Gene Body-Specific Methylation on the Active X Chromosome

Asaf Hellman, *et al.*

Science **315**, 1141 (2007);
DOI: 10.1126/science.1136352

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Gene Body–Specific Methylation on the Active X Chromosome

Asaf Hellman and Andrew Chess

Differential DNA methylation is important for the epigenetic regulation of gene expression. Allele-specific methylation of the inactive X chromosome has been demonstrated at promoter CpG islands, but the overall pattern of methylation on the active X (Xa) and inactive X (Xi) chromosomes is unknown. We performed allele-specific analysis of more than 1000 informative loci along the human X chromosome. The Xa displays more than two times as much allele-specific methylation as Xi. This methylation is concentrated at gene bodies, affecting multiple neighboring CpGs. Before X inactivation, all of these Xa gene body–methylated sites are biallelically methylated. Thus, a bipartite methylation-demethylation program results in Xa-specific hypomethylation at gene promoters and hypermethylation at gene bodies. These results suggest a relationship between global methylation and expression potentiality.

DNA methylation is essential for many developmental processes, including maintaining the silenced Xi state (*1*–*3*). Xi-specific methylation is seen at CpG islands (*4*–*9*), but the global distribution of Xi-specific methylation along the chromosome is unknown. Although silenced chromatin regions are usually hypermethylated, cytogenetic studies have suggested a global hypomethylation of the Xi (*10*); also, studies of mammalian hypoxanthine-guanine phosphoribosyltransferase (HPRT) genes have shown that at least one site within the transcribed region of the gene is methylated only on Xa in opposition to the Xi-specific methylation of the 5′ CpG island (*6*, *9*, *11*). This unexplained divergence led us to conduct a comprehensive analysis of the Xa and Xi allele–specific methylation patterns of the human X chromosome.

We modified the Affymetrix 500,000 (500K) single-nucleotide polymorphism (SNP) mapping array (*12*) to allow allele-specific analysis of DNA methylation (*13*). We digested genomic DNA with a cocktail of five methyl-sensitive restriction enzymes (MSREs) (*14*). Together, these frequent cutters recognize ~40% of all CG dinucleotides in the genome (*15*), allowing an efficient analysis of regions with both high and low GC content. After this pretreatment, fragments of 200 to 1100 base pairs containing the polymorphic sites were polymerase chain reaction (PCR) amplified, and the resulting amplicons were then labeled and hybridized to the array. Thus, an unmethylated MSRE site present on a given amplicon will lead to allele-specifically reduced intensity corresponding to the resident SNP (Fig. 1A). This allowed us to use either genotype calling or copy-number algorithms to identify transitions from a heterozygous state to a hemizygous state after the MSRE treatment. The assay is robust—replicate analyses of the same DNA never reveal allele-specific methylation status changing from one allele to the other [supporting online material (SOM) text 1]. Single locus validation included DNA sequencing of specific PCR products before and after MSRE treatment (table S1), and also bisulfite analyses (fig. S2).

Genotyping individuals from three generations of Centre d'Etude du Polymorphisme Humain (CEPH) pedigree 1332 (fig. S3) resolved the parental origins of 1948 heterozygous SNPs in the mother GM10849 and/or her daughter GM13130. Sequence analysis suggested that 1269 (65.1%) of them were suitable for our methylation assay, in that they had at least one methyl-sensitive restriction site on their respective amplicons (table S2). These informative SNPs were distributed along the chromosome (fig. S4) with an average distance of 120.1 kb between succeeding SNPs. Moreover, 351 SNPs mapped to 135 different genes (for an average of 2.6 SNPs per gene), representing ~9% of the known X-linked genes (table S2).

We resolved the active and inactive X copies in clones originated from single cells of GM10849

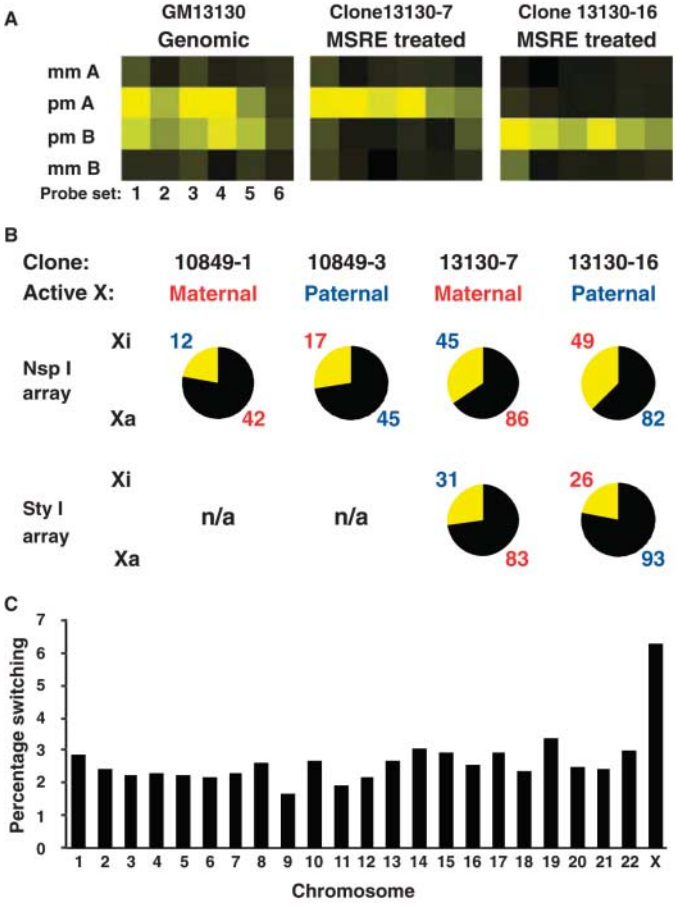


Fig. 1. Hypermethylation of the active X chromosome. (**A**) An example of hybridization intensities (brighter yellow indicates higher signal) for the six probe sets querying a particular SNP (rs16999756). When examining the perfect-match (pm) probes, one can observe both alleles in genomic DNA, but only one allele in each clone after MSRE treatment. This is due to methylation differences at an MSRE site present within the amplicon (fig. S1). Mismatch (mm) probes contain a single mismatch at a site distinct from the polymorphic site and serve as controls. (**B**) Presentations of Xa (black) and Xi (yellow) monoallelic methylation fractions in four clones from two individuals. The parental origin of Xa for each clone is indicated, and the numbers of occurrences of monoallelic methylation are marked blue (paternal) or red (maternal). Results from the Nsp I or the Sty I halves of the 500K array are presented. (**C**) Frequency of switching from heterozygous to hemizygous call after MSRE treatment for each chromosome. Data are presented for all informative 500K SNPs for GM13130 clone 7 and GM13130 clone 16. Potential monoallelic methylation at parentally imprinted regions, as well as technical artifacts resulting from MSRE site polymorphisms or allelic bias of the genotyping algorithm, were filtered out by requiring at least one clone switch from AB to A0, and one clone switch from AB to B0 per each SNP.

Center for Human Genetic Research and Department of Medicine, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA. E-mail: hellman@chgr.mgh.harvard.edu (A.H.); chess@chgr.mgh.harvard.edu (A.C.)

and GM13130 using reverse transcription (RT)–PCR to examine allele-specific mRNA expression (table S3). Replication timing analyses further confirmed the direction of X inactivation. We selected four clones, one paternal active (Xap) and one maternal active (Xam) from each individual, for further analysis. By definition, these clones must originate from four independent X inactivation events.

As a prelude to determining which heterozygous SNPs switch to a hemizygous state after digestion with the MSRE cocktail, we first filtered out SNPs that would show loss of heterozygosity as a result of polymorphism in their linked MSRE sites (fig. S1). We then analyzed the methylation status of the remaining 913 informative SNPs (table S4).

A pattern appeared when we related the SNPs to Xa and Xi: We always observed a significant excess of monoallelically methylated loci on the Xa. The average Xa:Xi ratio was 2.4 (Fig. 1B). The allele-specific methylation we observed is unique to the X chromosome, as shown by control analyses of autosomes, which show a vastly lower level of switching and no parent-of-origin effects (Fig. 1C and fig S5).

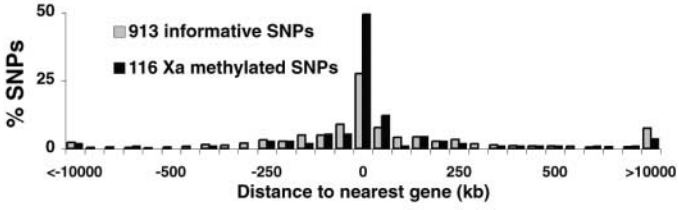
Because this global Xa hypermethylation is contrary to the reports of Xi-specific methylation at CpG islands, we sought to rule out the possibility that the clones we analyzed have unusual Xi methylation patterns. Two informative SNPs reside on amplicons that overlap with CpG islands: rs17139585 at phosphoglycerate kinase 1 and rs864570 at calcium/calmodulin-dependent serine protein kinase. As expected, these two SNPs are monoallelically methylated on their Xi alleles, indicating that our clones were not abnormal in their Xi methylation. Analysis of allele-specific methylation of the androgen receptor gene (*16*) further confirmed normal Xi promoter methylation (fig. S6).

To see if this methylation targets a unique set of loci, we analyzed overlapping among four clones by adding two more GM13130 clones (fig. S3). The extent of overlap among all combinations of three and four clones (table S5) was significantly higher than expected from a random distribution of Xa methylated amplicons along the chromosome ($P < 0.001$). Thus, a set of chromosomal locations are consistently Xa methylated.

We next examined whether these Xa-methylated sites related to genes. Out of the 116 SNPs that were Xa methylated in at least one Xam and one Xap clone, 57 (49.1%) were within known genes (Fig. 2 and table S6). This ratio is significantly higher ($P = 0.0009$) than the occurrence of gene SNPs in the entire informative set (251 out of 913 = 27.5%) (SOM text 2). Considering overlap in three and in four clones further highlighted gene exclusivity (table S7). All 17 amplicons displaying monoallelic Xa methylation in all four GM13130 clones were found either within genes or <50 kb from a gene (Fig. 2).

The CpG frequency in gene-body amplicons displaying Xa methylation is similar to the genome-wide average (table S8), well below the frequency seen in CpG islands. Nevertheless, nearby CpGs within different amplicons distantly present in the same gene (Fig. 2 and table S4), and neighboring CpGs within individual amplicons (fig. S7) tend to have the same Xa-specific methylation pattern.

Silencing tissue-specific genes does not explain Xa-specific gene-body methylation; indeed, mining public databases for tissue specificity and B cell line expression levels showed that both expressed and silenced genes undergo Xa-specific methylation (table S9). Furthermore, genes subject to Xa methylation do not share a common biological function or expression pattern (*17*). We cannot rule out a specific need for shutting down spurious transcription (e.g., from transposable elements) at active regions (i.e., genes), but the absence of a methylation bias toward repetitive elements embedded in genic amplicons (table S9) is a probable hint against this hypothesis. The list



SNP ID	Position	Gene	Gene relation (kb)	Clone 13130- 2	6	7	16
rs4826507	56693215	retro-UQCRB	upstream (46.2)	T	A	T	A
rs5931014	135698425	RBMX	upstream (10)	A	G	A	G
rs1638617	12576485	PRPS2	intron	C	G	C	G
rs2071211	21629701	MBTPS2	intron	C	T	C	T
rs5944690	24633379	POLA	intron	G	A	G	A
rs5972354	31036808	DMD	intron	C	T	C	T
rs17246798	37893629	RPGR	intron	G	A	G	A
rs6615850	95794852	DIAPH2	intron	T	A	T	A
rs5920651	96038343	DIAPH2	intron	T	G	T	G
rs5921728	96324533	DIAPH2	intron	A	G	A	G
rs5921827	96520713	DIAPH2	intron	G	A	G	A
rs42890	119360395	LAMP2	intron	T	G	T	G
rs209235	129073793	SUHW3	intron	T	C	T	C
rs5924843	149630958	CD99L2	intron	C	T	C	T
rs1212068	24323757	PDK3	downstream (11.6)	T	C	T	C
rs5951658	21699232	MBTPS2	downstream (38.7)	C	A	C	A
rs12557580	7900968	VCX2	downstream (46.8)	A	G	A	G

Fig. 2. Genes are preferable targets for Xa-specific methylation. (**Top**) Enrichment of Xa-methylated SNPs at gene regions. The distributions of Xa-methylated SNPs (black) and all informative SNPs (gray) are shown. SNPs were ranked according to distance from the nearest gene (table S6). (**Bottom**) The complete list of the SNPs showing Xa-specific methylations in all four clones from GM13130 is presented, with the methylated allele indicated for each clone (maternal in pink and paternal in blue).

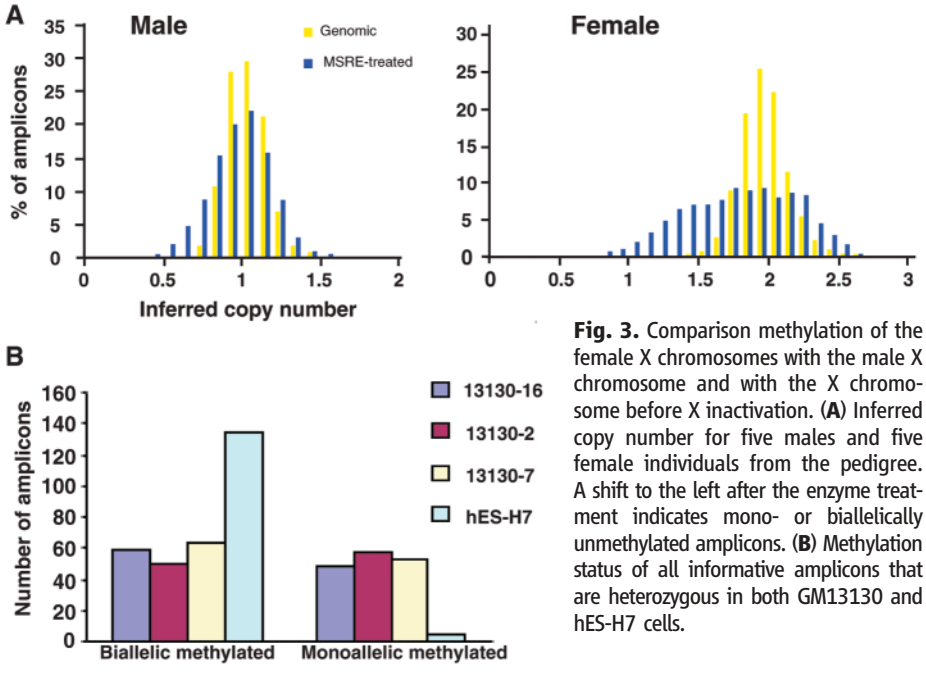


Fig. 3. Comparison methylation of the female X chromosomes with the male X chromosome and with the X chromosome before X inactivation. (**A**) Inferred copy number for five males and five female individuals from the pedigree. A shift to the left after the enzyme treatment indicates mono- or biallelically unmethylated amplicons. (**B**) Methylation status of all informative amplicons that are heterozygous in both GM13130 and hES-H7 cells.

of Xa-methylated genes contains several genes shown to escape X inactivation (*18*). However, given that these genes still have an elevated Xa expression level (*18*), the overall correlation between gene-body methylation and expression potentiality is maintained.

Male-to-female comparisons showed a more complete methylation level on the male X (Fig. 3A and fig. S8). Furthermore, the sites that are gene-body Xa-methylated in females are among the most highly methylated sites in males, highlighted even on the almost-complete methylation background of the male X chromosome (fig. S9).

An Xa-specific methylation present in somatic cells could reflect either methylation occurring only on Xa or demethylation occurring on Xi. The human embryonic stem (ES) cell line hES-H7 represents a stage of development just before X inactivation (*19*) and has been shown to stably maintain the appropriate characteristic methylation pattern for this stage, including allele-specific methylation (*20*). At this stage of development, the genome has presumably already undergone global demethylation and the wave of de novo methylation (*2*, *21*). Hence, we analyzed these cells using the 500K array. Out of the 116 amplicons shown to be gene-body Xa methylated, 50 also have a heterozygous genotype in H7 cells. We found that all 50 are biallelically methylated. When examining all 154 amplicons informative both for H7 and for 13130 clones, we observed that only five are monoallelically methylated in H7 cells, whereas the expected one-third are monoallelically methylated in the somatic clones (Fig. 3B). Bisulfite sequencing further verified biallelic methylation (fig. S10). Thus, given that biallelic methylation is the beginning state, demethylation of the Xi must account for the Xa-specific monoallelic pattern observed in somatic cells.

A simple model may explain both the Xa versus Xi and the gene versus intergenic differential methylation we observed: Constantly inactive regions, such as gene-poor regions and the entire Xi, may be more prone to loss of methylation maintenance (even if originally highly methylated). The resulting methylation decrease, for the entire Xi and for Xa intergenic regions, would thus highlight Xa gene body–specific methylation. At the same time, promoter CpG islands, which are protected from methylation on Xa, would remain more methylated on Xi.

In contrast to the widely held view that X chromosome allele–specific methylation is restricted to CpG islands on the inactive X, our global allele-specific methylation analyses uncovered extensive methylation specifically affecting transcribable regions (gene bodies) on the active X whether it is in the male or the female. One aspect of sex chromosome dosage compensation is the requirement for a chromosome-wide, likely epigenetic mechanism with the ability to double X-linked gene expression when necessary (i.e., in somatic cells but not in haploid germline cells). Indeed, such a phenomenon was recently described in mammals (*22*). Our finding of global elevation of methylation levels at gene bodies of both male and female active X chromosomes hints at such a chromosome-wide epigenetic control. Another example of a possible role for methylation in (potentially) active chromatin regions recently came from plants, in which extensive specific methylation of gene bodies was discovered (*23*). These results, together with the findings introduced here, should prompt reevaluation of the role of global DNA methylation that occurs away from gene promoters as well as the apparently complex relationship with chromatin activity.

Note added in proof. A second manuscript reporting gene body methylation in plants was recently published (*29*).

References and Notes

1. A. Bird, *Genes Dev.* **16**, 6 (2002).
2. T. Hashimshony, J. Zhang, I. Keshet, M. Bustin, H. Cedar, *Nat. Genet.* **34**, 187 (2003).
3. T. Mohandas, R. S. Sparkes, L. J. Shapiro, *Science* **211**, 393 (1981).
4. S. L. Gilbert, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13825 (1999).
5. E. Heard, *Curr. Opin. Cell Biol.* **16**, 247 (2004).
6. L. F. Lock, D. W. Melton, C. T. Caskey, G. R. Martin, *Mol. Cell. Biol.* **6**, 914 (1986).
7. L. F. Lock, N. Takagi, G. R. Martin, *Cell* **48**, 39 (1987).
8. S. F. Wolf, D. J. Jolly, K. D. Lunnen, T. Friedmann, B. R. Migeon, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2806 (1984).
9. P. H. Yen, P. Patel, A. C. Chinault, T. Mohandas, L. J. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1759 (1984).
10. E. Viegas-Pequignot, B. Dutrillaux, G. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7657 (1988).
11. Another recent study came to a conclusion that there is hypomethylation at gene-poor regions of the X chromosome, leading to an overall hypomethylation of Xi (*24*). Their conclusions are based entirely on the assumption that the single X chromosome of males is identical to Xa in female cells. In any case, our direct allele-specific analyses reveal only a modest hypomethylation of Xi in gene-poor regions, but a strong signature of Xi hypomethylation in gene bodies.
12. Details about the Affymetrix 500K SNP mapping array are available online (www.affymetrix.com).
13. A number of medium- to high-throughput assays have been described to analyze DNA methylation (*15*, *23*–*28*). Our methodology is most similar to the assay that used the 10K array (*28*).
14. The cocktail includes Aci I, BsaH I, Hha I, Hpa II, and HpyCH4 IV.
15. A. Schumacher *et al.*, *Nucleic Acids Res.* **34**, 528 (2006).
16. R. C. Allen, H. Y. Zoghbi, A. B. Moseley, H. M. Rosenblatt, J. W. Belmont, *Am. J. Hum. Genet.* **51**, 1229 (1992).
17. Materials and methods are available as supporting material on *Science* Online.
18. L. Carrel, H. F. Willard, *Nature* **434**, 400 (2005).
19. L. M. Hoffman *et al.*, *Stem Cells* **23**, 1468 (2005).
20. P. J. Rugg-Gunn, A. C. Ferguson-Smith, R. A. Pedersen, *Nat. Genet.* **37**, 585 (2005).
21. D. Frank *et al.*, *Nature* **351**, 239 (1991).
22. D. K. Nguyen, C. M. Disteche, *Nat. Genet.* **38**, 47 (2006).
23. X. Zhang *et al.*, *Cell* **126**, 1189 (2006).
24. M. Weber *et al.*, *Nat. Genet.* **37**, 853 (2005).
25. M. Bibikova *et al.*, *Genome Res.* **16**, 383 (2006).
26. I. Keshet *et al.*, *Nat. Genet.* **38**, 149 (2006).
27. C. M. Valley, H. F. Willard, *Curr. Opin. Genet. Dev.* **16**, 240 (2006).
28. E. Yuan *et al.*, *Cancer Res.* **66**, 3443 (2006).
29. D. Zilberman, M. Gehring, R. K. Tran, T. Ballinger, J. Henikoff, *Nat. Genet.* **39**, 61 (2007).

30. We thank H. Cedar, D. Housman, and J. Lee for discussions and comments and the staff of Harvard Medical School–Partners Healthcare Center for Genetics and Genomics Microarray Facility for Affymetrix array experiments. Support came from the NIH (to A.C.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/315/5815/1141/DC1
Materials and Methods
SOM Text
Figs. S1 to S10
Tables S1 to S9

16 October 2006; accepted 22 January 2007
10.1126/science.1136352

Reversal of Neurological Defects in a Mouse Model of Rett Syndrome

Jacky Guy,[1] Jian Gan,[2] Jim Selfridge,[1] Stuart Cobb,[2] Adrian Bird[1]*

Rett syndrome is an autism spectrum disorder caused by mosaic expression of mutant copies of the X-linked *MECP2* gene in neurons. However, neurons do not die, which suggests that this is not a neurodegenerative disorder. An important question for future therapeutic approaches to this and related disorders concerns phenotypic reversibility. Can viable but defective neurons be repaired, or is the damage done during development without normal MeCP2 irrevocable? Using a mouse model, we demonstrate robust phenotypic reversal, as activation of MeCP2 expression leads to striking loss of advanced neurological symptoms in both immature and mature adult animals.

Mutations in the X-linked *MECP2* gene are the primary cause of Rett syndrome (RTT), a severe autism spectrum disorder with delayed onset that affects 1 in 10,000 girls (*1*). *MECP2* mutations are also found in patients with other neurological conditions, including learning disability, neonatal encephalopathy, autism, and X-linked mental