRs from the paternally methylated DMRs. The distribution of the DMRs on the SOMs appeared to be dependant at least partly on the frequencies of CpG, ApT, TpA, TpApT+ApTpA, GpCpApA+TpTpGpC and other oligonucleotides containing CpG.

A possible link between epigenetic modifications of parasitic DNA and imprinting has been previously postulated (Barlow 1993; Yoder et al. 1997). One

piece of evidence comes from the fact that many mouse transgenes are imprinted by DNA methylation. Transgenes with prokaryotic sequences such as chloramphenicol acetyltransferase (CAT) (Reik et al. 1987), pBR322 (Hadchouel et al. 1987; Sapienza et al. 1987; Swain et al. 1987), and pUC18 (Sasaki et al. 1991) sequences, show methylation imprinting upon maternal transmission. Furthermore, such methylation is established in growing oocytes (Ueda et al. 1992), just like that of the endogenous DMRs (Lucifero et al. 2004). Endogenous DMRs may look like foreign DNA and, because of this, they may become methylated in the germline by the host defense machinery. The fact that the *Meg1/Grb10* and *U2af1-rs1* DMRs are mapped to the prokaryote territory in the DegeTetra-SOM provides a new piece of evidence for this interesting hypothesis.

It is also known that many imprinted genes (at least 8 genes) are derived by transposition events (Walter and Paulsen 2003). Two of them (*Peg10* and *Rtl1*) are in fact derived from retrotransposons. Because the SOM analysis can identify horizontally transferred “alien” sequences (Kanaya et al. 2001), I thought that it may be possible to identify the retrotransposon-derived DMRs. However, among the three, only *Peg10* is known to have a germline DMR (Ono et al. 2003) and this DMR was mapped in the CpG island area of the mouse territory. This is most probably consistent with the fact that this DMR is located 6.5 kb away from the retrotransposon-derived exon. Therefore, whether the DMRs associated with the retrotransposon-derived imprinted genes have

features distinct from the mouse genome sequences remains an open question.

In conclusion, the SOM analyses revealed that the germline DMRs on the whole have similar oligonucleotide patterns, which overlap with those of the genomes from other species. Specifically, some germline DMRs had a feature specific to prokaryotic DNA on the DegeTetra-SOM. Further studies are needed to establish whether the parasitic DNA-like sequences and the host defense mechanism are indeed involved in genomic imprinting.

Conclusions

In this thesis, I determined the boundaries and extents of the 15 mouse germline DMRs by bisulfite sequencing in 12.5-dpc embryos and sperm. The results showed that the average size of the DMRs is 2.7 kb and that their average G+C content is 54.2%. I also found that the DMRs have several different methylation patterns at the boundaries.

Then, oligonucleotide content analyses of the determined DMR sequences revealed that the DMRs show a content value intermediate between that of the whole genome and CpG islands for most oligonucleotides. I also found that the paternally methylated DMRs contain less CpGs than the maternally methylated DMRs. One possible explanation for this sexual dimorphism is that the paternally methylated DMRs are more mutable than the maternally methylated DMRs. However, it is also possible that the differential CpG content is one of the features recognized by the *de novo* methylation machinery in the germline. Some oligonucleotides such as TpGpC+GpCpA, GpCpApA+TpTpGpC and TpGpCpA were overrepresented in the DMRs, but their biological significance is currently unknown.

Using the germline DMR sequences, I also carried out SOM analyses and found that many germline DMRs have features distinct from typical mouse sequences. For example, some DMRs such as the *Meg1/Grb10* and *U2af1-rs1* DMRs had prokaryote-like sequence features.

Altogether, these results provide a basis to identify the structural

characteristics specific to the germline DMRs. My next goal is to determine the methylation status of the germline DMRs in oocytes and study the characteristics of the germline DMRs in more detail. Further studies on for the DMRs should give us clues to the understanding of the molecular mechanisms of genomic imprinting.

Acknowledgments

I am deeply gategul to Prof. Hiroyuki Sasaki for his help, advice and discussion throughout this work. I am deeply gategul to Drs. Takashi Sado and Kenichiro Hata for their advice and discussion. I am deeply gategul to Prof. Toshimichi Ikemura and Dr. Takashi Abe for oligonucleotide content analysis and SOM analysis. I gratefully thank Prof. Yuji Kohara, Mr. Tadasu Shin-I and Ms. Kazuko Oishi for DNA sequencing. I also thank Dr. Yayoi Obata and Prof. Tomohiro Kono for communicating their results prior to publication. I would like to express my sincere appreciation to all the members of Prof. Sasaki's laboratory.

References

Abe, T., S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, and T. Ikemura. 2002. A novel bioinformatic strategy for unveiling hidden genome signatures of eukaryotes: self-organizing map of oligonucleotide frequency. *Genome Inform Ser Workshop Genome Inform* **13:** 12-20.

Abe, T., S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, and T. Ikemura. 2003. Informatics for unveiling hidden genome signatures. *Genome Res* **13:** 693-702.

Arnaud, P., D. Monk, M. Hitchins, E. Gordon, W. Dean, C.V. Beechey, J. Peters, W. Craigen, M. Preece, P. Stanier, G.E. Moore, and G. Kelsey. 2003. Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Hum Mol Genet* **12:** 1005-1019.

Barlow, D.P. 1993. Methylation and imprinting: from host defense to gene regulation? *Science* **260:** 309-310.

Bartolomei, M.S., A.L. Webber, M.E. Brunkow, and S.M. Tilghman. 1993. Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev* **7:** 1663-1673.

Bolshoy, A. 1995. CC dinucleotides contribute to the bending of DNA in chromatin. *Nat Struct Biol* **2:** 446-448.

Bourc'his, D., G.L. Xu, C.S. Lin, B. Bollman, and T.H. Bestor. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294:** 2536-2539.

Cattanach, B.M., C.V. Beechey, and J. Peters. 2004. Interactions between imprinting effects in the mouse. *Genetics* **168:** 397-413.

Coombes, C., P. Arnaud, E. Gordon, W. Dean, E.A. Coar, C.M. Williamson, R. Feil, J. Peters, and G. Kelsey. 2003. Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Mol Cell Biol* **23:** 5475-5488.

Davis, T.L., J.M. Trasler, S.B. Moss, G.J. Yang, and M.S. Bartolomei. 1999. Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* **58:** 18-28.

Engemann, S., M. Strodicke, M. Paulsen, O. Franck, R. Reinhardt, N. Lane, W. Reik, and J. Walter. 2000. Sequence and functional comparison in the Beckwith-Wiedemann region: implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet* **9:** 2691-2706.

Ferguson-Smith, A.C., H. Sasaki, B.M. Cattanach, and M.A. Surani. 1993. Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* **362:** 751-755.

Fitzpatrick, G.V., P.D. Soloway, and M.J. Higgins. 2002. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* **32:** 426-431.

Frommer, M., L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy, and C.L. Paul. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* **89:** 1827-1831.

Gardiner-Garden, M. and M. Frommer. 1987. CpG islands in vertebrate genomes. *J Mol Biol* **196:** 261-282.

Gentles, A.J. and S. Karlin. 2001. Genome-scale compositional comparisons in eukaryotes. *Genome Res* **11:** 540-546.

Greally, J.M. 2002. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* **99:** 327-332.

Hadchouel, M., H. Farza, D. Simon, P. Tiollais, and C. Pourcel. 1987. Maternal inhibition of hepatitis B surface antigen gene expression in transgenic mice correlates with de novo methylation. *Nature* **329:** 454-456.

Handa, V. and A. Jeltsch. 2005. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J Mol Biol* **348:** 1103-1112.

Hata, K., M. Okano, H. Lei, and E. Li. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129:** 1983-1993.

Hayashizaki, Y., H. Shibata, S. Hirotsune, H. Sugino, Y. Okazaki, N. Sasaki, K. Hirose, H. Imoto, H. Okuizumi, M. Muramatsu, and et al. 1994. Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nat Genet* **6:** 33-40.

Hikichi, T., T. Kohda, T. Kaneko-Ishino, and F. Ishino. 2003. Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res* **31:** 1398-1406.

Holliday, R. and G.W. Grigg. 1993. DNA methylation and mutation. *Mutat Res* **285:** 61-67.

Hurst, L.D., G. McVean, and T. Moore. 1996. Imprinted genes have few and small introns. *Nat Genet* **12:** 234-237.

Kanaya, S., M. Kinouchi, T. Abe, Y. Kudo, Y. Yamada, T. Nishi, H. Mori, and T.

Ikemura. 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome. *Gene* **276:** 89-99.

Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, and H. Sasaki. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429:** 900-903.

Karlin, S. 1998. Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr Opin Microbiol* **1:** 598-610.

Karlin, S., A.M. Campbell, and J. Mrazek. 1998. Comparative DNA analysis across diverse genomes. *Annu Rev Genet* **32:** 185-225.

Ke, X., N.S. Thomas, D.O. Robinson, and A. Collins. 2002. The distinguishing sequence characteristics of mouse imprinted genes. *Mamm Genome* **13:** 639-645.

Kelsey, G., D. Bodle, H.J. Miller, C.V. Beechey, C. Coombes, J. Peters, and C.M. Williamson. 1999. Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. *Genomics* **62:** 129-138.

Koide, T., K. Moriwaki, K. Uchida, A. Mita, T. Sagai, H. Yonekawa, H. Katoh, N. Miyashita, K. Tsuchiya, T.J. Nielsen, and T. Shiroishi. 1998. A new inbred strain JF1 established from Japanese fancy mouse carrying the classic piebald allele. *Mamm Genome* **9:** 15-19.

Lefebvre, L., S. Viville, S.C. Barton, F. Ishino, and M.A. Surani. 1997. Genomic structure and parent-of-origin-specific methylation of Peg1. *Hum Mol Genet* **6:** 1907-1915.

Li, E., C. Beard, and R. Jaenisch. 1993. Role for DNA methylation in genomic

imprinting. *Nature* **366:** 362-365.

Li, E., T.H. Bestor, and R. Jaenisch. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:** 915-926.

Li, J.Y., D.J. Lees-Murdock, G.L. Xu, and C.P. Walsh. 2004. Timing of establishment of paternal methylation imprints in the mouse. *Genomics* **84:** 952-960.

Li, L.L., I.Y. Szeto, B.M. Cattanach, F. Ishino, and M.A. Surani. 2000. Organization and parent-of-origin-specific methylation of imprinted Peg3 gene on mouse proximal chromosome 7. *Genomics* **63:** 333-340.

Lin, S.P., N. Youngson, S. Takada, H. Seitz, W. Reik, M. Paulsen, J. Cavaille, and A.C. Ferguson-Smith. 2003. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* **35:** 97-102.

Liu, J., M. Chen, C. Deng, D. Bourc'his, J.G. Nealon, B. Erlichman, T.H. Bestor, and L.S. Weinstein. 2005. Identification of the control region for tissue-specific imprinting of the stimulatory G protein alpha-subunit. *Proc Natl Acad Sci U S A* **102:** 5513-5518.

Liu, J., S. Yu, D. Litman, W. Chen, and L.S. Weinstein. 2000. Identification of a methylation imprint mark within the mouse Gnas locus. *Mol Cell Biol* **20:** 5808-5817.

Lopes, S., A. Lewis, P. Hajkova, W. Dean, J. Oswald, T. Forne, A. Murrell, M. Constancia, M. Bartolomei, J. Walter, and W. Reik. 2003. Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet* **12:** 295-305.

Lucifero, D., M.R. Mann, M.S. Bartolomei, and J.M. Trasler. 2004. Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum Mol Genet* **13:** 839-849.

Lucifero, D., C. Mertineit, H.J. Clarke, T.H. Bestor, and J.M. Trasler. 2002. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* **79:** 530-538.

McGrath, J. and D. Solter. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37:** 179-183.

Nelson, J.R., Y.C. Cai, T.L. Giesler, J.W. Farchaus, S.T. Sundaram, M. Ortiz-Rivera, L.P. Hosta, P.L. Hewitt, J.A. Mamone, C. Palaniappan, and C.W. Fuller. 2002. TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. *Biotechniques* **Suppl:** 44-47.

Neumann, B., P. Kubicka, and D.P. Barlow. 1995. Characteristics of imprinted genes. *Nat Genet* **9:** 12-13.

Nussinov, R. 1984. Doublet frequencies in evolutionary distinct groups. *Nucleic Acids Res* **12:** 1749-1763.

Okamura, K., Y. Hagiwara-Takeuchi, T. Li, T.H. Vu, M. Hirai, M. Hattori, Y. Sakaki, A.R. Hoffman, and T. Ito. 2000. Comparative genome analysis of the mouse imprinted gene impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting. *Genome Res* **10:** 1878-1889.

Ono, R., H. Shiura, H. Aburatani, T. Kohda, T. Kaneko-Ishino, and F. Ishino. 2003. Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. *Genome Res* **13:** 1696-1705.

Phillips, G.J., J. Arnold, and R. Ivarie. 1987. Mono- through hexanucleotide composition of the Escherichia coli genome: a Markov chain analysis. *Nucleic Acids Res* **15:** 2611-2626.

Plass, C., H. Shibata, I. Kalcheva, L. Mullins, N. Kotelevtseva, J. Mullins, R. Kato, H. Sasaki, S. Hirotsune, Y. Okazaki, W.A. Held, Y. Hayashizaki, and V.M. Chapman. 1996. Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat Genet* **14:** 106-109.

Reik, W., A. Collick, M.L. Norris, S.C. Barton, and M.A. Surani. 1987. Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328:** 248-251.

Reik, W. and J. Walter. 2001. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2:** 21-32.

Rocha, E.P., A. Viari, and A. Danchin. 1998. Oligonucleotide bias in Bacillus subtilis: general trends and taxonomic comparisons. *Nucleic Acids Res* **26:** 2971-2980.

Sapienza, C., A.C. Peterson, J. Rossant, and R. Balling. 1987. Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328:** 251-254.

Sasaki, H., T. Hamada, T. Ueda, R. Seki, T. Higashinakagawa, and Y. Sakaki. 1991. Inherited type of allelic methylation variations in a mouse chromosome region where an integrated transgene shows methylation imprinting. *Development* **111:** 573-581.

Shemer, R., Y. Birger, A.D. Riggs, and A. Razin. 1997. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proc Natl Acad Sci U S A* **94:** 10267-10272.

Shibata, H., T. Ueda, M. Kamiya, A. Yoshiki, M. Kusakabe, C. Plass, W.A. Held, S. Sunahara, M. Katsuki, M. Muramatsu, and Y. Hayashizaki. 1997. An oocyte-specific methylation imprint center in the mouse U2afbp-rs/U2af1-rs1 gene marks the establishment of allele-specific methylation during preimplantation development. *Genomics* **44:** 171-178.

Shibata, H., Y. Yoda, R. Kato, T. Ueda, M. Kamiya, N. Hiraiwa, A. Yoshiki, C. Plass, R.S. Pearsall, W.A. Held, M. Muramatsu, H. Sasaki, M. Kusakabe, and Y. Hayashizaki. 1998. A methylation imprint mark in the mouse imprinted gene Grf1/Cdc25Mm locus shares a common feature with the U2afbp-rs gene: an association with a short tandem repeat and a hypermethylated region. *Genomics* **49:** 30-37.

Smith, R.J., P. Arnaud, G. Konfortova, W.L. Dean, C.V. Beechey, and G. Kelsey. 2002. The mouse Zac1 locus: basis for imprinting and comparison with human ZAC. *Gene* **292:** 101-112.

Stoger, R., P. Kubicka, C.G. Liu, T. Kafri, A. Razin, H. Cedar, and D.P. Barlow. 1993. Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73:** 61-71.

Surani, M.A., S.C. Barton, and M.L. Norris. 1984. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308:** 548-550.

Swain, J.L., T.A. Stewart, and P. Leder. 1987. Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. *Cell* **50:** 719-727.

Takada, S., M. Paulsen, M. Tevendale, C.E. Tsai, G. Kelsey, B.M. Cattanach, and A.C. Ferguson-Smith. 2002. Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting

control from comparison with Igf2-H19. *Hum Mol Genet* **11:** 77-86.

Thorvaldsen, J.L., K.L. Duran, and M.S. Bartolomei. 1998. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev* **12:** 3693-3702.

Tremblay, K.D., J.R. Saam, R.S. Ingram, S.M. Tilghman, and M.S. Bartolomei. 1995. A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat Genet* **9:** 407-413.

Ueda, T., K. Abe, A. Miura, M. Yuzuriha, M. Zubair, M. Noguchi, K. Niwa, Y. Kawase, T. Kono, Y. Matsuda, H. Fujimoto, H. Shibata, Y. Hayashizaki, and H. Sasaki. 2000. The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* **5:** 649-659.

Ueda, T., K. Yamazaki, R. Suzuki, H. Fujimoto, H. Sasaki, Y. Sakaki, and T. Higashinakagawa. 1992. Parental methylation patterns of a transgenic locus in adult somatic tissues are imprinted during gametogenesis. *Development* **116:** 831-839.

Walter, J. and M. Paulsen. 2003. The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum Mol Genet* **12 Spec No 2:** R215-220.

Williamson, C.M., S.T. Ball, W.T. Nottingham, J.A. Skinner, A. Plagge, M.D. Turner, N. Powles, T. Hough, D. Papworth, W.D. Fraser, M. Maconochie, and J. Peters. 2004. A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. *Nat Genet* **36:** 894-899.

Wutz, A., O.W. Smrzka, N. Schweifer, K. Schellander, E.F. Wagner, and D.P. Barlow. 1997. Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389:** 745-749.

Yatsuki, H., K. Joh, K. Higashimoto, H. Soejima, Y. Arai, Y. Wang, I. Hatada, Y. Obata, H. Morisaki, Z. Zhang, T. Nakagawachi, Y. Satoh, and T. Mukai. 2002. Domain regulation of imprinting cluster in Kip2/Lit1 subdomain on mouse chromosome 7F4/F5: large-scale DNA methylation analysis reveals that DMR-Lit1 is a putative imprinting control region. *Genome Res* **12:** 1860-1870.

Yoder, J.A., C.P. Walsh, and T.H. Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13:** 335-340.

Yoon, B.J., H. Herman, A. Sikora, L.T. Smith, C. Plass, and P.D. Soloway. 2002. Regulation of DNA methylation of Rasgrf1. *Nat Genet* **30:** 92-96.

Website references

http://genome.ucsc.edu/; UCSC Genome Bioinformatics Site

http://www.ncbi.nlm.nih.gov/; GenBank (NCBI)

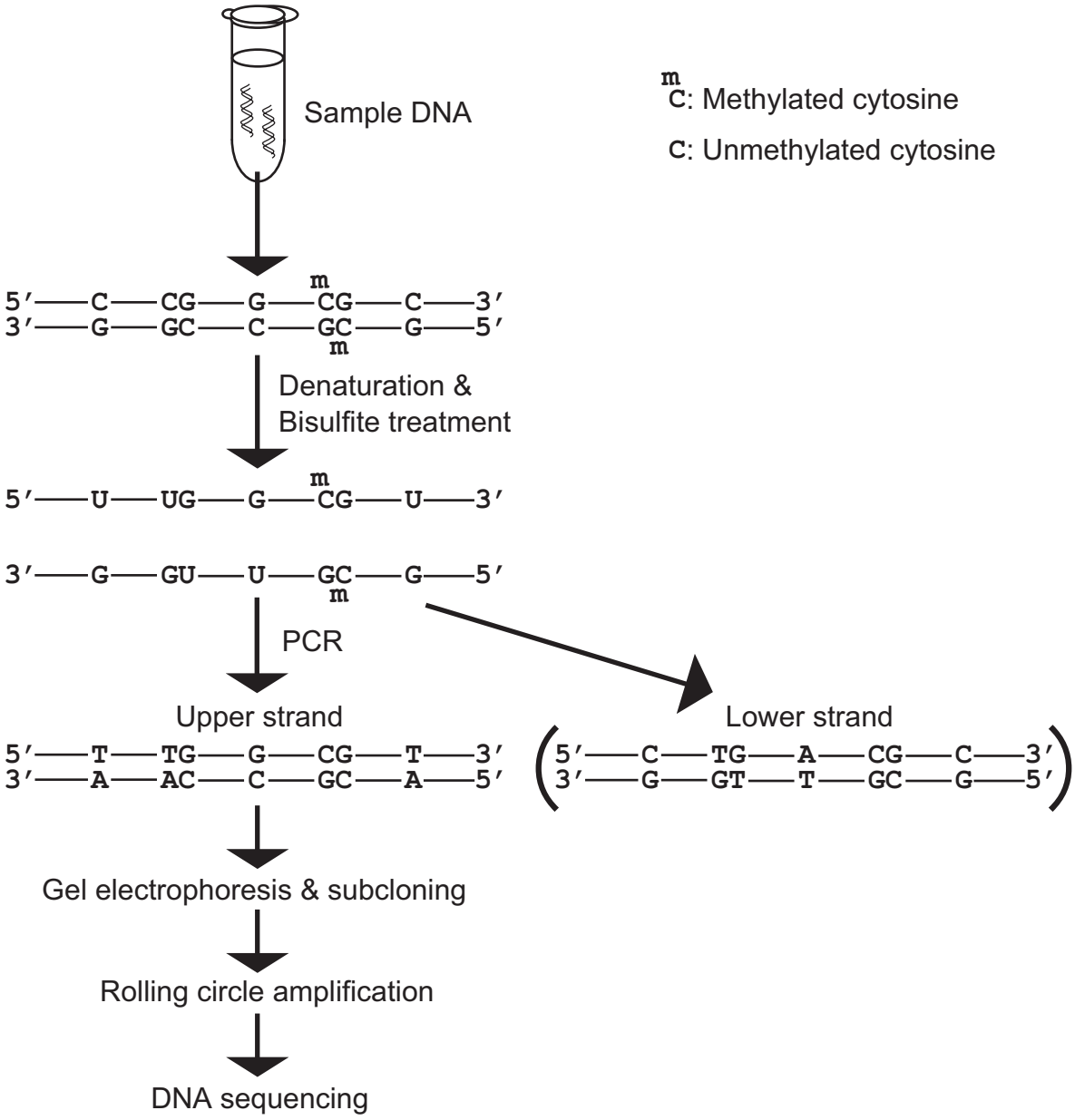

Figure 1. Scheme of bisulfite sequencing analysis

Figure 1. Scheme of bisulfite sequencing analysis

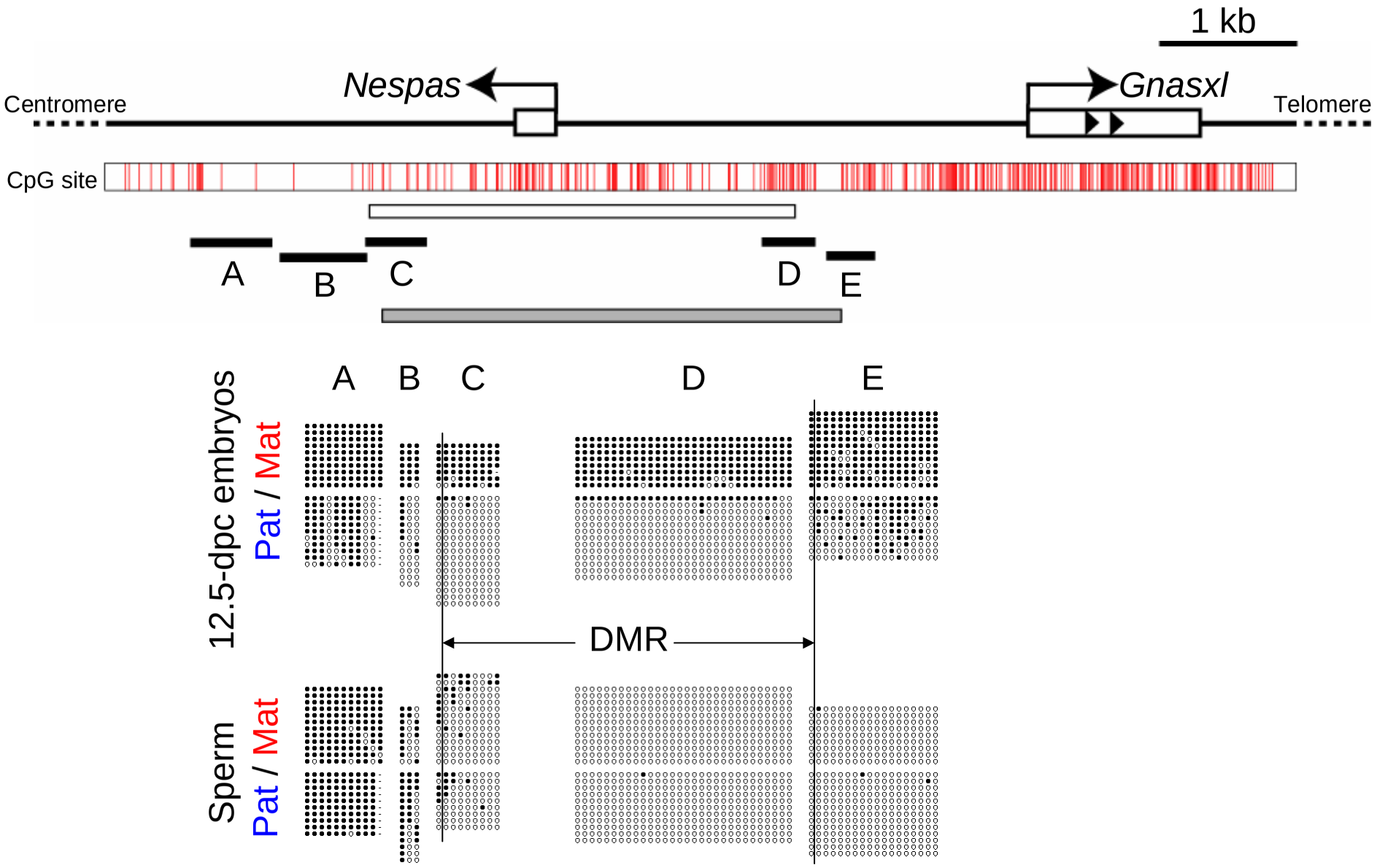

Figure 2 (a) Bisulfite methylation analysis of the *Nespas-Gnasxl* DMR

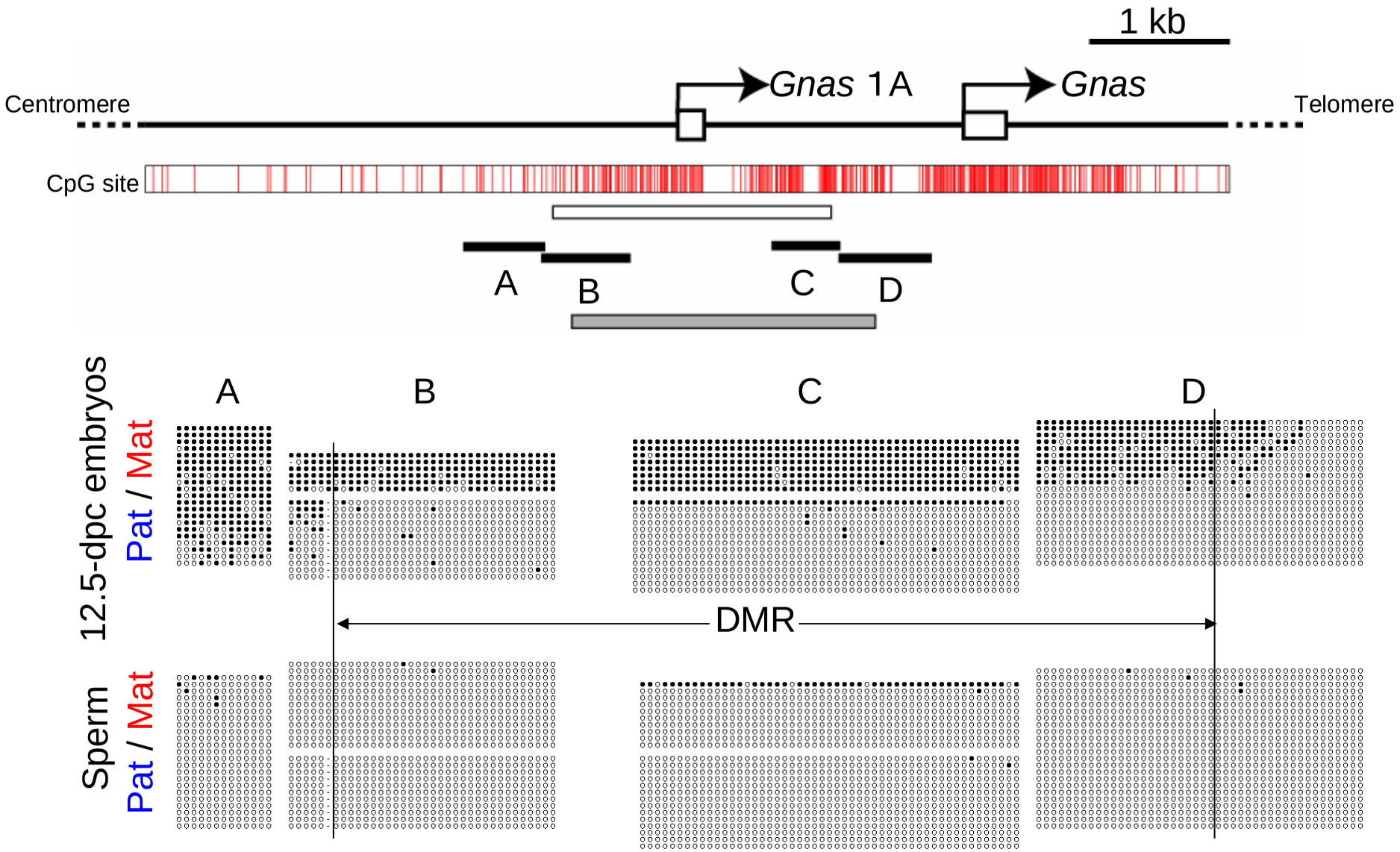

Figure 2 (b) Bisulfite methylation analysis of the *Gnas* 1A DMR

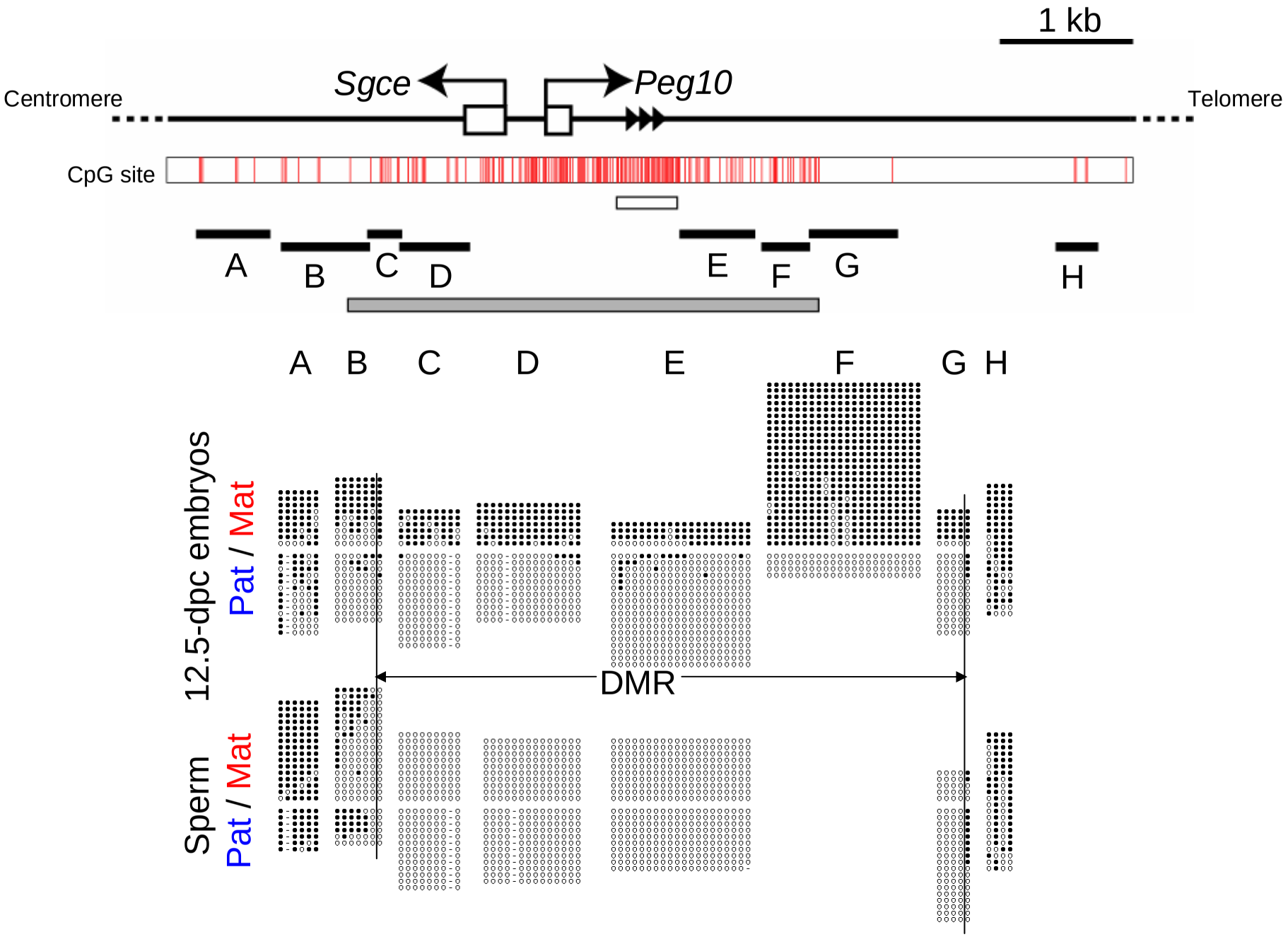

Figure 2 (c) Bisulfite methylation analysis of the *Peg10* DMR

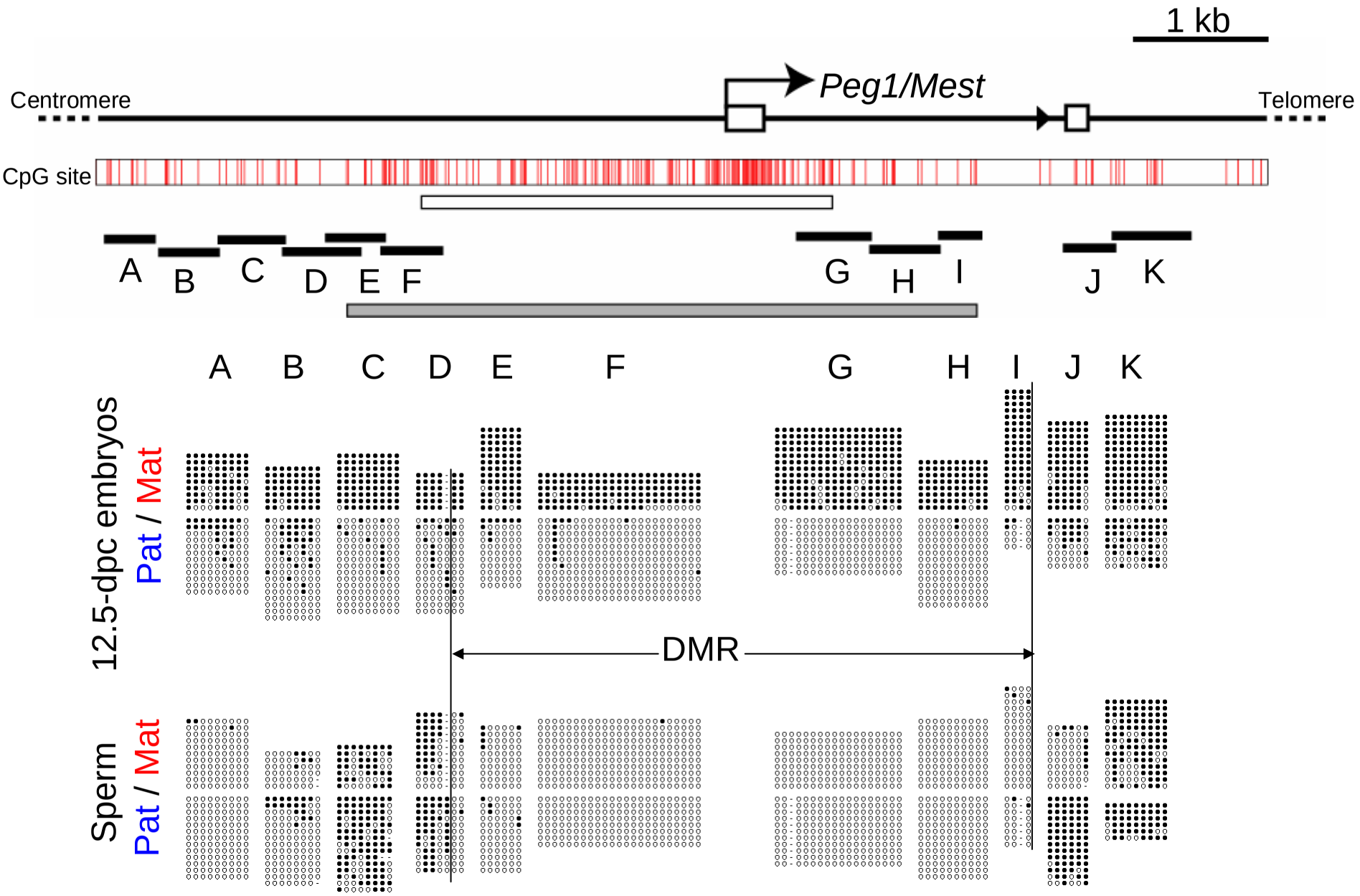

Figure 2 (d) Bisulfite methylation analysis of the *Peg1/Mest* DMR

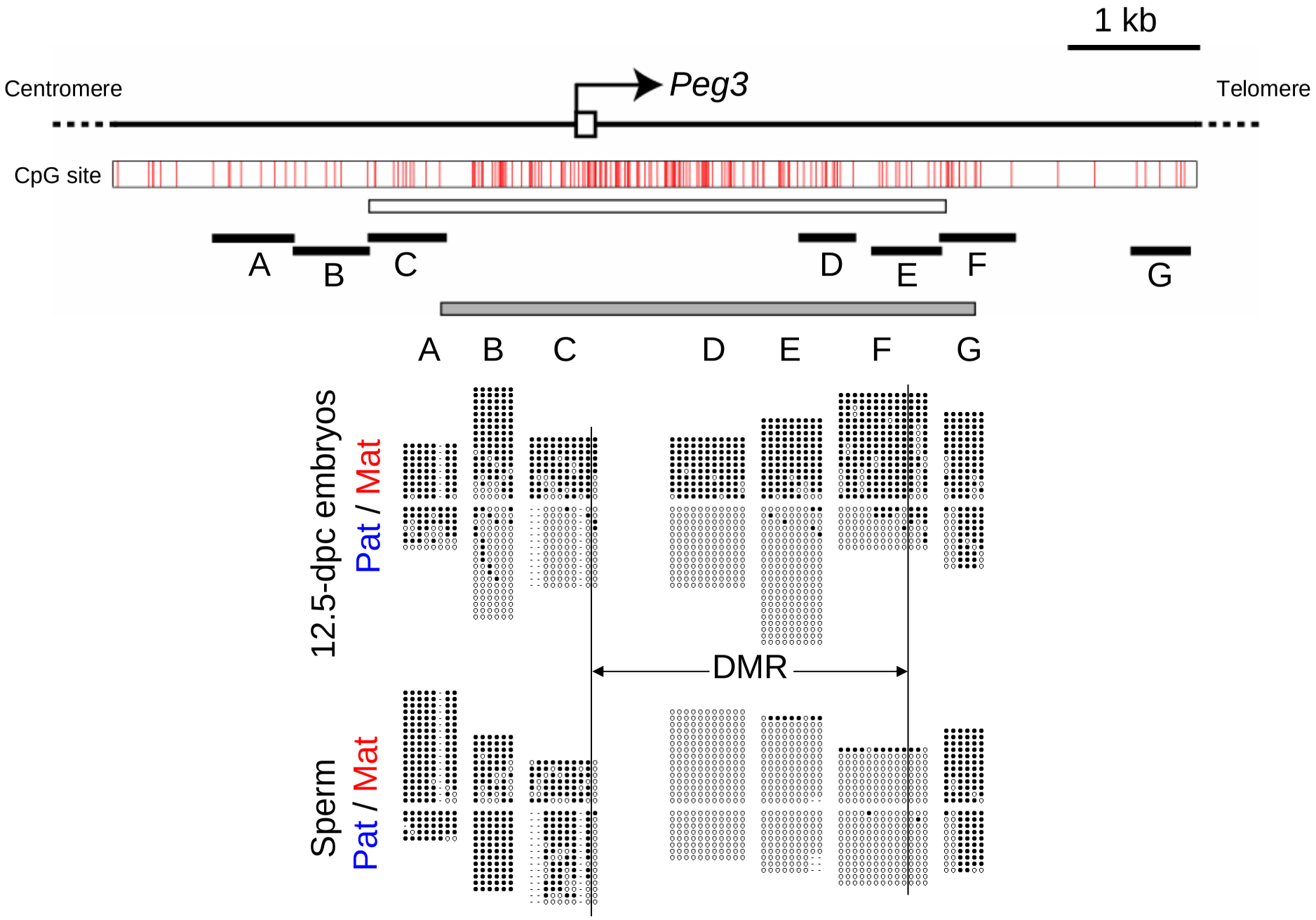

Figure 2 (e) Bisulfite methylation analysis of the *Peg3* DMR

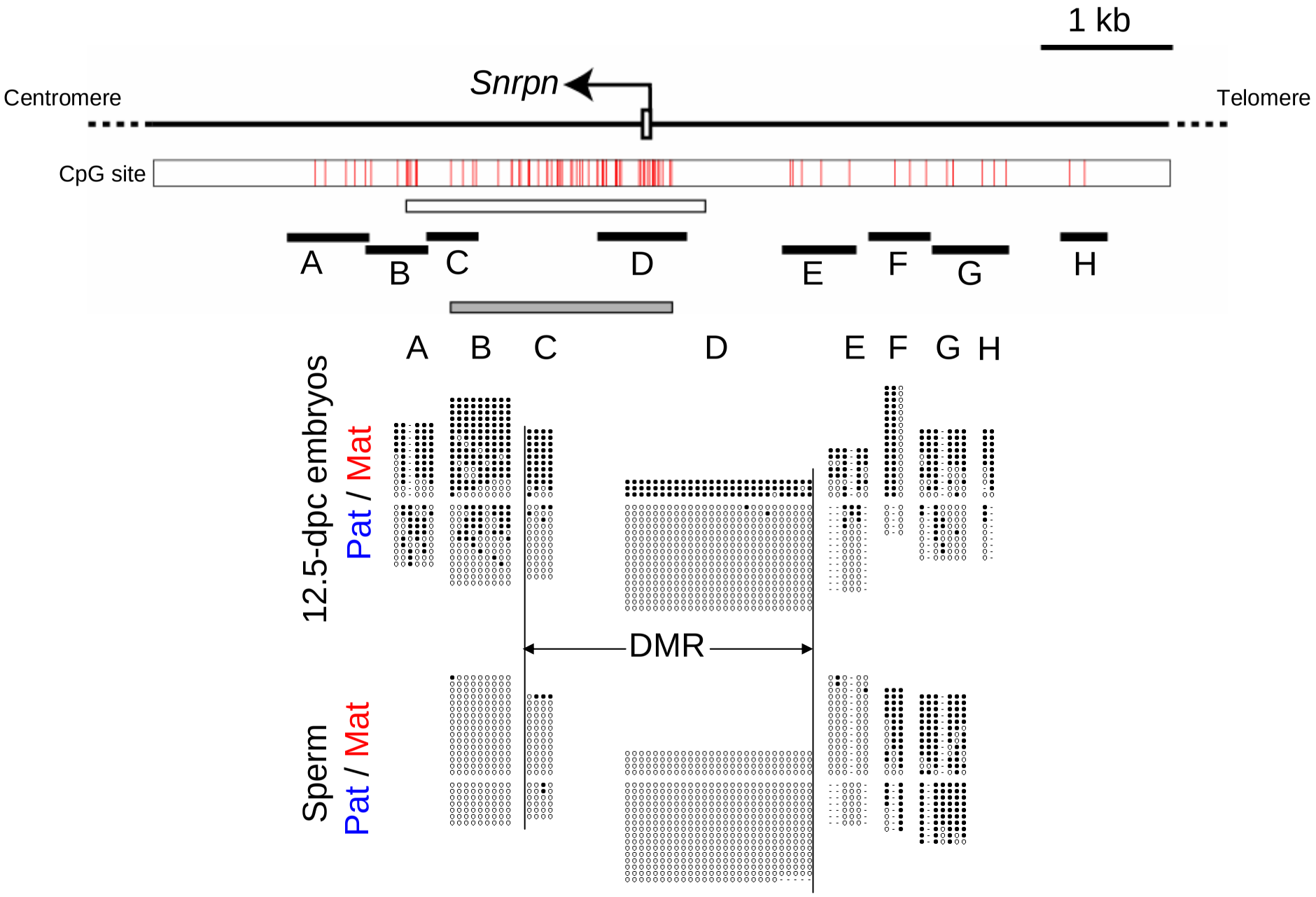

Figure 2 (f) Bisulfite methylation analysis of the *Snrpn* DMR1

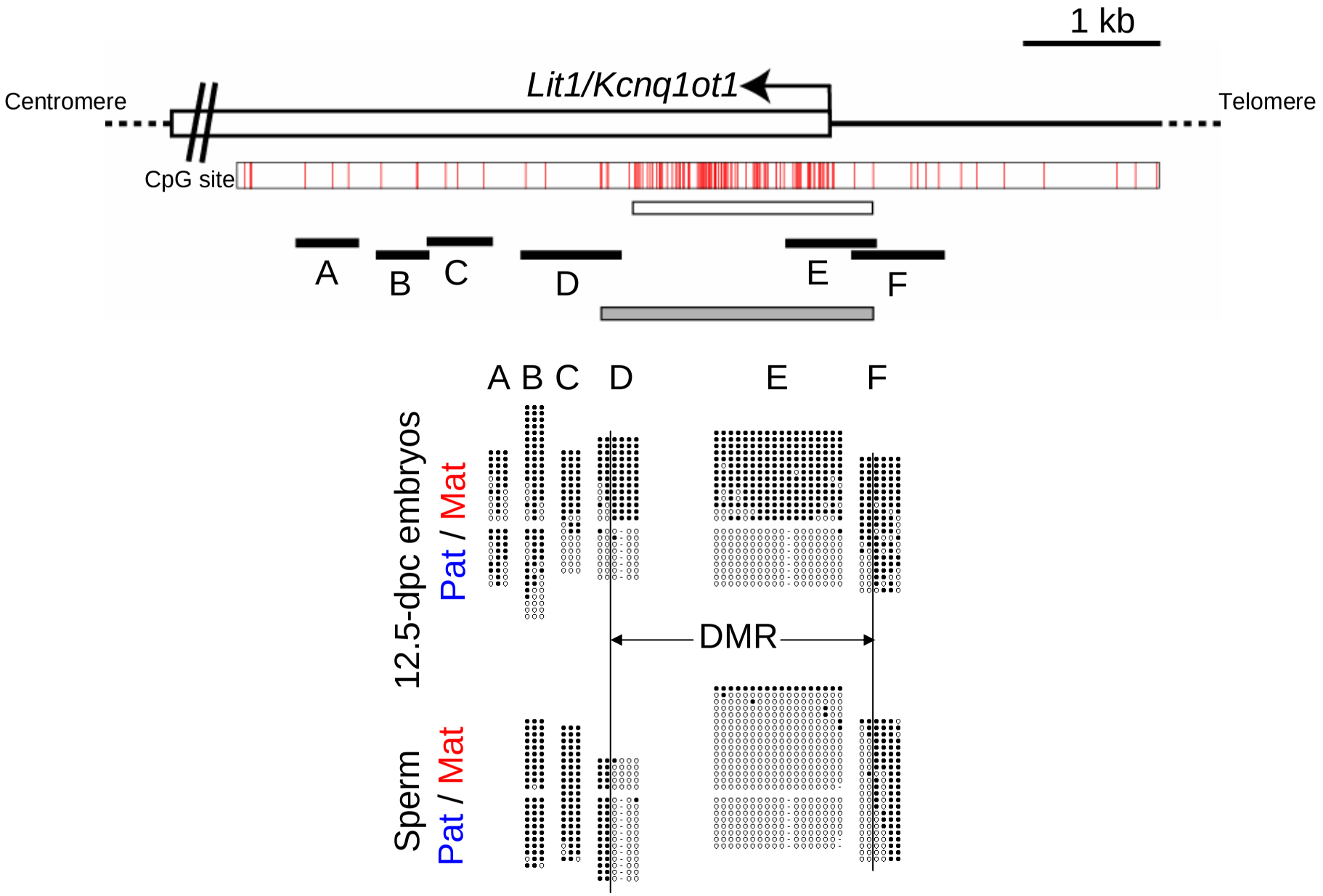

Figure 2 (g) Bisulfite methylation analysis of the *Lit1/Kcnq1ot1* DMR

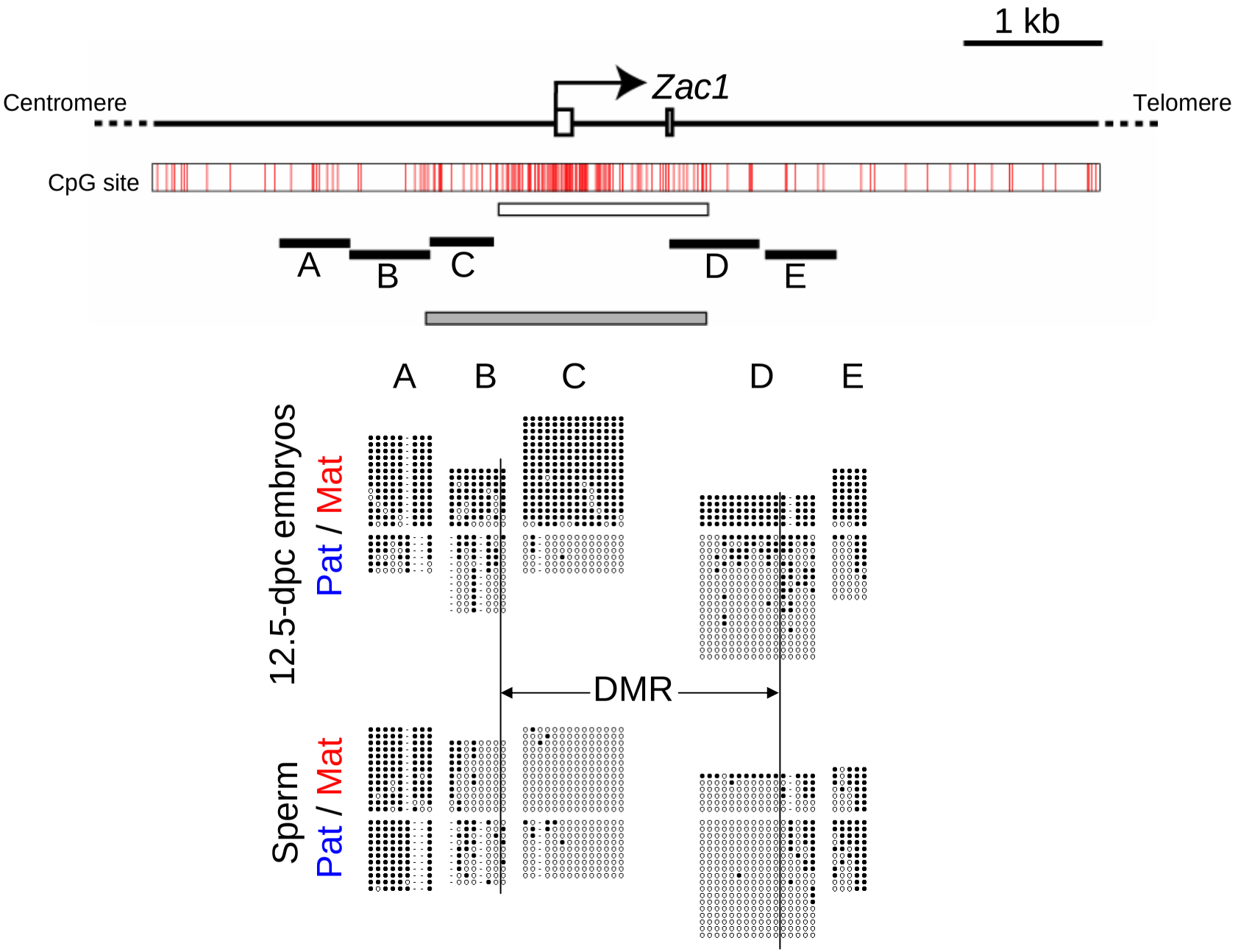

Figure 2 (h) Bisulfite methylation analysis of the *Zac1* DMR

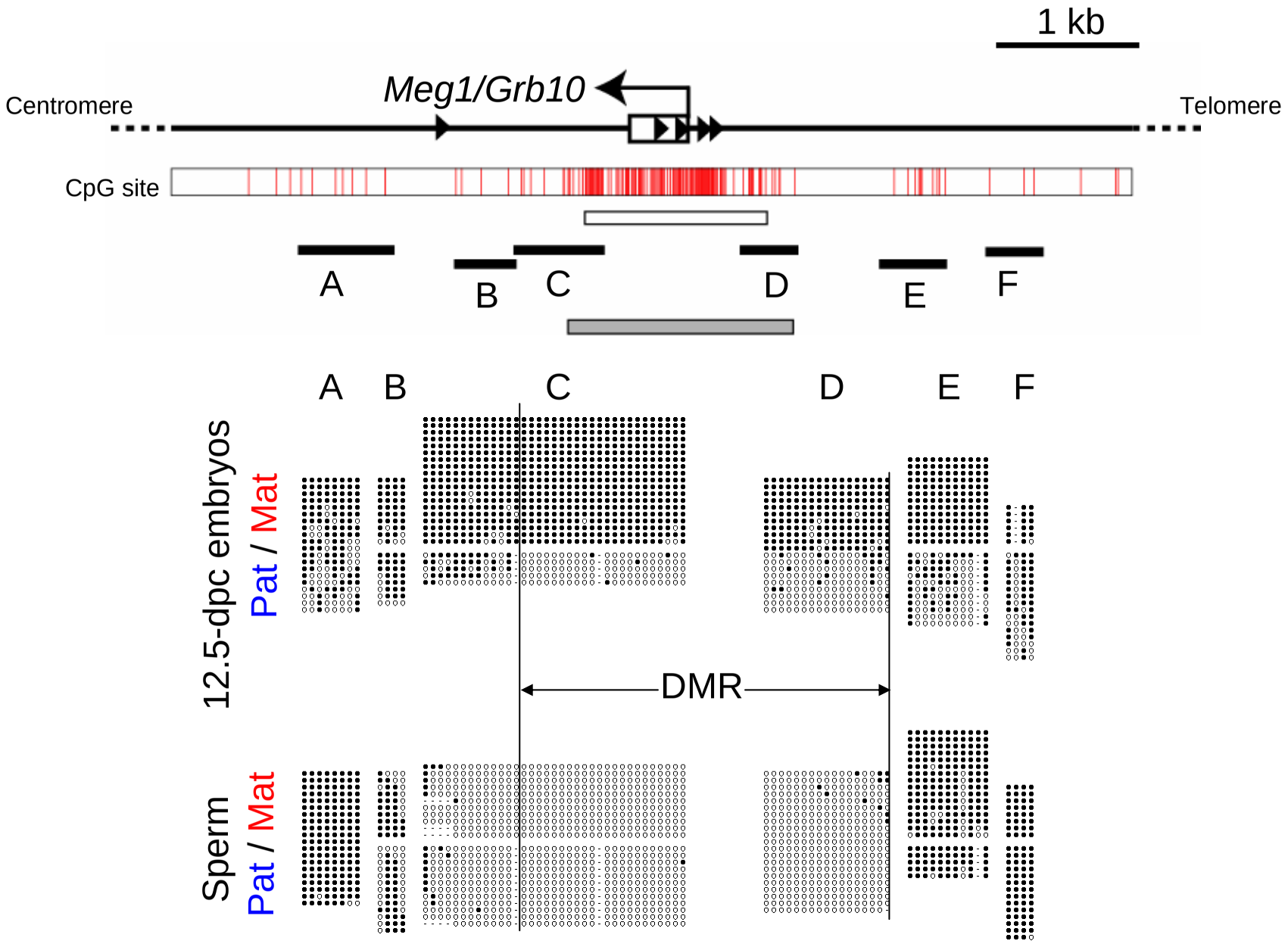

Figure 2 (i) Bisulfite methylation analysis of the *Meg1/Grb10* DMR

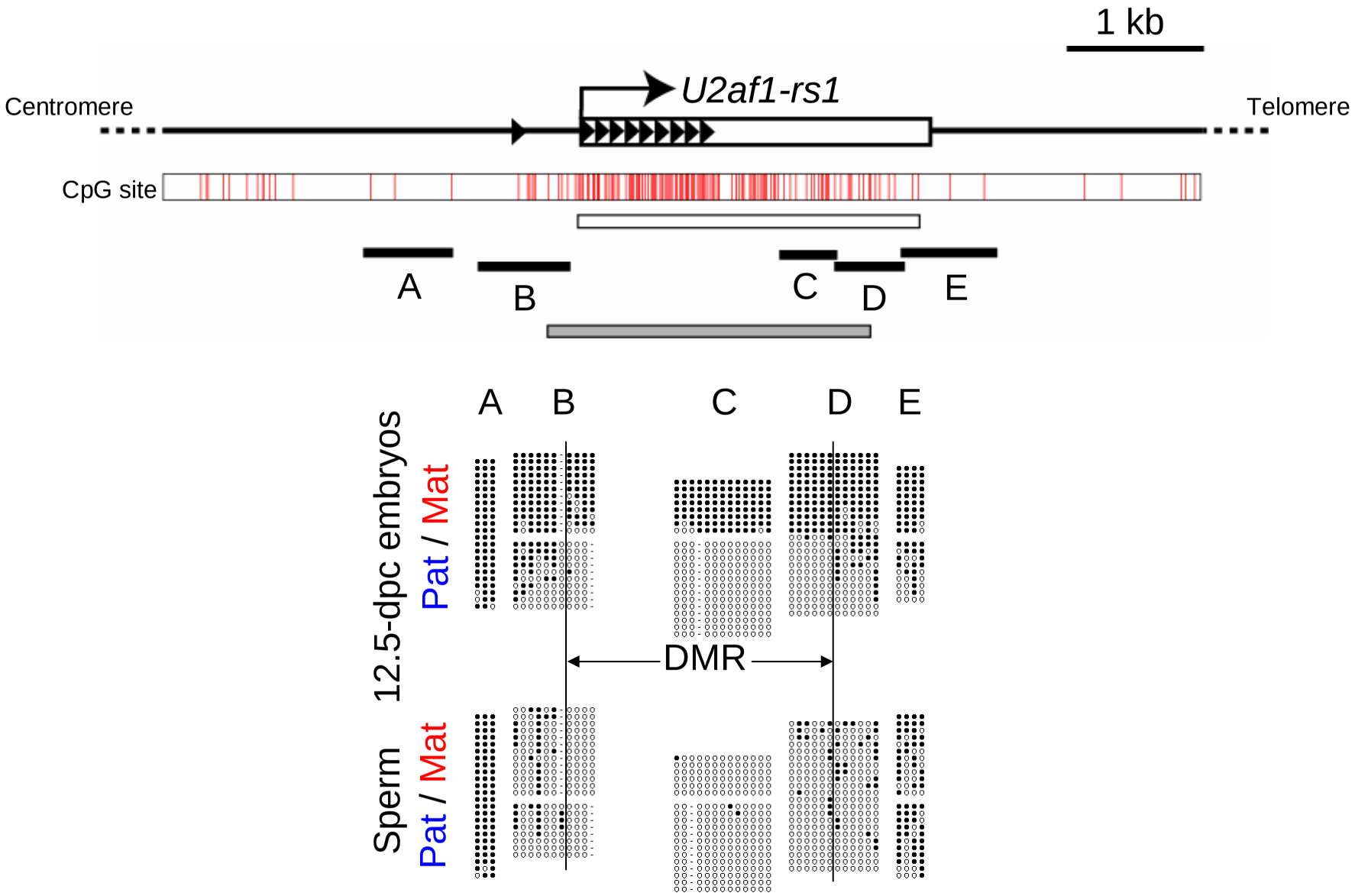

Figure 2 (j) Bisulfite methylation analysis of the *U2af1-rs1* DMR

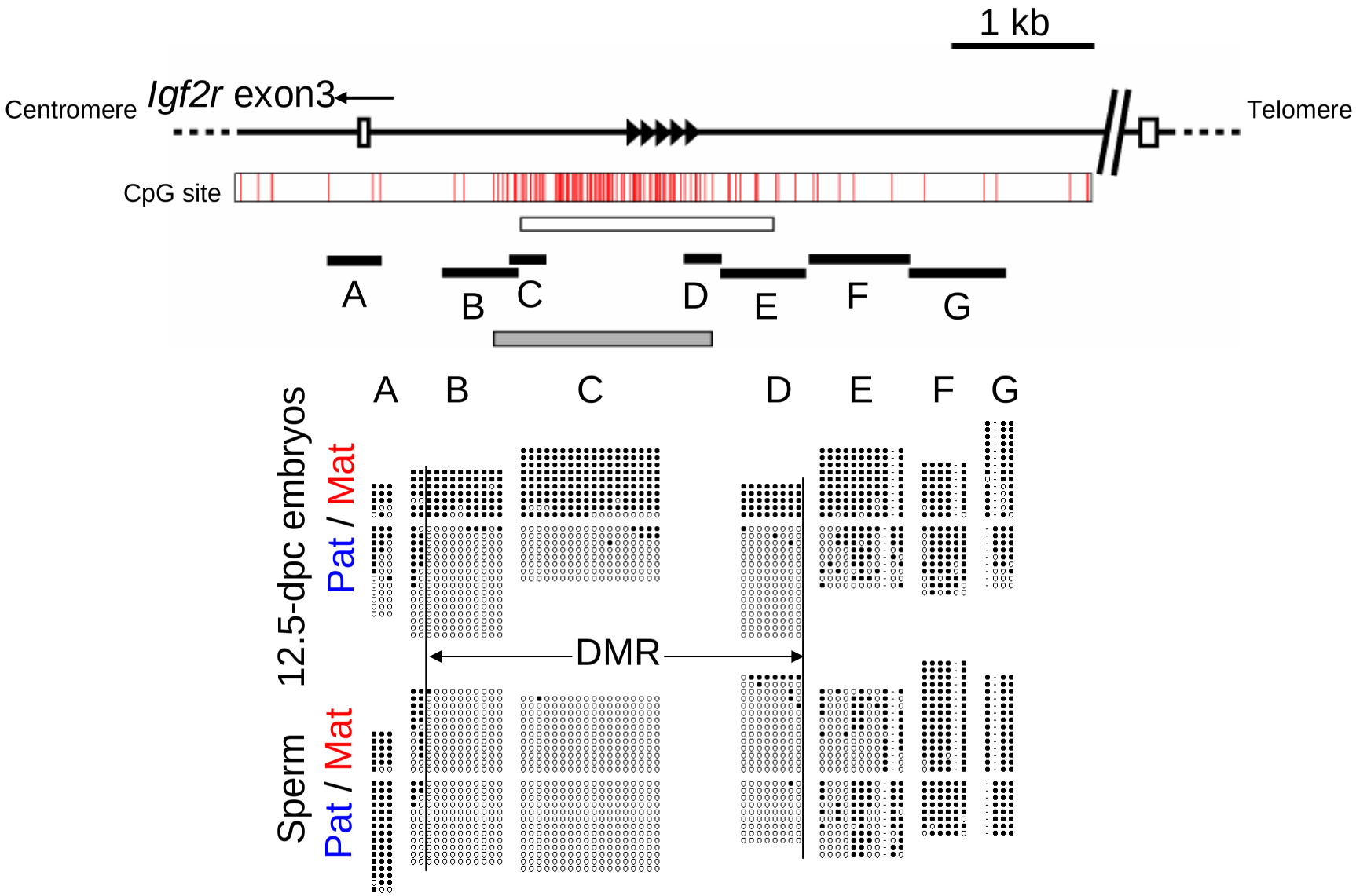

Figure 2 (k) Bisulfite methylation analysis of the *Igf2r* DMR2

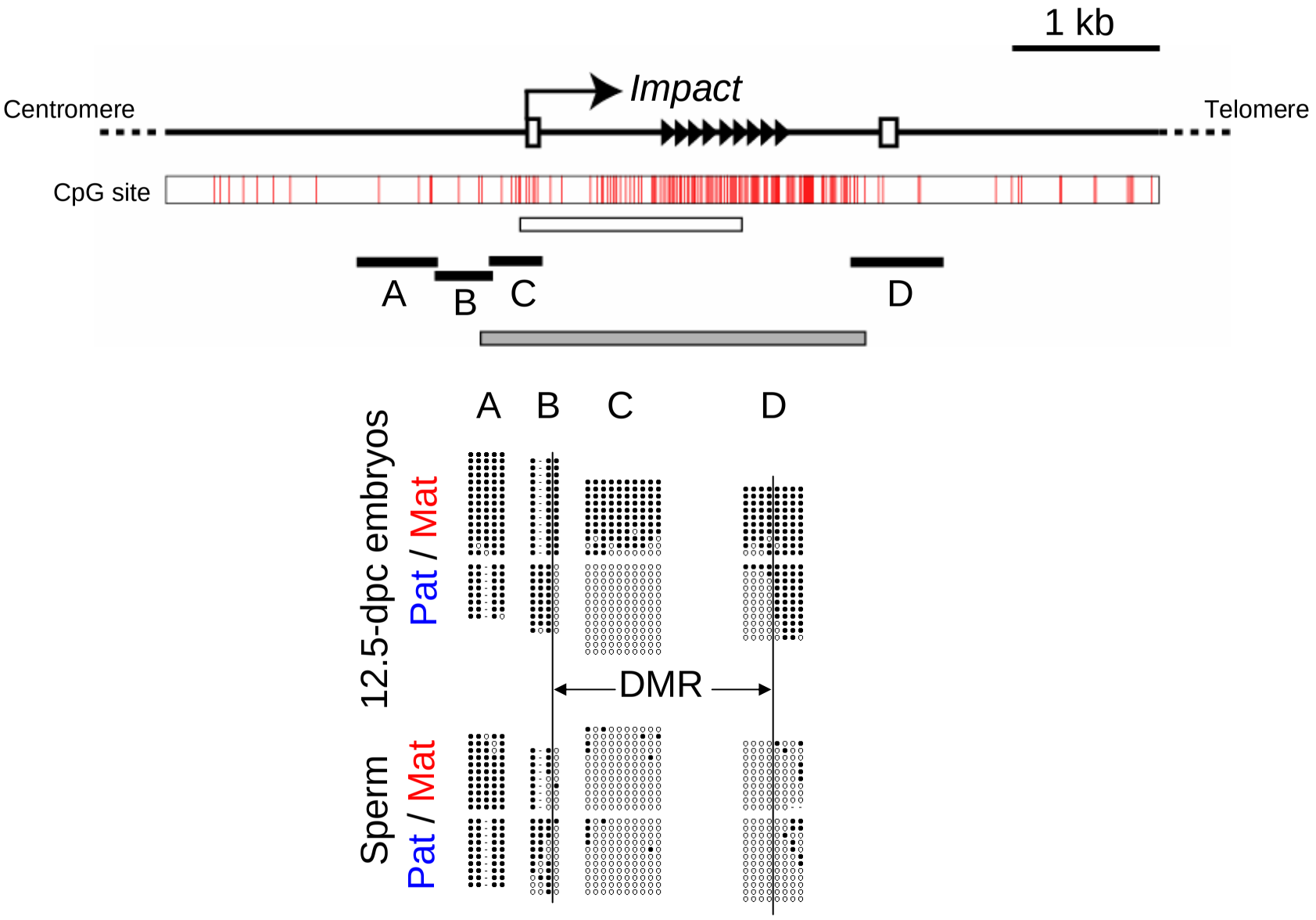

Figure 2 (l) Bisulfite methylation analysis of the *Impact* DMR

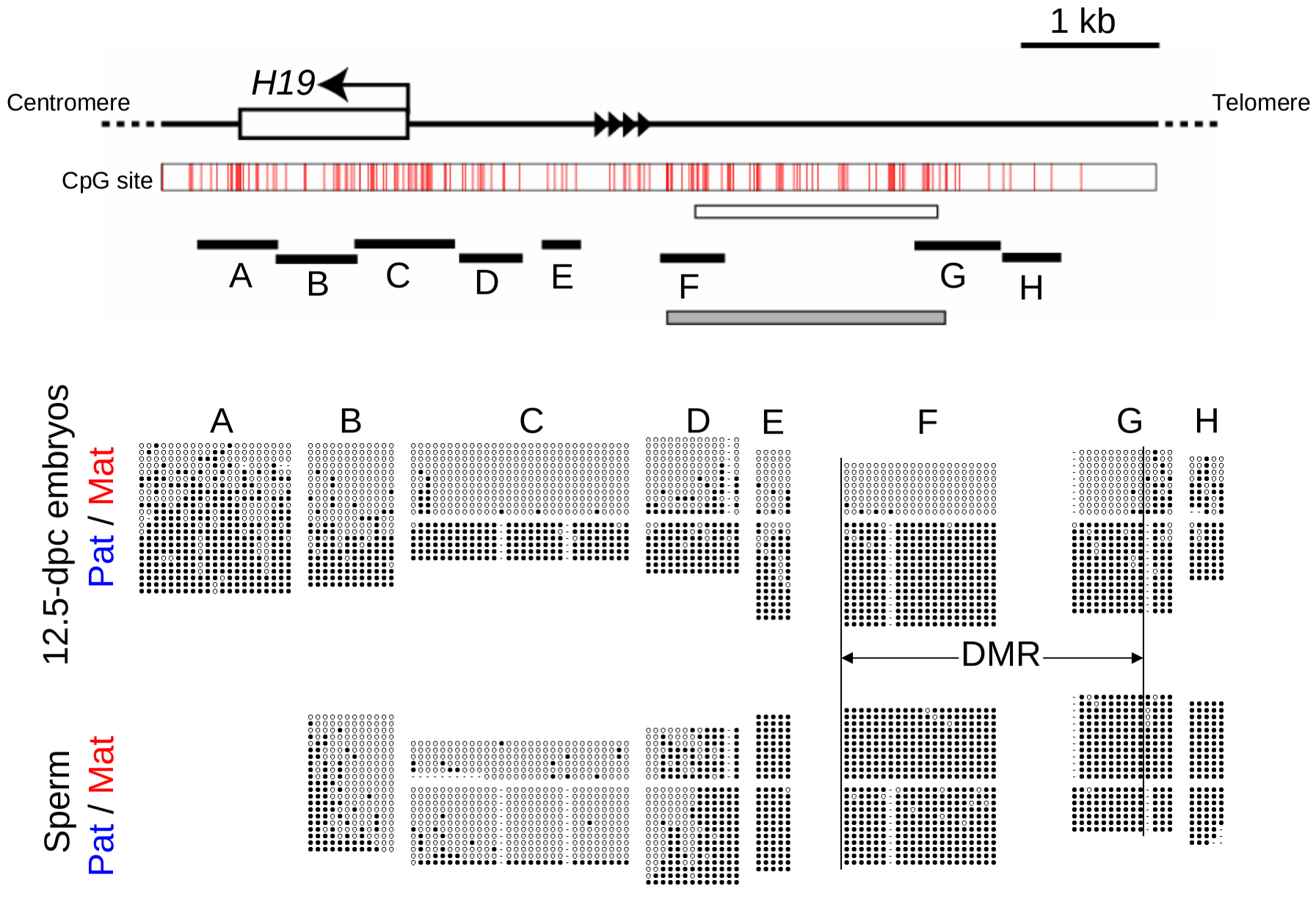

Figure 2 (m) Bisulfite methylation analysis of the *H19* DMR

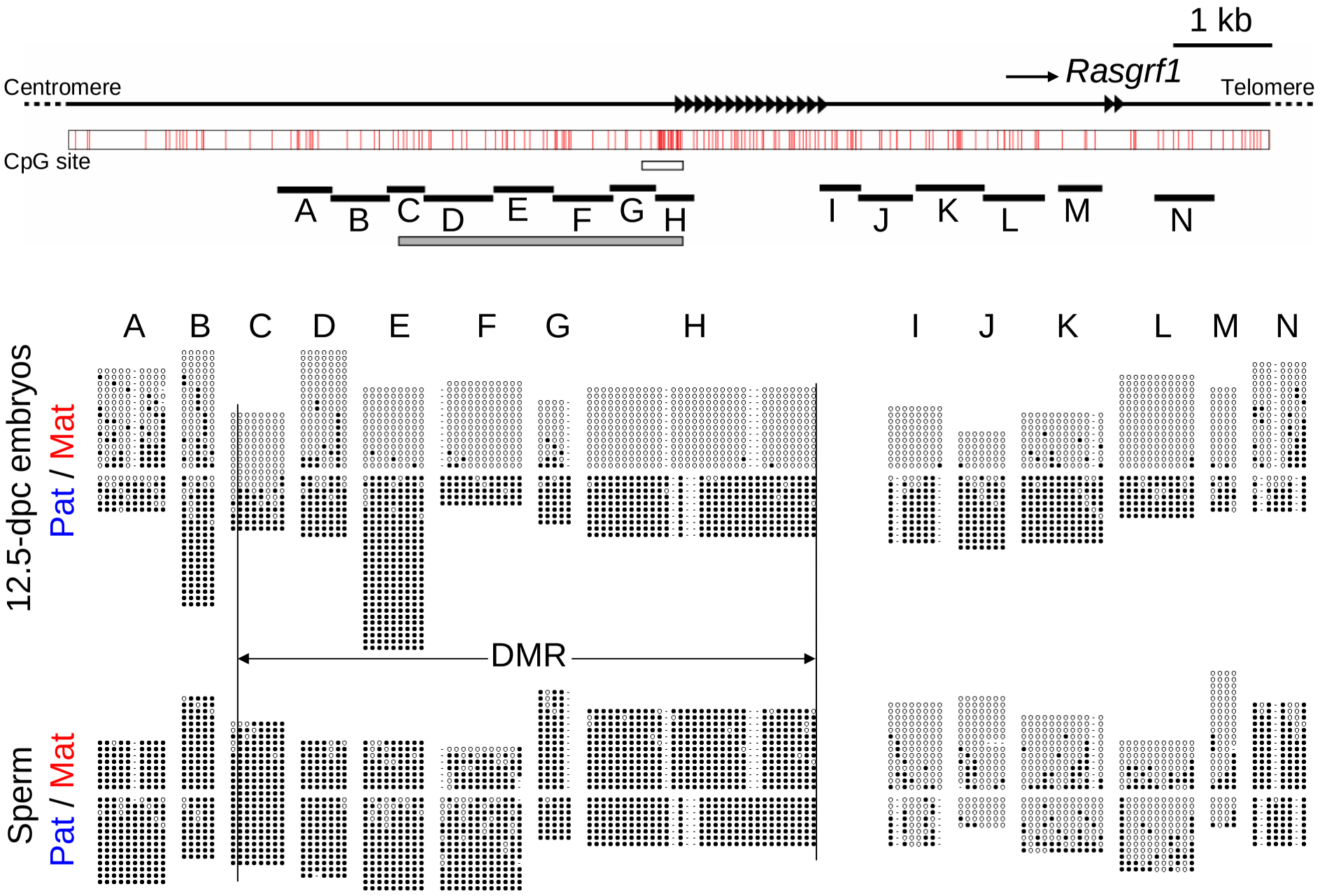

Figure 2 (n) Bisulfite methylation analysis of the *Rasgrf1* DMR

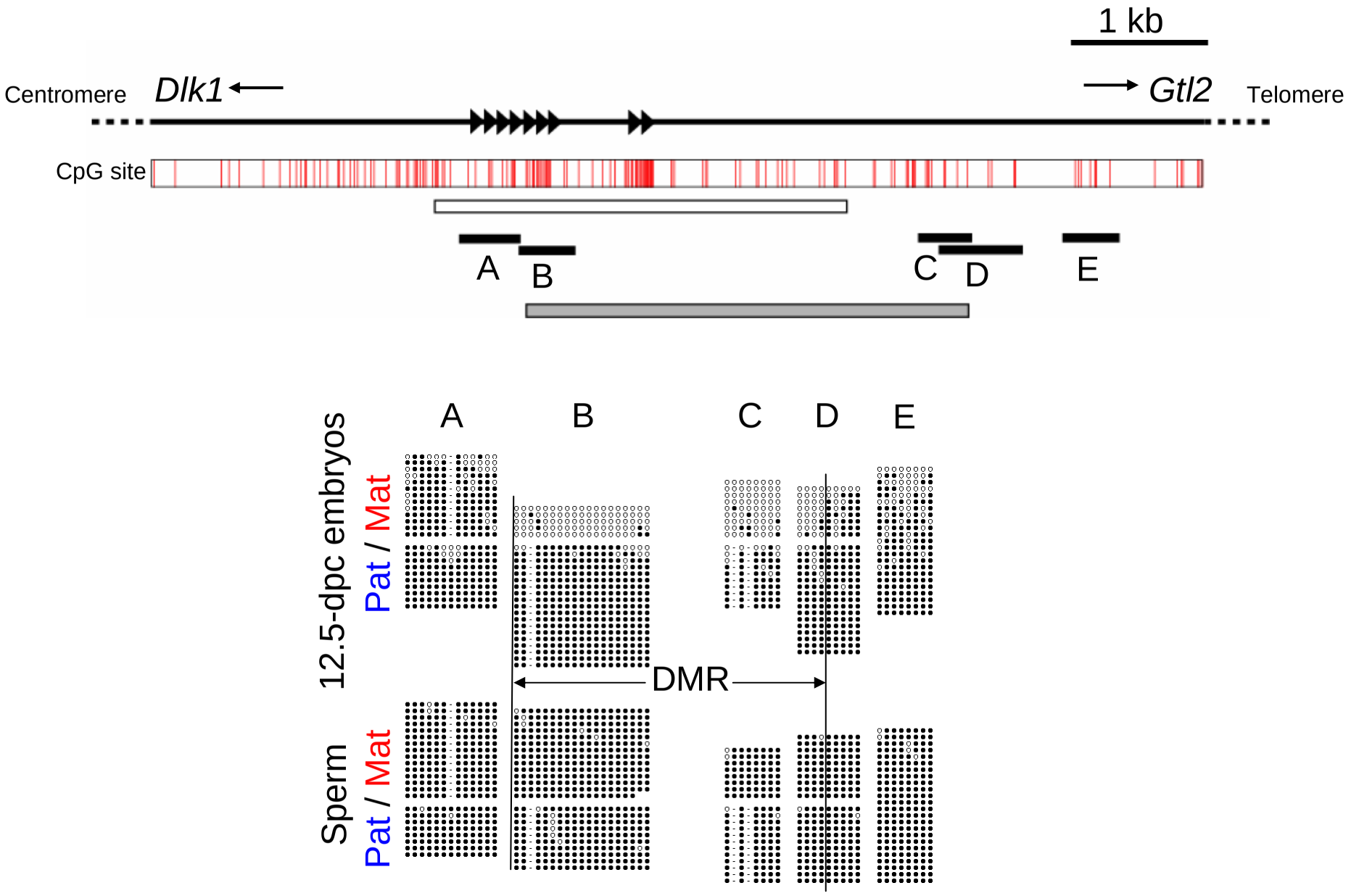

Figure 2 (o) Bisulfite methylation analysis of the IG-DMR

Figure 2. Bisulfite methylation analysis of the 15 germline DMRs in 12.5-dpc mouse embryos and sperm: (a) *Nespas-Gnasxl* locus; (b) *Gnas* 1A locus; (c) *Peg10*; (d) *Peg1/Mest* locus; (e) *Peg3* locus; (f) *Snrpn* locus; (g) *Lit1/Kcnq1ot1* locus; (h) *Zac1* locus; (i), *Meg1/Grb10* locus; (j) *U2af1-rs1* locus; (k) *Igf2r* locus; (l) *Impact* locus; (m) *H19* locus; (n) *Rasgrf1* locus; (o) *Dlk1-Gtl2* locus. Open boxes represent the gene exons. Arrows indicate the transcription start sites and directions of transcription. Filled triangles represent tandem repeats. Red vertical lines below the exon-intron organization indicate the positions of CpGs. Open horizontal bars represent the previously confirmed regions of the DMRs. Filled horizontal bars indicate the PCR amplicons, and gray horizontal bars represent the extent of the DMR determined in this study. Results of bisulfite sequencing are shown below the map (top, 12.5 dpc-embryos; bottom, sperm). Filled circles indicate methylated CpGs and open circles unmethylated CpGs. Hyphens indicate that these particular CpGs are missing due to an SNP or could not be analyzed due to a technical problem. The maternal and paternal alleles were distinguished by SNPs, if available, between C57BL/6 and JF1. Mat, the maternal (C57BL/6) allele; Pat, the paternal (JF1) allele.

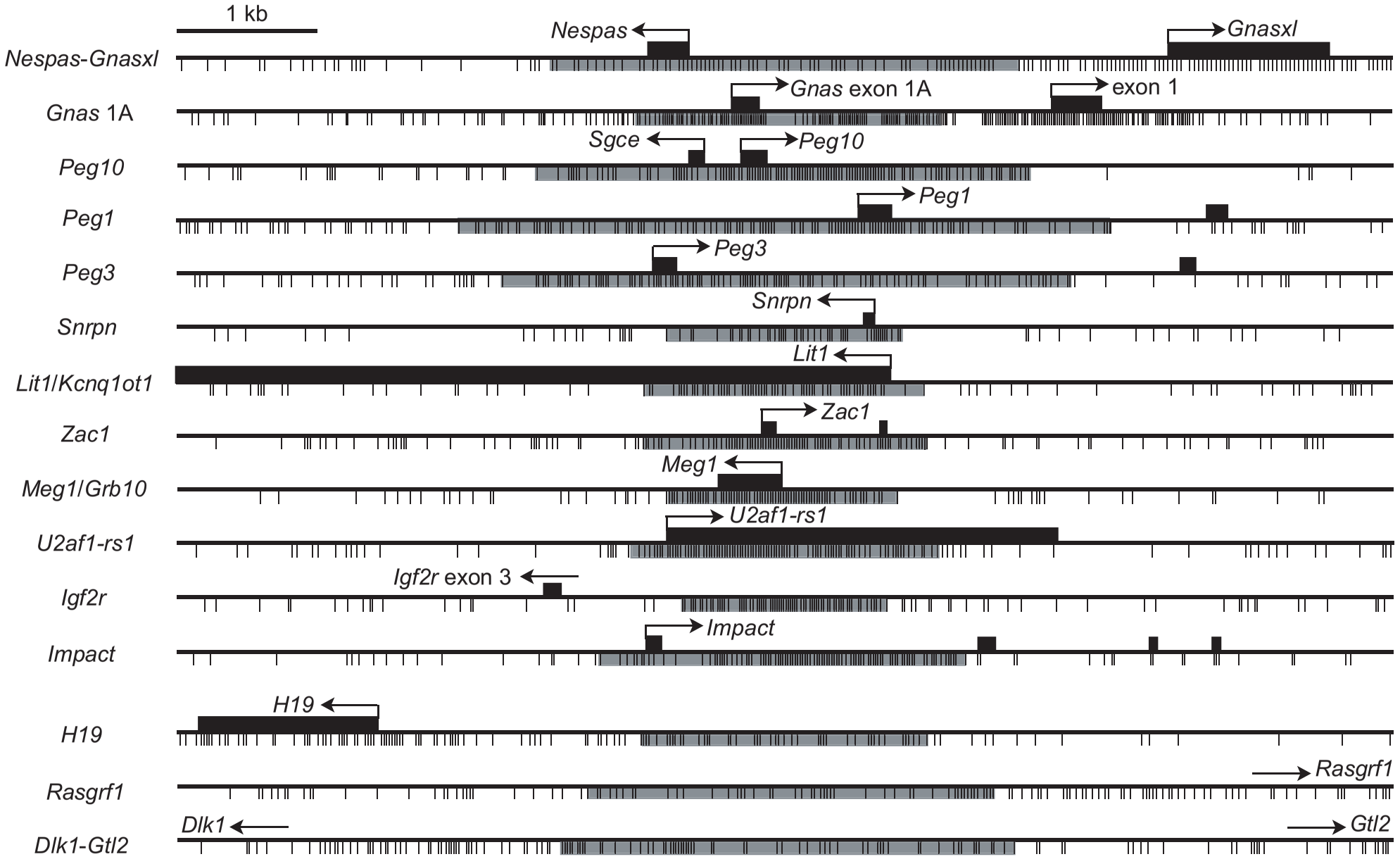

Figure 3. Schematic representation of the 15 DMRs

Figure 3. Schematic representation of the extents of the 15 germline DMRs. CpG positions are shown by vertical lines. Closed boxes represent the exons. Arrows indicate the transcription start sites and directions of transcription. Gray boxes represent the extents of the DMRs determined in this study.

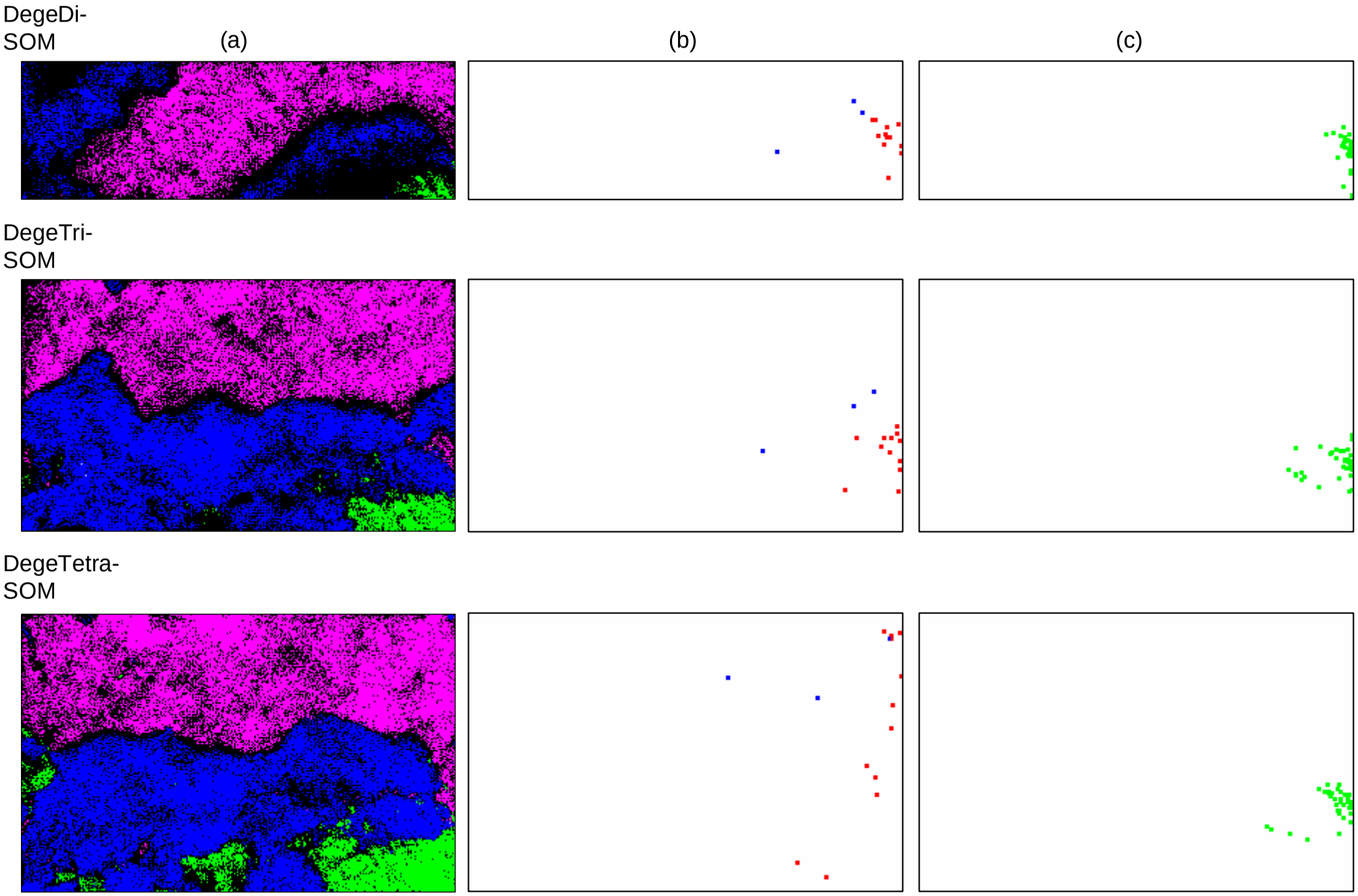

Figure 4. SOMs mapping of the germline DMRs

Figure 4. SOM mapping of the germline DMRs. (a) SOMs for 2-kb sequences of 10 eukaryotes genome and 143 prokaryotes genome. Top, DegeDi-SOM; middle, DegeTri-SOM; bottom, DegeTetra-SOM. Lattice points that include sequences from more than one species are indicated in black, those that contain no genomic sequences are indicated in white, and those containing sequences from a single species are indicated in color as follows: mouse *Mus musculus domesticus* (■), the remaining 9 eukaryotes (■), 143 prokaryotes (■). (b) Mapping of the 15 germline DMRs on the SOMs. 12 maternally methylated DMRs (■); 3 paternally methylated DMRs (■). (c) Mapping of 49 non-imprinted mouse CpG islands (■).

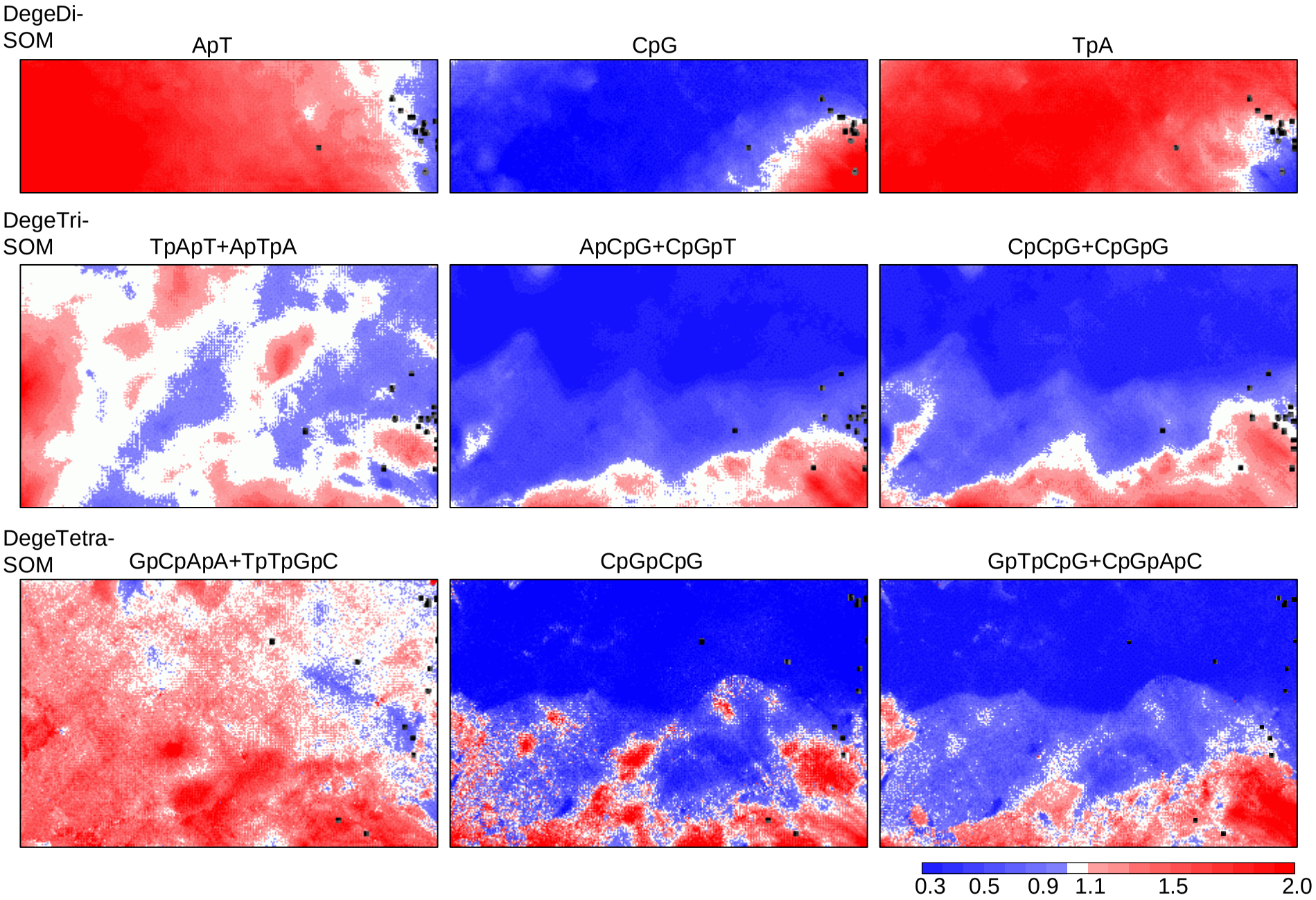

Figure 5. Levels of individual oligonucleotides in the SOMs

Figure 5. Levels of individual oligonucleotides in the SOMs. Top, DegeDi-SOM; middle, DegeTri-SOM; bottom, DegeTetra-SOM. Levels of some oligonucleotides and of some complimentary oligonucleotide pairs for each lattice point in the SOMS in Figure 4 are shown as examples of component planes. The observed/expected ratio is indicated in colors presented at the bottom. of the figure. Data of all component planes are presented by our URL (DegeDi-SOM, http://lavender.genes.nig.ac.jp/takaabe/KO/DS/GIF/DS.html; DegeTri-SOM, http://lavender.genes.nig.ac.jp/takaabe/KO/TRS/GIF/TRS.html; DegeTetra-SOM, http://lavender.genes.nig.ac.jp/takaabe/KO/TS/GIF/TS.html). Lattice points mapped for the 15 germline DMRs were indicated by black dots.

Table 1. Fifteen mouse germline DMRs.

Gene/DMR	Chromosome	References
Maternally methylated DMRs		
Nespas-Gnasxl	2	Kelsey et al., 1999; Coombes et al., 2003
Gnas 1A	2	Liu et al., 2000; Williamson et al., 2004; Liu et al., 2005
Peg10	6	Ono et al., 2003
Peg1/Mest	6	Lefebvre et al., 1997; Lucifero et al., 2002
Peg3	7	Li et al., 2000; Lucifero et al., 2002
Snrpn DMR1	7	Shemer et al., 1997
Lit1/Kcnq1ot1 (KvDMR1)	7	Engemann et al., 2000; Fitzpatrick et al., 2002; Yatsuki et al., 2002
Zac1	10	Smith et al., 2002
Meg1/Grb10	11	Arnaud et al., 2003; Hikichi et al., 2003
U2af1-rs1	11	Hayashizaki et al., 1994; Shibata et al., 1997
Igf2r DMR2	17	Stoger et al., 1993; Wutz et al., 1997
Impact	18	Okamura et al., 2000; personal communication (Obata et al.)
Paternally methylated DMRs		
H19	7	Bartolomei et al., 1993; Ferguson-Smith et al., 1993; Tremblay et al., 1995; Thorvaldsen et al., 1998; Ueda et al., 2000
Rasgrf1	9	Plass et al., 1996; Shibata et al., 1998; Yoon et al., 2002
Dlk1-Gtl2 (IG-DMR)	12	Takada et al., 2002; Lin et al., 2003

Table 2. Summary of the reported developmental methylation change in the 15 germline DMRs

Gene/DMR	DNA samples								
	Sperm	Oocytes	Androgenetic embyo	Parthenogenetic embryo	Paternal disomy	Maternal disomy	Preimplantational embryo	Postimplanation embryo	Adult somatic cells
Maternally methylated DMRs									
Nespas-Gnasxl	-	+	-	+	-	+	p-/m+	p-/m+	
Gnas 1A	-	+					p-/m+		p-/m+
Peg10	-	+						p-/m+	
Peg1/Mest	-	+		+			p-/m+	p-/m+	
Peg3	-	+			-	+		p-/m+	
Snrpn DMR1	-	+	-	+			p-/m+	p-/m+	p-/m+
Lit1/Kcnq1ot1 (KvDMR1)	-	+					p-/m+	p-/m+	p-/m+
Zac1	-	+			-	+			
Meg1/Grb10	-	+			-	+	p-/m+		p-/m+
U2af1-rs1	-	+					p-/m+	p-/m+	p-/m+
Igf2r DMR2	-	+					p-/m+	p-/m+	p-/m+
Impact	-	+							p-/m+
Paternally methylated DMRs									
H19	+	-					p+/m-	p+/m-	p+/m-
Rasgrf1	+	-					p+/m-		p+/m-
Dlk1-Gtl2 (IG-DMR)	+	-			+	-			p+/m-

+, Methylated; -, Unmethylated; p, paternal allele; m, maternal allele
References are listed in Table 1.

Table 3. Primer pairs for bisulfite analysis

Locus/Amplicon No.	Top primer (5' to 3')	Bottom primers (5' to 3')	Amplicon length (C57BL/6) (bp)	Annealing temperature (℃)
Nespas-Gnasxl A	GGGAGTTGAGGAGATGGGTAAAT	AAATAACCTCTAATACAAAATTACCCCTAACTAA	664	60
Nespas-Gnasxl B	GGGGATGTTATTGGTATTTGGTAGGTTTAG	CTACCACCTAAATTTTAAACACTAAAAATTAC	671	60
Nespas-Gnasxl C	GTAATTTTTAGTGTTTAAAATTTAGGTGGTAG	CACACCTTCACCTATCTAAAACAAAAAATCC	479	60
Nespas-Gnasxl D	GAAGAATTAGATGGGGAGGGAGG	CTATCACCTTCCTAATTACACTTACCCC	449	60
Nespas-Gnasxl E	GATTATTTATTTTTATAAAGTTTGGAAATTGAGGTTGGAG	CTTCTAAAAATCCACTAATCTCACTACC	383	60
Gnas 1A A	GTGTGTGTTGTAATGAAGAGAAGTTATAATTAGT	ACCCACAAATACCTTATACCCAAATAAAC	646	60
Gnas 1A B	GTTTATTTGGGTATAAGGTATTTGTGGGT	CCTAAACCTACCTCCCAATTTTCC	635	60
Gnas 1A C	GGTATTATGTTGAAGATGGTTATGAAGTTTAAAGTT	AACATCATCTTCTTCATCAATATCACTATCC	519	60
Gnas 1A D	GGATAGTGATATTGATGAAGAAGATGATGTT	CCCTCCTCCTACAACAACAAAAC	703	60
Peg10 A	TTGATTAGAATTATATTTTTATGTGATAATTTTGAGG	CAAATTTTAAAAATCACCCATCTTTCAAAAC	546	60
Peg10 B	GTAAAGGAATAAGGTATTTAAATATTGATTGGG	AACTCTTAAAAAACTCTCAATAAACAACTTAA	651	60
Peg10 C	GTTTTAAGTTAAGGAAATGTTTGGAGTTGAG	CTACAAACAACACCCATTACCCTAC	284	60
Peg10 D	GTAGGGTAATGGGTGTTGTTTGTAG	TCCCCAAACTTAACCCCTAACC	553	60
Peg10 E	GTATTTAATTTGGAAAGTTGTAGGAGAG	CTCCCAACCACCAAATCCCT	603	60
Peg10 F	AGTTATTTTGTAGTTTTGTTGTAGTTAGTTTTAGATTTGT	AAACTATACAACCAAACTCAAACAAACAAA	394	60
Peg10 G	TTTGTTTGTTTGAGTTTGGTTGTATAGTTT	CAACTCAACTCAAAACAACAAAACTCC	702	60
Peg10 H	AGTGGTGAAGGTTATAGATTTTTAATTGGTGAT	ACAACCTAACTCAAATCAAAAAAATACATACCCT	338	60
Peg1/Mest A	GGTAGGTAGAGGGTTTAGGTTGTTGG	TCATTTACTACTCACCACAAACCCTCT	348	60
Peg1/Mest B	TGAGTGGTTAAGATGGTTTTTATTTGTGGA	CAAAATCAATACTACACAAACTATTCCCTACCT	445	60
Peg1/Mest C	AGGTAGGGAATAGTTTGTGTAGTATTGATTTTG	CCAACAACTAAACCCTCTTACTCCT	533	60
Peg1/Mest D	AGGAGTAAGAGGGTTTAGTTGTTGGT	ACCCATTTTAAACACTATAACAAATTAAACAAACCC	623	60
Peg1/Mest E	AGTAGAGTATTTTTAGTGTTGTTTGAGGAGT	ACCTTCACATCCTTCTATACCATAATTCAC	478	60
Peg1/Mest F	GTGAATTATGGTATAGAAGGATGTGAAGGT	CTCTCAACCACAAAAACAATTCCAATCCAA	503	60
Peg1/Mest G	GAGGTGGTGGTGAAGTAATTTAGG	CCCAACCATTCTCAACTTTAATTACCTTA	627	60
Peg1/Mest H	TAAGGTAATTAAAGTTGAGAATGGTTGGG	CCCCTTAAAAACCTCAACTACCTACC	531	60
Peg1/Mest I	GGTAGGTAGTTGAGGTTTTTAAGGGG	CCCAACTTACCATACCTAAATTTCAAAAATCC	342	60
Peg1/Mest J	TGATTTTTGGAAGTGTATAATGTTATTTTGGTTGGT	AAATACTATATAACTTACCTAACACTAAAAAC	415	60
Peg1/Mest K	GTTTTTAGTGTTAGGTAAGTTATATAGTATTT	TCAAATTATTCAAACAATAAAATACCTTACCAACAATACT	622	60
Peg3 A	GTTTGATGTTGTTTTAAGATATTGAGAGGTAT	TCTTACACAACCTCTCCCTAAATAACC	665	60
Peg3 B	GGTGGTTATTTAGGGAGAGGTTGTG	CTTAAATCTTTAACTATTCATTCTACTCCCC	637	60
Peg3 C	GGGGAGTAGAATGAATAGTTAAAGATTTAAG	CTTTATTCTCTATACCCATAAATTTCCTTAAC	635	60
Peg3 D	GGTGAAGAAGGGAGTGGTAGG	CCTTCAATAATCACTTCAAACACACC	456	60
Peg3 E	GGGGTAAGTTTTAGGAAGGGG	CCCTACCCCTACAACCAACCC	559	60
Peg3 F	GGGTTGGTTGTAGGGGTAGGGT	CCACACCCCTTTCCCAAAATTCTA	628	60
Peg3 G	TGAGAAAAGAAGTATTTTTATGGAGTATTATGGGAG	ACCTACCCTACCCTCTACACCT	482	60
Snrpn DMR1 A	GAGAAGGTGGGGTGGATTATTAGT	ATATCTTTACCCCTCCTTTCCTACAATATC	666	60
Snrpn DMR1 B	GATATTGTAGGAAAGGAGGGGTAAAGATAT	CTATACAATACTTACTTTTACCCACTCATTAAA	510	60
Snrpn DMR1 C	TTTAATGAGTGGGTAAAAGTAAGTATTGTATAG	ACTTCTAATTTAAAACCTTAAAACTCTAACCCATA	435	60
Snrpn DMR1 D	GGTAATATAGTTTTGATTTTTAGGAGTTTAGAGGT	CCTTTTTTCCTCTCCCACATAATAAAAATCTA	728	60
Snrpn DMR1 E	GAGGGTTTGTAAGAATTGTTAATGAGGTATTT	CCAACCCAACCATAACACATCCTAT	620	60

Snrpn DMR1 F	GAGAGGGTATTTTGTAAGTTTGGGTT	TCAAACTTACATAAATAAACCTACCTCACTTAA	505	60
Snrpn DMR1 G	AGTTTGGTTTATAATATGAGTTTTAGGATAGTTAGGATT	CCTCACTATATTTTTAAAAATCTTTCAAAACTATAACCACAA	632	60
Snrpn DMR1 H	GTGTATTAGGGGTATGTTTTTTGTTTTTAGAGGT	CCCCAACTCAAACTAAACCCTATACCC	466	60
Lit1/Kcnq1ot1A	GGAGGTTTTATTGATATGTGGTGTTATTTTGATT	AAACAAATTTTTAAAACCCCACATTTCTTTATAACAATAATA	471	60
Lit1/Kcnq1ot1B	AGATTTTTTTAGGAAAGGTTATAGATGGTTTGAGG	ACAAACCAAAAATATTTTCCTAACCACCTCC	394	60
Lit1/Kcnq1ot1C	GGTGGTTAGGAAAATATTTTTGGTTTGT	AAATAATCCATTTCTACACCTATTTTCCAAATAC	620	60
Lit1/Kcnq1ot1D	GTTGGGTGGGGTAAGTTTTAGGG	ACCTCCAACCATAAAATTAATTTCATTACTACCTC	785	60
Lit1/Kcnq1ot1E	GTTGGGAAGGATTATGTAGAGAAAAGTATATT	CCAAAACCAAAAACATACTCATCTTTAACC	589	60
Lit1/Kcnq1ot1F	GTATTTGGATTGGGTGAGTTTGTATAGGAT	CTCCCCTTCCCCTAAATCCTATTTAAAATC	726	60
Zac1A	AGTGTTTAATTTGAGAATTGATAGTAGAGTTAAGATGT	CCTACCCCCATCCCACATATAACATA	555	60
Zac1B	GGGGTTGGGAGAGTAGGTTG	ACACTCTTTCTCCTATCCTAACCCAAA	607	60
Zac1C	TTTGGGTTAGGATAGGAGAAAGAGTGT	CCATCCCCACAAAAATAATCCACCC	517	60
Zac1D	TTGGGTTGGAGTTTGTGTTTGTTTTTTTGG	CATTTCTTCCCCTCAACACTAAAAAAAATCTATAC	679	60
Zac1E	AGAATTTGGTAGGTAAGGTTATTGTTTAGGG	AACTAAATAAAATATATCCAACCAACACCAAAATTACTT	549	60
Meg1/Grb10 A	GGTGAAAGAAGAGGATTTATATTGTTTGG	ACTTCTTACCACTCCCCATATCTAAAC	727	60
Meg1/Grb10 B	TTGATGGATATGAGGGGTTAAAAGG	TTAACTACCCTTCTCATCAACTACC	466	60
Meg1/Grb10 C	TTTGTAATGGTTTAAGTGGTGGGTAGTTT	ACAATCAAAATATAAACAAATAATCTAACTACTTCTCTAAAAATCT	677	60
Meg1/Grb10 D	TGTTGGGAAGTTTGTTTTGTTGAGT	TCACTCAAAACTCCCCTCCTC	441	60
Meg1/Grb10 E	GAAAAAGTTTGAGTATGGTTAAGAGAAAAGT	CCACCCTAAAACCTCTACTCCC	521	60
Meg1/Grb10 F	AGTATAGTTAGTTTAAGGAGAGAAAATTTGGGTTT	CACAAAATCTCAAATTTCATCCCTAATTTCCA	431	60
U2af1-rs1A	AAAGATTTTTTAGTTGAGTGATAGTATTGTAGTTTT	TCTCAACTCTTCCTACACCATACCTA	656	60
U2af1-rs1B	GTTTGTGTAAGTATAAGAGTTTATGGGGAT	TCCTACAATAAAAATACTCAACCCTAAACC	679	60
U2af1-rs1C	GATTATATGAAGAGTGGTTGTTGAGGG	TCATAATCATCCCTTCTACACTACTCC	436	60
U2af1-rs1D	GGAATGGAGTAGTGTAGAAGGGATGATTATGAT	ACCAAAATAATAACTTCTTAACTTACTATAATATTCCTCATA	522	60
U2af1-rs1E	GAGGAATATTATAGTAAGTTAAGAAGTTATTATTTTGGTTT	ACAAATTCCAAACCAACCAAAAATACATAA	712	60
Igf2r DMR2 A	TGGTGGTTTTGGGTGGTAGTGA	AATTATCTACTCTCCCTTTAAAAAAACAACCCTA	414	60
Igf2r DMR2 B	GGTGGTTAGGAAAATTTTTATAGGAGTGGT	ACCCTCATCTACAAAATCCTCCCC	545	60
Igf2r DMR2 C	GGGGAATTGAGGTAAGTTAGGGTTTT	TCTTATAACCCAAAAATCTTCACCCTAAC	271	60
Igf2r DMR2 D	GAAGGGTTTTGTGATTAGGGTTAA	AACACCTTCATATACCCCTAAACAC	286	60
Igf2r DMR2 E	TGTGTTTAGGGGTATATGAAGGTGTT	TCCCAAAATATACCAACCAATACAACAA	616	60
Igf2r DMR2 F	AGTTAGAGTTATTAGAGAAGGAGGAGATTT	AACTACCATAATATATACTAAACCCAAAACCAA	711	60
Igf2r DMR2 G	TTGGTTTTGGGTTTAGTATATATTATGGTAGTT	AACTATCCTTAACTCTCTTTATAATATCTCAATCTTAAA	698	60
Impact A	TTTTATGGTATTGGTATAGAGATAGGTAGGT	ACCTCCTTCCATATCTTATAAACATTCAAAA	598	60
Impact B	TTTTGAATGTTTATAAGATATGGAAGGAGGT	AAACTCTCATACTCATCCTTCAAAACAA	403	60
Impact C	TGTTTTGAAGGATGAGTATGAGAGTTTT	CCCTACTAAAACCTTACCTACCTCTAAC	405	60
Impact D	GTGGTATAGATGGGTAGGGTTAG	CATCTTAATTTCCCCACCTTTCCAA	675	60
H19 A	AGTGGGTGGATAATTAGGTGGTTGA	ACAAACACACCATTCCCATAAAACACTAC	619	60
H19 B	TCTACCTTAAAACTCCTCAAACTAACCAAATCTA	AACAAATAAATCTCCTTCTTCTCTCTTAACCTC	611	60
H19 C	GAGGTTAAGAGAGAAGAAGGAGATTTATTTGTT	CAACTTCAATATAACAAAATACCTAAAACACCACC	752	62
H19 D	AATGGGAATTTGAGGGTATGGGTTTT	ACCCTTAAATCCTCCTCCCATCTAAAAC	490	60
H19 E	GTTATTGTTGTGGGTGGATTTTAAGTTATGATATTG	ACCACACCAATAAATAACTTACATATAATAACCAACC	289	60
H19 F	AATGGTTGAATTTTAGTTTTTGTTTTTATGGTT	ACCAATACAATCCCACATACTTTATCATAAAA	506	60
H19 G	TGATTTTTTTGAGGTATTGAATTTGGGTGATT	ACTTTCTACCCTAACAATCCCCAAATTAAC	670	60
H19 H	GGTAGAAAGTATAATGAGGGTAGTAAGTTTTGG	TCCTAAAACTCACTTTATAAACCAAACTAACCTC	648	62
Rasgrf1 A	AGGTAGGAGTATTAGTTGTGTTTGGGT	CCCAATTATCATCAATTAAAAAATCATTAATCTAAAAACCAAAC	590	60

Rasgrf1 B	TGGTTTTTAGATTAATGATTTTTTAATTGATGATAATTGGGT	ATTCTATCCTTTAAAATAACCAATAAAAACACAATACACT	653	60
Rasgrf1 C	AGTGTATTGTGTTTTTATTGGTTATTTTAAAGGATAGAAT	AAACCATCACAAAAAACCACACAACTC	416	60
Rasgrf1 D	GAGTTGTGTGGTTTTTTGTGATGGTTT	TCCCTAATAACCAAAAAAAAACATTTTAAATCCCC	744	60
Rasgrf1 E	GGGATTTAAAATGTTTTTTTTTGGTTATTAGGGAT	ACATTCTCAACAAAAACAATAACCTACCTA	659	60
Rasgrf1 F	TGTTTTTGTTGAGAATGTATGTGTTTAAGAGAGT	ACCCAACTCTACATAAACCAAACAAAATAATACC	648	60
Rasgrf1 G	GGTATTATTTTGTTTGGTTTATGTAGAGTTGGGT	AACTTTATAAACTCTCTAAAAAATCCCCAAAATTCC	492	60
Rasgrf1 H	GGAATTTTGGGGATTTTTTAGAGAGTTTATAAAGT	CAAAAACAACAATAATAACAAAAACAAAAACAATAT	400	60
Rasgrf1 I	TTGTTGTTTTTGTTTTATATTGTTTTTGTTTTAGATATTATTGTTG	CAAATCTAATCTTCCTCATCAAAATATCTCCAACTA	439	60
Rasgrf1 J	TAGTTGGAGATATTTTGATGAGGAAGATTAGATTTG	AACCATCCTAATTAACAAAACAAAACCC	607	60
Rasgrf1 K	TTTTTGTTTGTTTTTAGTATGTTAAGATTGTAAATGTGAATT	TTCCCAACCACATATAAATTCACATACCTT	736	60
Rasgrf1 L	AAGGTATGTGAATTTATATGTGGTTGGGAA	TCCATTCCTCCCTCCAAAAAAACAC	644	60
Rasgrf1 M	AAATTTGAATTTGGAATATTTATTAGTTTGGGGGA	ACCCTAAAAACAAAACAAATCAAAAACTCTCT	475	60
Rasgrf1 N	TGGGAGGAAGGATTGTGTATATATGGAT	ACTTCCAAAACACTCTCTCTACTTTCTCTA	636	60
IG-DMR A	AGGGAATAGGATGTATTATGGAGTAATGT	AAACACCTAACCTATATTACAATACATCCATA	502	60
IG-DMR B	ATATGGATGTATTGTAATATAGGTTAGGTGTT	AATACACCATAACATAAACATAAAAATCCACAA	418	60
IG-DMR C	GTGTTGTAATTAGTTATAGTGGAGAGTATGTT	TCCCCCCACCAAAAACTAACC	428	60
IG-DMR D	TTATTATTTTGTGGTTATTTTTATTTAGAGAAAGTTTTATTGTTGT	TCCTTAAAACCCTCAAAAAACCAAATTATAATCTC	640	60
IG-DMR E	AGGAAGGGTAGTAGGGAATTTGTT	ATCCAACCCAACCAAATCCATAAAC	364	60

Table 4. Bisulfite sequencing of mouse germline DMRs

Target DMR	15 germline DMRs (12 maternally methylated DMRs; 3 paternally methylated DMRs)
PCR primers	127 PCR primer pairs synthesized (103, worked; 24, failed)
Amplicon (PCR product) size	Average, 551 bp (Minimum, 227 bp; maximum, 785 bp)
Total length analyzed (Sum of amplicon size)	56,769 bp
No. of amplicon with SNPs	89
Total no. of SNPs identified	228 (substitution, 151; deletion, 42; insertion, 35)
No. of clones sequenced	24 clones/amplicon
Total no. of sequence runs	~6200[†]
Total length sequenced	~2.8 Mb[†]

[†]Including those for both 12.5 dpc-embryo and sperm

Table 5. Size and GC contents of fifteen germline DMRs

Gene/DMR	Chromosome	Data from this study	
		Size (kb)	GC content (%)
Maternally methylated DMRs			
Nespas-Gnasxl	2	3.5	52.2
Gnas 1A	2	2.2	61.3
Peg10	6	3.6	55.9
Peg1/Mest	6	4.8	50.7
Peg3	7	4.2	53.4
Snrpn DMR1	7	1.7	55.6
Lit1 /*Kcnq1ot1* (KvDMR1)	7	2.1	53.6
Zac1	10	2.0	53.4
Meg1/Grb10	11	1.6	63.6
U2af1-rs1	11	2.3	54.0
Igf2r DMR2	17	1.5	60.7
Impact	18	2.7	55.4
Average		2.7	55.8*
Paternally methylated DMRs			
H19	7	2.1	50.3
Rasgrf1	9	3.0	43.6
Dlk1-Gtl2 (IG-DMR)	12	3.3	49.3
Average		2.8	47.7*
Average of DMRs		2.7	54.2
Whole mouse genome			41.7

Significant differences between the values of the maternally and paternally methylated DMRs are shown as *P < 0.05

Table 6. Randomly selected 49 non-imprinted mouse CpG islands.

Reference	Gene	Accession number	Chromosome	Length (bp)
Gardiner-Garden	Hist2h4	V00753	3	698
& Frommer	Car2	-	3	683
	Rpl32	K02060	6	1,168
	Rps16	M11408	7	1,218
	Mt1	J00605	8	957
	Mt2	K02236	8	1020
	Aprt	M11310	8	786
	Acta1	M12347	8	1,877
	Pomc	AH005319	12	1,168
	c-fos	V00727	12	1992
	Dhfr	M10071	13	1235
	Wnt1	K02593	15	3,158
	Rpl30	K02928	15	897
	c-myc	L00038	15	3,306
	MHC class II H2-IA-beta	K00008	17	1163
	Hprt	K01507	X	490
GenBank	Ncl	M22089	1	1785
	alpha-2 adrenergic receptor	M94583	2	1462
	Bc10, nn2	AF303656	2	2961
	Fjx1	AJ009634	2	1823
	Gclm	AF246994	3	811
	Rps3a	Z83368	3	1365
	Thioredoxin	D21855	4	850
	Mac25	AB042198	5	1,462
	Cenpc	AF012708	5	700
	Cyln2	AJ228864	5	687
	Spr	U78076	6	720
	Vasp	X98475	7	535
	Dmahp	X84814	7	1795
	Nktr	U63544	9	2054
	Bsg	D82019	10	1257
	Itgae, Gsg2	AF289866	11	1005
	Az1	D88497	11	955
	Sez6	Y09922	11	1529
	Nfh	Z31012	11	1482
	Sp4	AB019147	12	2037
	Apex	D38077	14	1069
	Krt7	AF509890	15	1035
	cytokeratin endo A	D90360	15	1165
	dTP2	X95711	16	1067
	Tcp1	D10606	17	803
	Rxrb	D21830	17	1272
	gMCK2-beta	X80685	17	1863
	Ring3	AL009226	17	1558
	Tcte2	U46150	17	845
	Lmnb1	D50070	18	1300
	Dp1/Tb2	D55682	18	788
	Zfx	L19715	X	2071
	G6pd	X53617	X	1692
Average of CpG islands				1,339

Table 7. Dinucleotide content of the germline DMRs

	Dinucleotide content (%)									
	ApA + TpT	ApC + GpT	ApG + CpT	CpA + TpG	CpC + GpG	GpA + TpC	ApT	CpG	GpC	TpA
Maternally methylated DMRs										
Nespas-Gnasxl	14.33	10.01	14.44	14.07	16.12	12.13	4.46	3.85	7.00	3.59
Gnas 1A	10.52	8.65	14.53	11.99	18.89	11.45	2.50	7.98	11.14	2.36
Peg10	13.10	9.74	13.16	13.90	18.10	11.20	4.06	5.38	8.41	2.95
Peg1/Mest	16.02	10.34	13.58	13.97	15.53	11.03	4.66	3.83	6.90	4.13
Peg3	12.58	9.26	14.92	15.25	16.30	12.22	4.94	3.46	7.78	3.29
Snrpn DMR1	10.20	11.24	15.33	16.95	16.95	10.78	3.80	3.23	8.30	3.23
Lit1 /*Kcnq1ot1* (KvDMR1)	11.67	12.05	14.27	14.37	14.90	13.11	4.19	5.06	6.75	3.62
Zac1	14.62	9.78	13.45	12.57	16.33	11.54	4.35	5.57	7.87	3.91
Meg1/Grb10	7.86	10.64	12.13	13.80	18.75	9.72	2.85	9.53	12.25	2.48
U2af1-rs1	11.80	11.36	13.61	13.74	15.86	10.34	4.64	5.39	8.17	5.08
Igf2r DMR2	10.18	8.68	14.09	12.07	21.00	12.39	3.20	6.78	9.26	2.35
Impact	11.11	9.33	15.85	16.85	12.00	10.85	4.15	5.37	11.59	2.89
Average	12.00	10.09	14.11	14.13*	16.73	11.40	3.98*	5.45**	8.79**	3.32*
Paternally methylated DMRs										
H19	12.67	11.44	12.72	16.56	14.67	11.49	6.41	3.23	6.36	4.46
Rasgrf1	18.06	10.12	14.44	15.21	9.41	11.69	6.87	2.31	6.20	5.70
Dlk1-Gtl2 (IG-DMR)	12.97	12.62	14.67	15.21	13.90	11.01	5.19	2.80	5.91	5.73
Average	14.57	11.39	13.94	15.66*	12.66	11.39	6.15*	2.78**	6.15**	5.29*
Average of DMRs	12.51	10.35	14.08	14.43	15.91	11.40	4.42	4.92	8.26	3.71
Average of CpG islands[a]	8.68	9.68	14.26	12.08	20.26	12.30	2.56	7.80	9.96	2.34
Whole mouse genome	18.19	10.67	14.69	14.92	10.44	12.39	7.36	0.84	4.11	6.38
Prokaryotic sequences from imprinted transgenes										
CAT	18.06	10.93	8.65	14.72	11.68	11.23	8.80	4.86	5.46	5.61
pBR322	11.77	10.69	12.06	13.35	13.26	12.43	5.85	7.55	8.69	4.36
pUC18	14.23	11.10	12.51	12.92	12.33	11.99	5.74	6.44	7.64	5.10
Average	14.68	10.90	11.08	13.66	12.42	11.88	6.80	6.28	7.26	5.02

[a]The values were obtained from randomly selected 49 non-imprinted mouse CpG islands.
Significant differences between the values of the maternally and paternally methylated DMRs are shown as
*P < 0.05 and **P < 0.01.

Table 8. Trinucleotide content of the germline DMRs

	Trinucleotide content (%)															
	TTT + AAA	TTC + GAA	TTA + TAA	TTG + CAA	TCT + AGA	TCC + GGA	TCA + TGA	TCG + CGA	TAT + ATA	TAC + GTA	TAG + CTA	TGT + ACA	TGC + GCA	TGG + CCA	AAG + CTT	GAG + CTC
Maternally methylated DMRs																
Nespas-Gnasxl	5.50	3.62	1.88	3.33	4.31	3.79	2.61	1.42	1.62	1.77	1.91	3.27	3.88	4.31	4.02	4.26
Gnas 1A	3.25	3.74	0.98	2.54	4.28	2.85	1.69	2.63	0.94	1.11	1.69	2.32	4.50	3.48	4.24	4.64
Peg10	4.99	3.39	1.57	3.15	2.95	3.81	2.76	1.68	1.02	1.60	1.71	2.87	3.73	4.55	3.31	3.84
Peg1/Mest	5.93	3.77	3.01	3.32	3.73	3.21	2.82	1.26	1.24	1.74	2.26	3.23	4.04	3.88	3.86	3.19
Peg3	4.32	2.91	1.84	3.51	4.54	3.08	3.58	1.03	1.86	1.00	1.89	2.86	4.03	4.77	3.84	3.68
Snrpn DMR1	2.88	2.31	1.21	3.81	4.04	4.04	2.25	0.46	0.98	1.96	2.31	4.27	4.79	5.65	2.60	4.38
Lit1 /Kcnq1ot1 (KvDMR1)	3.67	3.38	1.64	2.99	3.04	3.71	3.86	2.51	1.50	1.83	2.27	2.94	2.94	4.63	3.14	4.53
Zac1	5.77	2.74	2.79	3.33	3.18	3.72	2.40	2.25	1.37	1.66	2.01	2.94	3.47	3.77	3.42	4.11
Meg1/Grb10	1.80	2.35	1.24	2.48	1.98	3.53	1.42	2.79	0.74	1.30	1.67	3.41	4.02	4.95	1.98	3.84
U2af1-rs1	3.14	3.36	3.23	2.08	3.45	3.71	2.08	1.11	2.34	3.49	1.11	2.30	5.61	3.76	2.83	2.74
Igf2r DMR2	2.15	4.05	1.04	2.94	2.68	5.09	2.22	2.42	0.72	0.85	2.09	1.50	5.61	2.74	2.94	5.03
Impact	4.00	1.74	1.63	3.74	2.07	1.33	3.74	3.71	1.04	1.00	2.11	4.04	6.26	2.82	2.78	5.41
Average	3.95	3.11	1.84	3.10	3.35	3.49	2.62**	1.94*	1.28*	1.61	1.92	3.00*	4.41	4.11	3.25	4.14**
Paternally methylated DMRs																
H19	3.23	3.37	1.71	4.37	3.13	3.04	3.85	1.47	2.90	1.99	2.33	3.37	4.46	4.89	3.32	2.80
Rasgrf1	7.47	4.39	3.18	3.02	4.76	2.61	3.52	0.80	3.92	1.78	2.51	4.05	4.16	3.49	3.92	2.75
Dlk1-Gtl2 (IG-DMR)	3.76	3.43	2.21	3.58	3.52	3.19	3.31	0.98	2.71	3.37	3.16	3.94	3.49	4.47	3.37	3.37
Average	4.82	3.73	2.37	3.65	3.80	2.95	3.56**	1.09*	3.18*	2.38	2.67	3.79*	4.04	4.28	3.54	2.97**
Average of DMRs	4.12	3.24	1.94	3.21	3.44	3.38	2.81	1.77	1.66	1.76	2.07	3.15	4.33	4.14	3.30	3.90
Average of CpG islands	2.55	2.84	1.18	2.13	3.25	4.42	2.22	2.41	0.77	1.24	1.50	2.27	3.53	3.88	3.03	5.01
Whole mouse genome	7.74	4.34	4.01	4.07	5.10	3.21	3.90	0.36	4.46	2.43	2.83	4.59	2.47	3.44	4.27	3.22

Significant differences between the values of the maternally and paternally methylated DMRs are shown as *P < 0.05 and **P < 0.01.

Table 8. Continued

	Trinucleotide content (%)															
	CAG + CTG	AGG + CCT	GGG + CCC	CGG + CCG	ATG + CAT	GTG + CAC	ACG + CGT	GCG + CGC	AAT + ATT	GAT + ATC	AGT + ACT	GGT + ACC	GTT + AAC	GCT + AGC	GAC + GTC	GGC + GCC
Maternally methylated DMRs																
Nespas-Gnasxl	4.26	4.55	4.78	2.49	2.95	3.53	1.27	2.46	2.49	1.85	2.08	3.39	2.32	3.50	2.40	4.17
Gnas 1A	3.97	4.06	6.42	4.95	1.87	3.61	1.78	6.51	1.11	1.07	1.65	2.90	1.92	4.55	2.01	6.73
Peg10	4.30	4.25	6.82	2.48	3.04	3.42	1.79	4.75	2.24	1.82	2.01	3.06	2.57	3.95	2.15	4.42
Peg1/Mest	4.27	4.25	5.00	2.41	2.67	3.71	1.24	2.72	3.23	2.18	2.24	3.63	3.01	3.36	1.89	3.69
Peg3	5.51	4.42	4.87	2.24	3.03	3.20	1.05	2.55	2.43	2.55	2.15	3.20	1.98	3.82	3.08	5.16
Snrpn DMR1	6.06	4.21	4.96	2.13	2.65	4.44	0.81	2.94	2.25	1.73	2.54	3.63	2.48	4.56	2.36	4.33
Lit1/Kcnq1ot1 (KvDMR1)	4.34	4.20	3.14	2.94	2.46	4.58	2.32	2.27	2.03	2.41	2.65	4.15	2.85	4.39	2.80	3.91
Zac1	3.91	3.42	5.72	3.42	2.25	3.08	1.86	3.52	2.64	2.45	2.64	2.35	2.79	4.21	2.25	4.55
Meg1/Grb10	4.64	3.72	5.26	4.83	2.54	4.15	3.34	7.99	1.36	1.05	1.05	2.85	2.72	5.39	2.48	7.12
U2af1-rs1	6.94	3.76	4.20	4.16	2.21	2.52	2.12	3.32	2.03	2.70	3.40	3.54	3.80	3.01	1.55	4.42
Igf2r DMR2	4.05	5.68	8.42	4.18	1.17	3.92	1.70	5.16	1.76	2.74	1.83	3.66	3.33	3.92	0.59	3.85
Impact	5.56	3.00	2.67	3.52	3.71	3.85	1.41	2.04	2.04	1.52	2.59	1.30	2.30	8.19	2.19	6.71
Average	4.82	4.13	5.19	3.31**	2.55	3.67	1.72	3.85**	2.14*	2.01	2.24	3.14	2.67	4.40	2.14	4.92**
Paternally methylated DMRs																
H19	4.27	3.56	4.51	1.71	3.99	3.94	1.33	1.85	3.32	2.61	2.80	3.94	2.80	3.23	2.71	3.18
Rasgrf1	5.26	2.88	1.94	1.11	3.65	3.28	0.97	1.68	3.75	2.41	2.88	2.21	2.92	3.92	2.14	2.65
Dlk1-Gtl2 (IG-DMR)	4.77	3.85	3.88	1.70	2.89	3.97	1.58	1.28	2.71	2.06	3.64	3.46	3.13	3.67	2.15	3.37
Average	4.77	3.43	3.44	1.51**	3.51	3.73	1.29	1.60**	3.26*	2.36	3.11	3.20	2.95	3.61	2.33	3.07**
Average of DMRs	4.81	3.99	4.84	2.95	2.74	3.68	1.64	3.40	2.36	2.08	2.41	3.15	2.73	4.24	2.18	4.55
Average of CpG islands	4.65	4.58	6.51	5.46	1.59	3.52	2.19	5.85	1.09	1.57	2.02	3.18	2.03	4.33	2.87	6.32
Whole mouse genome	3.67	3.48	2.31	0.30	3.84	2.82	0.33	0.21	4.97	3.00	3.22	2.30	3.18	2.20	2.01	1.71

Significant differences between the values of the maternally and paternally methylated DMRs are shown as *P < 0.05 and **P < 0.01.

Table 9. Tetrainucleotide content of the germline DMRs

	Trinucleotide content (%)																
	AAAA + TTTT	GAAA + TTTC	TAAA + TTTA	CAAA + TTTG	AGAA + TTCT	GGAA + TTCC	TGAA + TTCA	CGAA + TTCG	ATAA + TTAT	GTAA + TTAC	CTAA + TTAG	ACAA + TTGT	GCAA + TTGC	TCAA + TTGA	CCAA + TTGG	AAGA + TCTT	GAGA + TCTC
Maternally methylated DMRs																	
Nespas-Gnasxl	2.03	1.33	0.84	1.30	1.36	1.01	0.81	0.43	0.55	0.49	0.38	0.81	0.98	0.49	1.04	1.04	1.48
Gnas 1A	1.52	0.71	0.18	0.85	1.87	0.67	0.40	0.80	0.31	0.22	0.27	0.76	0.76	0.27	0.76	1.92	1.25
Peg10	2.13	1.05	0.58	1.24	0.83	0.80	1.05	0.72	0.28	0.33	0.41	0.83	0.66	0.44	1.21	0.75	1.05
Peg1/Mest	2.05	1.16	1.31	1.41	1.53	1.00	0.85	0.39	0.44	0.66	0.91	1.04	0.87	0.58	0.83	0.91	0.89
Peg3	1.53	0.96	0.81	1.03	1.07	0.72	0.84	0.29	0.69	0.26	0.55	0.81	1.10	0.69	0.91	1.46	0.88
Snrpn DMR1	1.15	0.35	0.29	1.10	1.15	0.81	0.35	0.00	0.35	0.29	0.23	0.87	1.38	0.52	1.04	0.52	0.87
Lit1 /Kcnq1ot1 (KvDMR1)	1.54	0.97	0.43	0.72	1.06	0.97	0.77	0.58	0.48	0.29	0.29	0.68	0.63	0.48	1.21	0.82	0.82
Zac1	2.45	0.98	1.17	1.17	0.83	0.98	0.59	0.34	0.44	0.39	0.69	0.73	0.83	0.64	1.13	0.69	1.03
Meg1/Grb10	0.74	0.37	0.31	0.37	0.68	0.68	0.31	0.68	0.25	0.37	0.25	0.31	0.81	0.12	1.24	0.50	0.50
U2af1-rs1	0.88	1.06	0.66	0.53	1.55	0.93	0.66	0.22	1.42	1.37	0.09	0.66	0.71	0.35	0.35	0.97	0.88
Igf2r DMR2	0.52	0.65	0.26	0.72	0.72	2.09	0.59	0.65	0.20	0.26	0.33	0.59	1.37	0.33	0.65	0.52	0.78
Impact	1.70	0.56	0.44	1.30	0.48	0.33	0.48	0.44	0.41	0.30	0.33	1.30	1.26	0.33	0.85	0.56	0.48
Average	1.52	0.85	0.61	0.98	1.10	0.92	0.64**	0.46*	0.48	0.44	0.39	0.78*	0.95	0.44*	0.94	0.89	0.91
Paternally methylated DMRs																	
H19	0.57	0.52	0.52	1.62	1.09	0.81	1.19	0.29	0.71	0.33	0.29	0.95	1.14	0.86	1.43	0.76	0.86
Rasgrf1	3.59	1.34	1.34	1.21	2.04	0.80	1.31	0.23	1.14	0.50	0.67	1.21	0.67	0.64	0.50	1.54	1.14
Dlk1-Gtl2 (IG-DMR)	0.98	1.01	0.66	1.10	1.13	1.04	0.98	0.27	0.48	0.63	0.87	1.22	0.81	0.66	0.90	0.98	0.93
Average	1.71	0.96	0.84	1.31	1.42	0.89	1.16**	0.26*	0.78	0.49	0.61	1.13*	0.87	0.72*	0.94	1.10	0.97
Average of DMRs	1.56	0.87	0.65	1.04	1.16	0.91	0.75	0.42	0.54	0.45	0.44	0.85	0.93	0.49	0.94	0.93	0.92
Average of CpG islands	0.80	0.75	0.36	0.61	0.86	1.04	0.49	0.45	0.22	0.31	0.31	0.47	0.57	0.42	0.72	0.78	1.10
Whole mouse genome	3.00	1.62	1.55	1.57	1.80	1.10	1.32	0.11	1.30	0.70	0.82	1.35	0.70	1.05	0.97	1.55	1.27

Significant differences between the values of the maternally and paternally methylated DMRs are shown as *P < 0.05 and **P < 0.01.

Table 9. Continued

	Trinucleotide content (%)																
	TAGA + TCTA	CAGA + TCTG	AGGA + TCCT	GGGA + TCCC	TGGA + TCCA	CGGA + TCCG	ATGA + TCAT	GTGA + TCAC	CTGA + TCAG	ACGA + TCGT	GCGA + TCGC	CCGA + TCGG	AATA + TATT	GATA + TATC	CATA + TATG	AGTA + TACT	GGTA + TACC
Maternally methylated DMRs																	
Nespas-Gnasxl	0.58	1.22	1.16	0.93	1.01	0.70	0.43	0.72	0.96	0.23	0.52	0.55	0.41	0.32	0.49	0.41	0.49
Gnas 1A	0.45	0.67	0.89	0.94	0.45	0.58	0.36	0.36	0.71	0.40	0.98	1.25	0.36	0.13	0.27	0.27	0.49
Peg10	0.25	0.91	0.86	1.57	0.83	0.55	0.86	0.75	0.72	0.44	0.58	0.47	0.39	0.22	0.30	0.50	0.19
Peg1/Mest	0.85	1.08	1.12	0.97	0.77	0.35	0.58	0.81	0.85	0.21	0.39	0.41	0.48	0.33	0.19	0.27	0.71
Peg3	0.57	1.62	1.10	0.88	0.69	0.41	0.93	0.76	1.19	0.24	0.26	0.38	0.53	0.48	0.53	0.24	0.48
Snrpn DMR1	0.58	2.08	1.04	1.10	1.44	0.46	0.29	0.52	0.92	0.00	0.40	0.00	0.40	0.23	0.23	0.46	0.69
Lit1 /Kcnq1ot1 (KvDMR1)	0.48	0.92	1.01	0.68	1.25	0.77	0.82	1.06	1.50	0.72	0.43	1.25	0.48	0.39	0.63	0.34	0.53
Zac1	0.29	1.17	1.03	1.27	0.83	0.59	0.39	0.64	0.73	0.29	0.93	0.83	0.59	0.24	0.34	0.64	0.44
Meg1/Grb10	0.19	0.81	1.05	1.05	0.93	0.50	0.37	0.25	0.68	0.43	1.30	0.93	0.12	0.12	0.37	0.12	0.25
U2af1-rs1	0.31	1.28	0.84	0.66	0.93	1.28	0.44	0.57	0.71	0.22	0.27	0.40	0.31	1.46	0.49	0.84	1.59
Igf2r DMR2	0.65	0.72	1.50	1.31	0.72	1.57	0.33	0.65	0.91	0.26	0.72	1.18	0.07	0.39	0.13	0.00	0.33
Impact	0.30	0.74	0.41	0.37	0.30	0.26	1.07	1.30	1.04	0.30	0.52	2.45	0.26	0.19	0.44	0.33	0.19
Average	0.46	1.10	1.00	0.98	0.85	0.67*	0.57	0.70*	0.91	0.31	0.61*	0.84*	0.37	0.38	0.37	0.37	0.53
Paternally methylated DMRs																	
H19	*0.33*	1.19	0.95	0.90	0.71	0.48	0.86	1.05	1.09	0.43	0.19	0.48	0.48	0.43	1.14	0.52	0.67
Rasgrf1	*0.70*	1.37	0.91	0.91	0.60	0.20	0.54	0.94	1.41	0.30	0.27	0.10	1.01	0.70	1.21	0.47	0.37
Dlk1-Gtl2 (IG-DMR)	*0.57*	1.04	1.04	0.95	0.95	0.24	0.75	0.90	1.01	0.24	0.42	0.18	0.48	0.30	0.63	0.93	0.84
Average	0.53	1.20	0.97	0.92	0.76	0.30*	0.71	0.96*	1.17	0.32	0.29*	0.25*	0.65	0.48	0.99	0.64	0.62
Average of DMRs	0.47	1.12	0.99	0.97	0.83	0.60	0.60	0.75	0.96	0.31	0.55	0.72	0.42	0.40	0.49	0.42	0.55
Average of CpG islands	0.34	1.02	1.10	1.49	0.88	0.97	0.35	0.68	0.77	0.39	0.85	0.81	0.21	0.21	0.21	0.30	0.39
Whole mouse genome	1.00	1.28	1.23	0.80	1.10	0.08	1.09	0.71	1.05	0.09	0.06	0.09	1.45	0.79	0.95	0.79	0.56

Table 9. Continued

	Trinucleotide content (%)																
	TGTA + TACA	CGTA + TACG	ATTA + TAAT	GTTA + TAAC	CTTA + TAAG	ACTA + TAGT	GCTA + TAGC	CCTA + TAGG	AACA + TGTT	GACA + TGTC	CACA + TGTG	AGCA + TGCT	GGCA + TGCC	CGCA + TGCG	ATCA + TGAT	GTCA + TGAC	CTCA + TGAG
Maternally methylated DMRs																	
Nespas-Gnasxl	0.70	0.17	0.35	0.23	0.46	0.32	0.32	0.70	0.93	0.55	1.10	0.98	1.19	0.43	0.46	0.43	0.90
Gnas 1A	0.22	0.13	0.27	0.22	0.31	0.27	0.45	0.54	0.85	0.49	0.76	1.25	1.61	1.12	0.18	0.36	0.76
Peg10	0.52	0.39	0.25	0.30	0.44	0.33	0.55	0.58	0.86	0.61	0.88	0.91	0.75	0.91	0.36	0.52	0.83
Peg1/Mest	0.46	0.31	0.52	0.56	0.62	0.37	0.39	0.64	1.02	0.56	1.20	1.14	1.04	0.95	0.50	0.52	0.95
Peg3	0.24	0.05	0.26	0.38	0.38	0.29	0.62	0.41	0.64	1.19	0.79	0.96	1.12	0.43	0.57	1.27	0.91
Snrpn DMR1	0.58	0.23	0.29	0.35	0.29	0.52	0.63	0.58	0.98	0.69	2.02	0.81	1.33	1.27	0.29	0.58	1.04
Lit1 /Kcnq1ot1 (KvDMR1)	0.68	0.29	0.34	0.39	0.48	0.34	0.63	0.82	0.77	0.34	1.16	1.25	0.58	0.43	0.58	0.68	1.83
Zac1	0.49	0.10	0.29	0.54	0.78	0.34	0.83	0.54	0.93	0.69	0.83	0.88	1.08	1.03	0.44	0.54	0.83
Meg1/Grb10	0.56	0.37	0.31	0.31	0.31	0.19	0.62	0.68	1.12	1.05	0.68	0.93	1.80	0.93	0.19	0.37	0.56
U2af1-rs1	0.53	0.53	0.53	1.64	0.40	0.18	0.22	0.40	0.88	0.40	0.49	0.71	1.15	2.26	0.31	0.49	0.62
Igf2r DMR2	0.33	0.20	0.33	0.26	0.20	0.20	0.72	0.52	0.20	0.07	0.91	1.11	1.76	1.31	0.46	0.07	1.11
Impact	0.22	0.26	0.52	0.33	0.33	0.48	0.85	0.48	1.04	0.96	1.82	2.48	1.15	1.30	0.48	1.30	1.48
Average	0.46	0.25	0.35	0.46	0.42*	0.32	0.57	0.57	0.85	0.63	1.05*	1.12	1.21	1.03**	0.40	0.59	0.98
Paternally methylated DMRs																	
H19	0.57	0.24	0.24	0.19	0.76	0.33	0.76	0.90	0.71	0.62	1.47	0.90	1.52	0.43	0.81	0.81	1.05
Rasgrf1	0.74	0.20	0.94	0.37	0.54	0.74	0.74	0.34	1.17	0.80	1.34	1.78	1.14	0.37	0.80	0.60	0.80
Dlk1-Gtl2 (IG-DMR)	1.01	0.60	0.39	0.51	0.66	1.07	0.90	0.63	0.72	0.69	1.52	1.01	1.28	0.18	0.54	0.78	1.01
Average	0.77	0.35	0.52	0.36	0.65*	0.71	0.80	0.62	0.87	0.70	1.45*	1.23	1.31	0.03**	0.72	0.73	0.95
Average of DMRs	0.52	0.27	0.39	0.44	0.46	0.40	0.62	0.58	0.85	0.65	1.13	1.14	1.23	0.89	0.46	0.62	0.98
Average of CpG islands	0.35	0.23	0.17	0.24	0.37	0.26	0.43	0.47	0.50	0.64	0.85	0.94	1.00	0.91	0.32	0.56	0.86
Whole mouse genome	1.02	0.06	1.08	0.58	0.81	0.71	0.45	0.67	1.42	0.86	1.29	0.88	0.67	0.07	0.99	0.63	0.97

Table 9. Continued

	Trinucleotide content (%)																
	ACCA + TGGT	GCCA + TGGC	CCCA + TGGG	AAAG + CTTT	GAAG + CTTC	CAAG + CTTG	AGAG + CTCT	GGAG + CTCC	CGAG + CTCG	ATAG + CTAT	GTAG + CTAC	ACAG + CTGT	GCAG + CTGC	CCAG + CTGG	AAGG + CCTT	GAGG + CCTC	CAGG + CCTG
Maternally methylated DMRs																	
Nespas-Gnasxl	0.75	1.04	1.51	1.71	1.01	0.84	1.42	1.51	0.43	0.43	0.52	0.87	1.27	1.16	1.48	1.16	1.22
Gnas 1A	0.54	1.12	1.38	0.89	2.19	0.85	1.52	1.25	1.12	0.40	0.49	0.40	1.96	0.89	0.67	1.47	1.38
Peg10	0.80	1.02	1.90	0.83	1.10	0.94	0.94	1.63	0.44	0.25	0.55	0.86	1.44	1.30	0.99	1.38	1.30
Peg1/Mest	1.14	0.79	1.18	1.58	0.91	0.75	1.02	0.89	0.33	0.23	0.50	0.68	1.51	1.22	1.29	1.24	1.08
Peg3	1.03	1.93	1.12	1.24	0.91	1.31	1.38	1.19	0.19	0.55	0.21	1.19	1.74	1.38	1.12	1.34	1.55
Snrpn DMR1	1.10	1.21	1.90	0.58	0.75	0.98	1.62	1.67	0.06	0.35	1.15	1.62	1.85	1.67	0.87	1.10	1.67
Lit1 /Kcnq1ot1 (KvDMR1)	1.40	0.97	1.01	0.72	0.92	1.01	0.77	1.16	0.77	0.29	0.53	0.82	1.11	0.92	0.97	1.69	0.72
Zac1	0.44	1.08	1.42	1.22	0.59	0.83	1.03	1.37	0.88	0.39	0.44	0.88	1.17	1.13	0.93	1.17	0.78
Meg1/Grb10	1.05	1.61	1.36	0.19	0.81	0.68	0.74	1.24	1.30	0.25	0.56	1.18	1.12	1.67	0.43	1.05	1.55
U2af1-rs1	1.11	0.84	0.88	0.75	1.06	0.62	0.84	1.11	0.18	0.35	0.49	0.88	3.72	1.64	0.88	0.71	1.77
Igf2r DMR2	0.46	0.91	0.65	0.78	0.98	0.98	1.11	1.70	1.11	0.33	0.52	0.33	1.89	0.91	1.44	2.55	1.18
Impact	0.44	0.82	1.26	0.89	0.30	1.26	0.74	0.33	2.85	0.19	0.48	0.78	2.93	0.82	0.52	0.33	1.67
Average	0.85	1.11	1.30	0.95	0.96	0.92	1.09	1.25*	0.81	0.33	0.54	0.87	1.81	1.23	0.97	1.27*	1.32
Paternally methylated DMRs																	
H19	1.47	1.19	1.52	1.00	0.62	0.95	0.67	0.76	0.33	1.24	0.43	0.62	1.33	1.24	1.09	0.71	0.86
Rasgrf1	0.94	1.11	0.84	1.21	1.48	0.70	1.21	0.57	0.17	0.74	0.64	1.24	1.54	1.07	0.77	0.47	1.31
Dlk1-Gtl2 (IG-DMR)	1.22	1.10	1.19	0.95	0.69	1.07	1.01	0.95	0.39	0.60	0.87	1.34	1.13	1.28	0.78	0.95	1.49
Average	1.21	1.13	1.18	1.05	0.93	0.91	0.96	0.76*	0.30	0.86	0.64	1.07	1.34	1.20	0.88	0.71*	1.22
Average of DMRs	0.93	1.12	1.28	0.97	0.95	0.92	1.07	1.16	0.70	0.44	0.56	0.91	1.71	1.22	0.95	1.16	1.30
Average of CpG islands	0.68	1.13	1.25	0.81	1.13	0.75	1.24	1.83	1.06	0.23	0.44	0.82	1.67	1.46	0.99	1.72	1.40
Whole mouse genome	0.87	0.62	0.85	1.55	1.05	0.87	1.35	0.81	0.10	0.82	0.55	1.07	0.67	0.87	1.06	0.81	0.94

Table 9. Continued

	Trinucleotide content (%)																
	AGGG + CCCT	GGGG + CCCC	CGGG + CCCG	ATGG + CCAT	GTGG + CCAC	ACGG + CCGT	GCGG + CCGC	AATG + CATT	GATG + CATC	AGTG + CACT	GGTG + CACC	CGTG + CACG	ATTG + CAAT	GTTG + CAAC	ACTG + CAGT	GCTG + CAGC	AACG + CGTT
Maternally methylated DMRs																	
Nespas-Gnasxl	1.42	1.45	0.41	1.16	0.96	0.29	0.87	0.96	0.75	0.72	1.19	0.52	0.43	0.75	0.64	1.19	0.20
Gnas 1A	1.07	2.14	1.83	0.71	1.12	0.58	2.10	0.36	0.62	0.67	1.34	0.85	0.36	0.49	0.31	1.61	0.18
Peg10	1.08	2.87	0.97	0.69	1.35	0.69	0.99	0.75	0.77	0.69	1.30	0.55	0.41	0.55	0.77	1.32	0.47
Peg1/Mest	1.24	1.80	0.77	0.79	1.04	0.39	1.10	1.12	0.54	0.83	1.16	0.52	0.56	0.60	0.68	1.43	0.31
Peg3	1.55	1.41	0.79	1.24	1.24	0.29	1.10	0.86	1.03	0.84	1.31	0.26	0.50	0.67	0.86	1.48	0.29
Snrpn DMR1	1.10	1.73	0.23	0.92	2.02	0.29	1.27	0.92	0.58	0.75	1.50	0.17	0.58	1.15	0.92	1.38	0.23
Lit1 /Kcnq1ot1 (KvDMR1)	0.77	0.87	0.48	0.72	1.79	0.68	0.53	0.58	0.48	1.01	1.21	1.21	0.58	0.68	0.97	1.74	0.48
Zac1	0.69	2.25	1.37	0.64	0.88	0.59	1.08	0.73	0.78	0.78	0.69	0.78	0.29	1.03	0.59	1.37	0.49
Meg1/Grb10	0.99	1.55	1.36	0.87	1.18	0.68	0.74	0.56	0.25	0.43	0.99	2.04	0.19	1.24	0.37	1.92	0.50
U2af1-rs1	0.53	1.86	0.93	0.62	1.15	0.88	0.97	0.66	0.35	0.53	0.66	0.84	0.40	0.53	2.34	1.55	0.44
Igf2r DMR2	3.27	3.92	0.59	0.20	0.98	0.39	1.89	0.33	0.20	0.85	1.31	0.85	0.33	0.91	0.78	1.37	0.39
Impact	0.70	0.33	0.37	0.70	0.44	0.04	0.59	0.78	0.78	1.04	0.48	0.52	0.41	0.78	0.78	2.37	0.22
Average	1.20	1.85	0.84**	0.77**	1.18	0.48	1.10**	0.72	0.59	0.76	1.09	0.76	0.42	0.78	0.84	1.56	0.35
Paternally methylated DMRs																	
H19	1.24	1.52	0.24	1.00	1.24	0.24	0.67	1.14	0.57	0.90	1.05	0.52	1.09	0.71	0.95	1.28	0.29
Rasgrf1	0.57	0.37	0.17	1.07	0.84	0.34	0.50	0.80	0.64	0.80	0.74	0.40	0.47	0.64	0.77	1.81	0.13
Dlk1-Gtl2 (IG-DMR)	0.93	1.55	0.21	1.07	1.22	0.72	0.51	0.63	0.63	1.04	0.75	0.66	0.72	0.69	1.07	1.16	0.24
Average	0.91	1.15	0.20**	1.02**	1.10	0.43	0.56**	0.86	0.61	0.92	0.84	0.53	0.76	0.68	0.93	1.42	0.22
Average of DMRs	1.14	1.71	0.71	0.83	1.16	0.47	0.99	0.74	0.60	0.79	1.04	0.71	0.49	0.76	0.85	1.53	0.32
Average of CpG islands	1.41	1.98	1.80	0.56	1.19	0.61	2.16	0.34	0.61	0.69	1.21	0.80	0.26	0.56	0.63	1.69	0.37
Whole mouse genome	0.82	0.54	0.09	0.94	0.67	0.08	0.05	1.21	0.80	0.82	0.61	0.10	0.97	0.67	0.83	0.62	0.11

Table 9. Continued

	Trinucleotide content (%)																
	GACG + CGTC	AGCG + CGCT	GGCG + CGCC	ATCG + CGAT	GTCG + CGAC	ACCG + CGGT	GCCG + CGGC	AAAT + ATTT	GAAT + ATTC	AGAT + ATCT	GGAT + ATCC	GTAT + ATAC	ACAT + ATGT	GCAT + ATGC	AAGT + ACTT	GAGT + ACTC	AGGT + ACCT
Maternally methylated DMRs																	
Nespas-Gnasxl	0.38	0.43	0.84	0.26	0.29	0.58	0.81	1.10	0.61	0.64	0.49	0.35	0.55	0.81	0.67	0.46	0.81
Gnas 1A	0.62	1.07	2.54	0.18	0.54	0.67	1.87	0.36	0.13	0.36	0.36	0.13	0.40	0.40	0.49	0.58	0.58
Peg10	0.39	1.08	1.38	0.25	0.28	0.55	0.41	0.94	0.63	0.61	0.61	0.33	0.63	0.86	0.36	0.55	0.88
Peg1/Mest	0.10	0.46	0.85	0.35	0.19	0.52	0.77	1.31	0.85	0.64	0.68	0.33	0.75	0.56	0.77	0.41	0.85
Peg3	0.45	0.38	1.27	0.36	0.19	0.26	0.79	0.93	0.74	1.07	0.55	0.38	0.29	0.57	0.38	0.62	0.81
Snrpn DMR1	0.17	0.52	0.23	0.00	0.40	0.35	1.10	0.81	0.58	0.58	0.87	0.17	0.58	0.87	0.52	0.58	0.98
Lit1 /Kcnq1ot1 (KvDMR1)	0.34	0.48	0.29	0.58	0.58	0.58	1.11	0.48	0.63	0.43	0.82	0.43	0.39	0.53	0.43	0.92	1.11
Zac1	0.49	0.73	0.98	0.49	0.54	0.49	0.98	1.32	0.73	0.73	0.78	0.34	0.49	0.73	0.88	0.83	0.73
Meg1/Grb10	0.43	1.43	1.80	0.25	0.56	0.43	2.54	0.37	0.50	0.19	0.43	0.12	0.81	0.50	0.43	0.06	0.37
U2af1-rs1	0.31	0.44	0.35	0.40	0.31	0.84	1.11	0.62	0.49	0.62	1.37	0.40	0.57	0.57	0.40	0.49	0.62
Igf2r DMR2	0.26	0.98	0.52	0.39	0.26	0.72	1.31	0.33	0.78	0.65	1.24	0.07	0.20	0.46	0.33	0.52	0.65
Impact	0.41	0.30	0.30	0.11	0.30	0.15	2.74	0.70	0.41	0.56	0.37	0.15	0.70	1.22	0.37	0.96	0.22
Average	0.36	0.69	0.95**	0.30	0.37**	0.51	1.29	0.77	0.59**	0.59	0.71	0.27	0.53	0.67	0.50**	0.58	0.72
Paternally methylated DMRs																	
H19	0.29	0.43	0.33	0.71	0.14	0.38	0.62	0.81	1.19	0.62	0.48	0.48	0.90	1.24	0.81	0.71	0.81
Rasgrf1	0.23	0.67	0.10	0.23	0.17	0.20	0.54	1.34	1.01	0.74	0.64	0.44	1.17	0.87	0.70	0.67	0.80
Dlk1-Gtl2 (IG-DMR)	0.09	0.42	0.27	0.18	0.15	0.69	0.57	0.66	0.95	0.66	0.69	0.93	0.45	0.63	0.75	0.75	0.78
Average	0.20	0.51	0.23**	0.38	0.15**	0.42	0.57	0.94	1.05**	0.67	0.60	0.61	0.84	0.91	0.75**	0.71	0.80
Average of DMRs	0.33	0.65	0.80	0.32	0.33	0.49	1.15	0.80	0.68	0.61	0.69	0.34	0.59	0.72	0.55	0.61	0.73
Average of CpG islands	0.71	1.08	2.08	0.32	0.60	0.77	1.78	0.36	0.32	0.44	0.54	0.18	0.31	0.41	0.49	0.66	0.69
Whole mouse genome	0.06	0.06	0.06	0.09	0.06	0.07	0.06	1.92	1.01	1.17	0.75	0.75	1.16	0.66	1.00	0.68	0.80

Table 9. Continued

	Trinucleotide content (%)																
	GGGT + ACCC	GTGT + ACAC	GCGT + ACGC	GATT + AATC	AGTT + AACT	GGTT + AACC	GTTT + AAAC	GCTT + AAGC	GACT + AGTC	GGCT + AGCC	GTCT + AGAC	GCCT + AGGC	GAAC + GTTC	GGAC + GTCC	GCAC + GTGC	GAGC + GCTC	GGGC + GCCC
Maternally methylated DMRs																	
Nespas-Gnasxl	1.25	1.04	0.46	0.55	0.52	0.67	0.67	0.84	0.43	1.22	0.90	1.16	0.67	0.78	0.81	1.16	1.16
Gnas 1A	1.12	0.76	0.71	0.13	0.36	0.54	0.49	1.16	0.36	1.43	0.54	1.52	0.71	0.58	1.38	1.34	2.23
Peg10	0.83	0.55	0.44	0.39	0.44	0.80	1.10	1.21	0.39	1.02	0.58	1.44	0.61	0.77	0.77	0.86	1.55
Peg1/Mest	1.12	0.77	0.31	0.73	0.50	1.18	1.00	0.89	0.64	1.06	0.54	1.04	0.85	0.64	1.10	0.64	1.10
Peg3	1.10	0.57	0.33	0.72	0.33	0.72	0.62	0.88	0.74	1.38	1.00	0.96	0.31	0.62	0.62	0.84	1.48
Snrpn DMR1	1.21	1.21	0.52	0.58	0.58	0.69	0.35	0.69	0.75	1.50	0.69	1.10	0.63	0.69	0.69	1.85	0.92
Lit1 /Kcnq1ot1 (KvDMR1)	1.06	1.06	0.43	0.58	0.63	0.97	0.92	0.92	0.68	1.50	0.77	1.30	0.87	0.77	0.68	1.11	0.53
Zac1	0.69	0.83	0.54	0.73	0.69	0.69	0.78	0.93	0.54	1.71	0.59	0.98	0.44	0.59	0.73	1.08	1.52
Meg1/Grb10	0.99	1.12	2.04	0.31	0.12	0.99	0.50	0.62	0.37	1.80	0.37	1.30	0.68	1.18	1.61	2.23	1.67
U2af1-rs1	0.97	0.18	0.31	0.62	1.55	0.93	0.88	0.57	0.49	1.15	0.44	1.77	0.75	0.31	0.62	0.66	0.71
Igf2r DMR2	1.83	0.39	0.91	1.11	0.91	1.83	0.52	0.65	0.07	0.78	0.20	0.26	1.63	0.07	1.89	1.18	1.37
Impact	0.48	1.26	0.70	0.33	0.70	0.33	0.70	1.33	0.52	3.85	0.30	1.67	0.48	0.30	0.85	3.63	1.48
Average	1.05	0.81	0.64	0.56	0.61	0.86	0.71	0.89	0.50*	1.53	0.58*	1.21	0.72	0.61	0.98	1.38**	1.31**
Paternally methylated DMRs																	
H19	1.28	0.90	0.43	0.86	0.71	1.09	0.86	0.67	0.67	0.95	0.76	0.57	1.05	1.00	0.76	0.52	0.81
Rasgrf1	0.27	0.44	0.23	0.60	0.97	0.64	1.34	0.91	0.64	0.94	0.77	0.60	0.57	0.60	1.07	0.47	0.40
Dlk1-Gtl2 (IG-DMR)	0.78	0.93	0.21	0.72	1.01	1.16	1.16	0.87	0.66	1.16	0.72	1.10	0.78	0.51	0.93	0.75	0.60
Average	0.78	0.75	0.29	0.72	0.90	0.96	1.12	0.81	0.65*	1.02	0.75*	0.76	0.80	0.70	0.92	0.58**	0.60**
Average of DMRs	1.00	0.80	0.57	0.60	0.67	0.88	0.79	0.88	0.53	1.43	0.61	1.12	0.74	0.63	0.97	1.22	1.17
Average of CpG islands	1.06	0.74	0.70	0.34	0.46	0.69	0.56	0.79	0.59	1.34	0.69	1.38	0.65	1.01	0.94	1.51	1.91
Whole mouse genome	0.56	1.01	0.06	0.85	1.00	0.65	1.27	0.67	0.62	0.53	0.78	0.63	0.66	0.54	0.44	0.46	0.41

Table 9. Continued

	Trinucleotide content (%)																
	GGTC + GACC	GGCC	TTAA	TCGA	TATA	TGCA	CTAG	CCGG	CATG	CGCG	ATAT	ACGT	AATT	AGCT	GTAC	GCGC	GATC
Maternally methylated DMRs																	
Nespas-Gnasxl	1.04	0.46	0.23	0.06	0.20	0.64	0.29	0.38	0.38	0.38	0.14	0.14	0.29	0.43	0.20	0.29	0.12
Gnas 1A	0.54	0.58	0.09	0.00	0.09	0.27	0.27	0.49	0.31	0.89	0.04	0.04	0.13	0.40	0.13	1.34	0.09
Peg10	0.77	0.63	0.28	0.08	0.06	0.58	0.25	0.17	0.61	0.69	0.08	0.11	0.36	0.47	0.19	1.35	0.22
Peg1/Mest	0.58	0.37	0.50	0.10	0.12	0.46	0.31	0.25	0.41	0.23	0.12	0.17	0.46	0.35	0.12	0.46	0.29
Peg3	0.69	0.69	0.17	0.05	0.17	0.76	0.29	0.24	0.31	0.24	0.12	0.10	0.17	0.55	0.07	0.43	0.17
Snrpn DMR1	0.75	0.63	0.17	0.00	0.06	0.69	0.29	0.29	0.46	0.46	0.06	0.00	0.17	0.87	0.17	0.35	0.17
Lit1 /Kcnq1ot1 (KvDMR1)	1.45	0.77	0.29	0.05	0.00	0.34	0.58	0.24	0.39	0.53	0.14	0.19	0.19	0.58	0.29	0.43	0.48
Zac1	0.54	0.39	0.64	0.10	0.10	0.24	0.24	0.44	0.20	0.39	0.10	0.20	0.29	0.44	0.24	0.49	0.34
Meg1/Grb10	0.62	0.87	0.19	0.06	0.06	0.19	0.31	1.24	0.68	1.92	0.06	0.06	0.19	0.62	0.12	1.92	0.19
U2af1-rs1	0.35	0.88	0.18	0.09	0.04	0.75	0.09	0.93	0.35	0.13	0.09	0.35	0.22	0.35	0.62	0.88	0.13
Igf2r DMR2	0.20	0.39	0.13	0.13	0.07	0.72	0.46	0.33	0.26	1.18	0.07	0.07	0.13	0.52	0.00	0.78	0.52
Impact	0.30	0.70	0.30	0.19	0.07	0.67	0.56	0.22	0.85	0.07	0.15	0.19	0.33	0.78	0.04	0.11	0.11
Average	0.65	0.62*	0.26	0.08	0.09	0.53	0.33	0.43**	0.43	0.59	0.10	0.13	0.24	0.53	0.18	0.74*	0.24
Paternally methylated DMRs																	
H19	1.14	0.19	0.19	0.19	0.43	0.81	0.19	0.14	0.57	0.33	0.24	0.10	0.43	0.48	0.38	0.29	0.38
Rasgrf1	0.47	0.23	0.44	0.07	0.50	0.44	0.23	0.07	0.50	0.27	0.80	0.03	0.67	0.27	0.10	0.34	0.23
Dlk1-Gtl2 (IG-DMR)	0.72	0.33	0.12	0.06	0.66	0.51	0.42	0.15	0.51	0.21	0.36	0.21	0.45	0.54	0.48	0.06	0.21
Average	0.78	0.25*	0.25	0.11	0.53	0.58	0.28	0.12**	0.53	0.27	0.47	0.11	0.52	0.43	0.32	0.23*	0.27
Average of DMRs	0.68	0.54	0.26	0.08	0.18	0.54	0.32	0.37	0.45	0.53	0.17	0.13	0.30	0.51	0.21	0.63	0.24
Average of CpG islands	0.91	0.89	0.15	0.18	0.08	0.37	0.26	0.87	0.24	0.85	0.07	0.20	0.11	0.53	0.17	1.03	0.23
Whole mouse genome	0.47	0.22	0.60	0.06	0.63	0.43	0.32	0.04	0.44	0.01	0.80	0.05	0.73	0.36	0.21	0.02	0.28

Table 10. Genome sequeces used for SOM analysis

Species	Sequence lengths (MB)
Mus musculus (Mouse)	2,638
Danio rero (Zeblafish)	1,169
Fugu rubripes (Fugu)	309
Drosophila melanogaster (Fruit fly)	115
Caenorhabditis elegans (Nematode)	100
Arabidopsis thaliana (Thale cress)	116
Oryza sativa japonica (Japanese rice)	359
Saccharomyces cerevisiae	12
Schizosaccharomyces pombe	12
Plasmodium falciparum	20
143 Prokaryotes	402

Table 11. Mapping 15 mouse germline DMR on the SOMs

Gene/DMR	DegeDi-SOM					DegeTri-SOM					DegeTetra-SOM				
	X axis	Y axis	Distributed genome segments			X axis	Y axis	Distributed genome segments			X axis	Y axis	Distributed genome segments		
			Prokaryotes	Eukaryotes	Mouse			Prokaryotes	Eukaryotes	Mouse			Prokaryotes	Eukaryotes	Mouse
Maternally methylated DMRs															
Nespas-Gnasxl	281	40	5	49	543	287	109	1	19	419	292	17	0	2	698
Gnas 1A	299	63	20	109	466	298	131	19	50	567	282	125	2	261	2
Peg10	289	52	38	143	142	291	119	355	94	43	281	113	0	331	0
Peg1/Mest	279	40	10	66	720	268	109	0	16	224	287	12	0	0	819
Peg3	289	45	0	26	242	292	109	0	3	273	298	13	0	8	1322
Snrpn DMR1	297	43	0	82	431	298	111	1	67	443	292	15	0	3	721
Lit1/Kcnq1ot1 (KvDMR1)	283	51	4	451	5	296	101	2	273	1	275	105	0	273	1
Zac1	288	50	19	81	390	285	115	2	51	148	293	63	0	0	596
Meg1/Grb10	290	80	1331	12	14	297	146	15	238	15	247	182	626	0	0
U2af1-rs1	291	52	7	127	198	296	106	5	333	13	227	172	437	0	0
Igf2r DMR2	299	58	3	26	562	298	125	157	63	489	292	79	0	5	429
Impact	287	57	124	325	10	260	145	451	457	1	299	43	0	0	631
Paternally methylated DMRs															
H19	272	35	365	201	292	280	77	0	265	115	241	58	0	0	332
Rasgrf1	213	62	31	745	13	203	118	0	529	32	179	44	0	3	361
Dlk1-Gtl2 (IG-DMR)	266	27	0	111	556	266	87	0	31	435	291	17	0	1	683