University of Massachusetts Medical School
eScholarship@UMMS

Hematology/Oncology | Pediatric Divisions

2-23-2010

X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils

Yukio Yasukochi
Yale University School of Medicine

Osamu Maruyama
Kyushu University

Milind C. Mahajan
Yale University School of Medicine

See next page for additional authors

Follow this and additional works at: http://escholarship.umassmed.edu/peds_hematology

 Part of the Hematology Commons, Oncology Commons, and the Pediatrics Commons

Repository Citation

Yasukochi, Yukio; Maruyama, Osamu; Mahajan, Milind C.; Padden, Carolyn; Euskirchen, Ghia M.; Schulz, Vincent; Hirakawa, Hideki; Kuhara, Satoru; Pan, Xing-Hua; Newburger, Peter E.; Snyder, Michael; and Weissman, Sherman M., "X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils" (2010). *Hematology/Oncology*. 128.
http://escholarship.umassmed.edu/peds_hematology/128

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Hematology/Oncology by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils

Authors
Yukio Yasukochi, Osamu Maruyama, Milind C. Mahajan, Carolyn Padden, Ghia M. Euskirchen, Vincent Schulz, Hideki Hirakawa, Satoru Kuhara, Xing-Hua Pan, Peter E. Newburger, Michael Snyder, and Sherman M. Weissman

Comments
Publisher PDF posted as allowed by the publisher's author rights policy at http://www.pnas.org/site/misc/authorfaq.shtml.

Rights and Permissions
Citation: Proc Natl Acad Sci U S A. 2010 Feb 23;107(8):3704-9. Epub 2010 Feb 2. Link to article on publisher's website

This article is available at eScholarship@UMMS: http://escholarship.umassmed.edu/peds_hematology/128

X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils

Yukio Yasukochi[a], Osamu Maruyama[b], Milind C. Mahajan[a], Carolyn Padden[c], Ghia M. Euskirchen[d], Vincent Schulz[e], Hideki Hirakawa[f], Satoru Kuhara[g], Xing-Hua Pan[a], Peter E. Newburger[c], Michael Snyder[d], and Sherman M. Weissman[a,1]

Departments of [a]Genetics, [d]Molecular, Cellular, and Developmental Biology, and [e]Pediatrics, Yale University School of Medicine, New Haven, CT 06520-8064; [b]Faculty of Mathematics, Kyushu University, Nishi-ku, Fukuoka-City 819-0395, Japan; [c]Department of Pediatrics, University of Massachusetts Medical School, Worcester, MA 01655; [f]Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan; and [g]Graduate School of Genetic Resources, Kyushu University, Higashi-ku, Fukuoka 812-8581, Japan

Contributed by Sherman M. Weissman, December 23, 2009 (sent for review October 6, 2009)

The DNA methylation status of human X chromosomes from male and female neutrophils was identified by high-throughput sequencing of HpaII and MspI digested fragments. In the intergenic and intragenic regions on the X chromosome, the sites outside CpG islands were heavily hypermethylated to the same degree in both genders. Nearly half of X chromosome promoters were either hypomethylated or hypermethylated in both females and males. Nearly one third of X chromosome promoters were a mixture of hypomethylated and heterogeneously methylated sites in females and were hypomethylated in males. Thus, a large fraction of genes that are silenced on the inactive X chromosome are hypomethylated in their promoter regions. These genes frequently belong to the evolutionarily younger strata of the X chromosome. The promoters that were hypomethylated at more than two sites contained most of the genes that escaped silencing on the inactive X chromosome. The overall levels of expression of X-linked genes were indistinguishable in females and males, regardless of the methylation state of the inactive X chromosome. Thus, in addition to DNA methylation, other factors are involved in the fine tuning of gene dosage compensation in neutrophils.

high-throughput sequencing | DNA methylation | X inactivation

The mammalian X chromosome is unique in many ways. Males have one copy and females have two copies of the X chromosome, although both genders have two copies of each autosome in diploid cells. To compensate for the difference in gene dosage of the X chromosome between the genders, genes on one of two X chromosomes in females are active and those on the other are subjected to X chromosome inactivation (1). Analysis of the active and inactive X chromosomes in rodent/human somatic cell hybrids and nonrandomly inactivated primary fibroblast cultures showed that about 15% of the genes on the X chromosome escape X chromosome inactivation to some degree and are expressed from both the active and inactive X chromosomes. Also, an additional 10–20% of genes on the X chromosome show reactivation in some but not all cell lines (2). Recent large-scale analysis of gene expression in lymphoblastoid cell lines using microarrays showed that only about 5% of X-linked genes have increased expression in females compared with males (3). This difference in the proportion of genes escaping X chromosome inactivation may be explained by the different approaches of the studies. For example, both X chromosomes in females might express a gene, but the level of expression might be down-regulated compared with the expression from the single X chromosome in males. Also, the selection of escaped genes could be cell type-specific or influenced by in vitro culture conditions.

DNA methylation appears to be a part of the mechanism for human X chromosome inactivation, because 5-azacytidine treatment can reactivate hypoxanthine-guanine phosphoribosyltransferase (HPRT) on the inactive X chromosome (4). The cluster of CpG dinucleotides 5′ to the HPRT gene is extensively hypomethylated on active alleles, whereas the inactive alleles are nonuniformly and less extensively hypomethylated at the 5′ CpG cluster (5). Recently, the chromosome-wide methylation status at promoters was analyzed using an immunoprecipitation method coupled with a promoter oligonucleotide microarray (6). Only CpG island-embedded promoters of genes that undergo X chromosome inactivation showed hypermethylation, whereas the promoters without CpG islands did not show differential DNA methylation in relation to their X chromosome inactivation status. In autosomes, the vast majority of promoters within CpG islands are hypomethylated, whereas those without CpG islands may be methylated (6–8). The observed differences in the promoters with CpG islands between autosomes and sex chromosomes in females have been hypothesized to be responsible for silencing of the inactive X chromosome. For promoters without CpG islands, inactivation was not reflected by changes in DNA methylation (6). Recently, Meissner et al. (9) reported the genome-wide analysis of methylation in pluripotent and differentiated cells using MspI digestion of the genomic DNA followed by Solexa (Illumina) sequencing of bisulfite-treated MspI fragments. CpG islands in female cell populations (ES cell-derived and primary astrocytes, respectively) showed an average of ~50% methylation, consistent with hypermethylation of high CpG density promoters on the inactivated X chromosome. These investigators also noted that there may be relatively rapid changes in the patterns of DNA methylation of cells placed into tissue culture.

To revisit the DNA methylation status and gene expression from X chromosomes, especially for the escapee genes on the inactive X chromosome in normal nondividing cells from females, we employed digestion with CpG methylation-sensitive and -insensitive restriction endonucleases (HpaII and MspI, respectively) and submitted the small restriction fragments for Solexa sequencing. We found that the majority of sites at the promoter regions on the inactive X chromosome in neutrophils polymorphonuclear leukocytes (PMNs) were composed of hypermethylated sites or a mixture of hyper- and hypomethylated sites but that most of the sites at the promoters of escapee genes were hypomethylated. The levels of expression of X-linked

Author contributions: X.-H.P. and S.M.W. designed research; Y.Y. and M.C.M. performed research; C.P. and P.E.N. contributed new reagents/analytic tools; O.M., G.M.E., V.S., H.H., S.K., M.S., and S.M.W. analyzed data; and Y.Y. and S.M.W. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: sherman.weissman@yale.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0914812107/DCSupplemental.

genes were indistinguishable in females and males, regardless of their state of methylation.

Results

Comparison of Methylation States at Promoter Regions Between Autosomes and X Chromosomes of Female and Male Neutrophils. To identify hypo- and hypermethylated regions on the female and male chromosomes, we first digested genomic DNA from male and female PMNs with Tsp509I (AATT recognition site). CpG-rich fragments obtained after Tsp5091 digestion were selected by agarose gel electrophoresis and extraction of fragments from 500–4,000 bp in length. These size-selected fragments were digested with HpaII, and the smaller fragments (70–350 bp) were isolated for sequencing. The larger fragments (500–4,000 bp) from the HpaII digestion were reisolated and cut with MspI, and, again, fragments from 70–350 bp were size-selected in an agarose gel and subjected to Solexa sequencing. Among the total Solexa sequences, those having a CCGG sequence at the 5′ end of an identified position on the X chromosome were selected for further analysis. The extent of hypo- or hypermethylation is reflected in the number of Solexa read sequences from either HpaII or MspI sites, respectively, at the identified sites. For the numbers of sequences from the HpaII and MspI sites to be compared with each other to determine their statistical significance, let f_1 and f_2 be the frequencies of Solexa read sequences starting at a given site of a chromosomal sequence in HpaII and MspI samples, respectively. From the frequency pair (f_1, f_2), the category of the site is determined. All frequencies are log-scaled, because Solexa read sequences are PCR-amplified. Suppose that the pair ($\log f_1$, $\log f_2$) is plotted on the *x*-*y* plane, and the closest line from the point is determined between the *x* axis, the *y* axis, and the line $y = x$, which correspond to C1 (hypomethylated), C2 (hypermethylated), and C3 (heterogeneously methylated), respectively. If the plotted point is located around the origin of the *x*-*y* coordinate, the category is set to C4 (insufficient). In addition, if the point is at equal distances from $y = x$ and the *x* axis or *y* axis, the category is assigned C5 (ambiguous). When we use multiple samples, a normalized frequency is calculated from the original frequencies of the samples. The normalization is done separately for females and males using the logarithms of the frequencies (see *Materials and Methods* for details).

First, we compared the methylation states at promoter regions (−3,000 to 1,000 bp from the transcription start site) on the female and male X chromosomes with those on the female and male autosomes. To avoid specific clusters of related genes on each chromosome, such as the tumor antigen cluster genes on the X chromosome, only sequences with unique matches between the sequence and the reference genome were used to identify the sequences at the promoter regions. In autosomes, 94% of the sites identified at promoter regions were hypomethylated and only 4% of the sites were hypermethylated in both females and males (Table 1). The DNA methylation states were almost indistinguishable between female and male autosomes. Interestingly, a small but significant number of the sites were identified as heterogeneously methylated in both females and males or heterogeneously methylated in females and hypomethylated in males. These might relate to monoallelic expression or to sexually different expression of genes (Table S1 as an example of the methylation status of the imprinted MEST gene for monoallelic expression).

In the male X chromosome, 89% of the sites in promoter regions were hypomethylated (C1) to almost the same degree as those in the autosomes (Table 1). In the female X chromosomes, 43% of the sites in promoters were hypomethylated (C1), whereas the hypermethylated state (C2, C2) at the same sites in both female and male promoters were at almost the same low level as that in the autosomes. A prominent difference in methylation either between the female X chromosomes and autosomes or between the female and male X chromosomes was that 46% of the sites in the promoter regions on the X chromosome were heterogeneously methylated (C3) in females (Table 1). Presumably, this reflects X chromosome inactivation, in which the promoters subjected to X chromosome inactivation were hypermethylated.

Table 1. DNA methylation states at promoter regions on autosomes and X chromosomes of female and male neutrophils

Categories					
Female	Male	Autosome*	%	X chromosome*	%
C1	C1	92,395	94.1	940	43.1
C1	C2	11	0.01	0	0
C1	C3	78	0.08	0	0
C1	C5	36	0.04	0	0
C2	C1	71	0.07	86	3.94
C2	C2	3,904	3.98	77	3.53
C2	C3	204	0.21	1	0.05
C2	C5	33	0.03	0	0
C3	C1	487	0.50	1,003	45.9
C3	C2	95	0.10	4	0.18
C3	C3	654	0.67	10	0.46
C3	C5	56	0.06	1	0.05
C5	C1	85	0.09	61	2.79
C5	C2	19	0.02	0	0
C5	C3	35	0.04	0	0
C5	C5	6	0.01	0	0
Total positions		98,169		2,183	

Only sequences with unique matches with 0, 1, and 2 errors in the match between the sequence and the reference genome were used to identify the sequences at the promoter regions (−3,000 to 1,000 bp from the transcription start site). The numbers of sequences from the HpaII and MspI sites are compared with each other to determine their statistical significance after normalization, and sequences are classified into five categories as described in *Materials and Methods*. Category 1 (C1) donates a hypomethylated state, whereas category 2 (C2) denotes a hypermethylated state. Category 3 (C3) defines a heterogeneously methylated state. Category 5 (C5) means a borderline either between C1 and C3 or between C2 and C3.
*Identified numbers of CCGG sites at the promoter on the autosome or the X chromosome.

Methylation States in Promoter Regions of the Female and Male X Chromosomes. Next, we compared the methylation states on the female and male X chromosomes in detail using the sequences with both unique and four or fewer matches to the reference genomic DNA, with zero to two errors. The sequences with the low multiple matches included genes in the pseudoautosomal regions and tumor antigen cluster genes that match to more than one genomic position. The promoter region was separated into five segments relative to the transcription start site (Table 2). Among the promoter sites for which the methylation status could be identified, 82% were in CpG islands. There were essentially no C3 sites (0.8%, 23 of 2,881 sites) in the male promoters regardless of the presence of CpG islands. C1 and C2 sites constituted 85% (2,007 of 2,354 sites) and 14.2% (335 of 2,354 sites), respectively, on the entire length of the male promoters (including promoter segments A–E) with CpG islands, indicating that the CpG island-embedded male promoters were extensively hypomethylated (Table 2). In contrast, among the sites outside CpG islands, we found a somewhat lesser fraction (61%, 323 of 527 sites) of the C1 category sites in the male promoters, whereas C2 sites were increased to 36% (191 of 527 sites). This finding is consistent with the previous reports that more methylation is observed in non-CpG island promoters than in CpG islands (6, 8). In the case of female promoters, C2 category sites (21%, 613 of 2,881 sites) were almost the same as those in the male promoters (18.3%, 526 of 2,881 sites) regardless of the

Table 2. Methylation status of promoters on female and male X chromosomes

Regions		Promoter A		Promoter B		Promoter C		Promoter D		Promoter E		Subtotal		
		CpG island												
Female	Male	(+)	(–)	(+)	(–)	(+)	(–)	(+)	(–)	(+)	(–)	(+)	(–)	Total
C1	C1	60	21	63	23	278	47	378	43	100	18	879	152	1,031
			(31%)				(5.3%)				(36%)			
C2	C1	4	1	2	0	26	5	36	6	8	1	76	13	89
			(2.6%)				(0.5%)				(3.1%)			
C3	C1	63	33	71	23	275	28	413	46	161	19	983	149	1,132
			(34%)				(5.2%)				(39%)			
C5	C1	7	2	1	1	24	5	32	0	5	1	69	9	78
			(2.4%)				(0.3%)				(2.7%)			
C2	C2	37	42	24	20	14	56	181	44	77	26	333	188	521
			(11.6%)				(6.5%)				(18.1%)			
Miscellaneous		0	3	0	2	5	3	9	7	0	1	14	16	30
			(0.5%)				(0.6%)				(1.0%)			
Total			2,354				527				2,881			
			(82%)						(18.3%)					

Sequences with unique and four or fewer matches with 0, 1, and 2 errors in the match between the sequence and the human genome were used to identify the number of sites at the promoter regions. The promoters are divided into five regions: promoter A (−3,000 to −1,001 bp from the transcription start site), promoter B (−1,000 to −501 bp), promoter C (−500 to −1 bp), promoter D (1–500 bp), and promoter E (501–1,000 bp). The symbols (+) and (–) in CpG islands indicate the promoters with and without CpG islands, respectively. Numbers in the promoters indicate identified numbers of the CCGG sites in the promoters. Numbers in parentheses indicate percentages of each category of promoters.

presence of CpG islands. Overall, C1 and C3 sites were almost equally distributed in the female promoters with CpG islands, at 31% (879 of 2,881 sites) and 35% (997 of 2,881 sites), and in the female promoters without CpG islands, at 5.3% (152 of 2,881 sites) and 5.6% (162 of 2,881 sites), respectively.

Methylation Status Outside of Promoter Regions. The intergenic and intragenic regions on the X chromosome were markedly different in their methylation states compared with the promoters. The sites without CpG islands were heavily hypermethylated to the same degree in both females and males. Overall, 60% and 65% of sites were hypermethylated (C2, C2) in the intergenic and intragenic regions, respectively (Table S2). We examined the sites in the intergenic and intragenic regions assigned on the basis of the human mRNA data from the University of California at Santa Clara (UCSC) Genome Browser. As shown in parentheses in Table S2, 19.4% (96 of 496) and 33% (76 of 230) of (C1, C1) sites in the intergenic and intragenic regions, respectively, corresponded to promoters (−3,000 to 1,000 bp) of annotated human mRNAs. Also, 18.6% (55 of 295) and 45% (80 of 178) of the (C3, C1) sites in the intergenic and intragenic regions, respectively, corresponded to these annotated promoters.

Characterization of the Promoter Methylation States in mRNAs Recorded in GenBank. We examined the distribution of methylation states at the promoters in males and females. At least one CCGG site could be evaluated in the promoters of 716 RefSeq genes, corresponding to about 58% of the RefSeq genes on the X chromosome. The promoters were mainly composed of fully methylated sites (C2, C2) (33%, 237 of 716), and a mixture of partly methylated (C3, C1) and unmethylated (C1, C1) (34%, 240 of 716) sites. Fig. 1*A* shows an example of a promoter that is hypermethylated in both female and male cells, a schematic of the CT47 (cancer/testis) gene family, and one of the tumor antigen gene clusters. The genes in the region containing G6PD were heterogeneously methylated (Fig. 1*B*), and the genes in pseudoautosomal region 1 and evolutionarily younger strata 4 and 5 were hypomethylated (Fig. 1*C*). We also examined the distribution of methylation sites at the promoters of human mRNAs recorded in GenBank but not annotated in RefSeq. The methylation states of 27 mRNAs not recorded in RefSeq were identified in the intergenic regions and were found to be composed mainly of a mixture of (C1, C1) and (C3, C1) (14 of 27 promoters; see Table S3 for details). The methylation states of 55 variants in the intragenic regions were evaluated and mainly constituted one set each of (C1, C1) (9 of 55 promoters) or (C2, C2) (17 of 55 promoters) and a mixture of (C1, C1) and (C3, C1) (12 of 55 promoters) (Table S3).

The promoters with (C1, C1) in the RefSeq genes as well as the mRNAs in GenBank were concentrated in gene clusters escaping from X chromosome inactivation, especially in pseudoautosomal region 1 and the younger strata of the X chromosome. Because the hypomethylated state is likely necessary for gene expression, C1 at the promoters on the female X chromosome may represent the genes that have escaped or have the proclivity to escape from X chromosome inactivation. Because the hypermethylated state is known to repress gene expression, C3 at the promoter in females should represent the genes stringently subjected to X chromosome inactivation.

To evaluate this hypothesis, we compared the methylation states at the promoters on the X chromosomes with the Xi-expression results from rodent/human somatic cell hybrids described by Carrel and Willard (2). We edited the Xi-gene expression results (excluding ESTs) in the cell hybrids using their data. The sites in A, B, C, D, and E segments of promoters were used for the comparison of the methylation status at the promoter and the Xi-expression in the hybrid cells. Seventy-four percent of the Ξ-genes were expressed in none or one and two of the four to nine cell hybrids tested (Table S4*A*). Ninety-five percent of genes with (C3, C1) (36 of 38 genes) and 86% of genes with the mixture of (C1, C1) and (C3, C1) (124 of 144 genes) belonged to the Xi-genes that rarely escaped inactivation, indicating a high correlation between hypermethylation states at the promoter in female cells and infrequent escape from X inactivation (Table S4 *A* and *B*). The promoters of 52% of 311 genes showed these methylation patterns, whereas 74% of the genes were stably inactivated in the hybrids. This low percentage of X-inactivated genes estimated by their methylation states was mainly attributable to uncounted genes within the remaining categories, which were also X-inactivated (Table S4*B*). However, 56% of genes with (C1, C1) corresponded to the Ξ-genes expressed in all or seven and eight of the hybrids tested (Table S4*A*). Considering only the promoter regions that had

(C1, C1) at more than two sites, 86% were among the Ξ-genes expressed in all or most of the hybrids. These genes represented 14.2% (44 of 311 of all measured genes), in good agreement with 15.9% of the Ξ-genes expressed in all or most of the hybrids (2).

We further examined whether there were any differences in the distribution of X-inactivated and escapee genes estimated by the methylation states at the promoter compared with inactivated and escapee genes in the hybrids. Two hundred twenty-nine X-inactivated and 57 escapee genes from a total 311 genes were detected in the hybrids (2). Of these, 70% of X-inactivated (160 of 229) and 77% of escapee (44 of 57) genes could be evaluated for the methylation state of their promoters. Fig. 2 shows the distribution of the X-inactivated and escapee genes, estimated by methylation status and Ξ-gene expression. The 59 genes belonging to category (C1, C1) at more than two promoter sites were compared with the escapee genes identified earlier by Ξ-expression (2). Similarly, 53 genes with category (C3, C1) and 210 with the mixture of categories (C3, C1) and (C1, C1) were also compared with the X-inactivated genes identified by the Ξ-expression method (2). These comparisons revealed that the distribution of X-inactivated and escapee genes was quite similar by both methods. The escapee genes tended to cluster and were mapped mainly to evolutionarily younger strata 3, 4, and 5 in addition to the pseudoautosomal region 1 (Fig. 2). In addition, a few escapee genes were scattered in the evolutionary oldest stratum 1. One cluster of the escapee genes identified by Ξ-expression was located in a gene-rich region of Xq28, including 5 Ξ-genes that were expressed in most of the hybrids. However, these five genes were not in category (C1, C1) at two or more sites. Three additional genes in category (C1, C1) that mapped in this region were not clustered (Table S5). Of these three genes, one was Ξ-inactivated and two showed variable Ξ-expression.

VAMP7 is exceptional in that it is located in the pseudoautosomal region 2 and is expected to escape X inactivation, but the Ξ-gene was expressed in only one of nine hybrids. The methylation states showed a mixture of two sites of category (C3, C3) and one site of category (C1, C1) at the female and male promoters, consistent with the Ξ-expression results (2) as well as with another report showing that VAMP7 is inactivated on both the inactive X chromosome and the Y chromosomes (10).

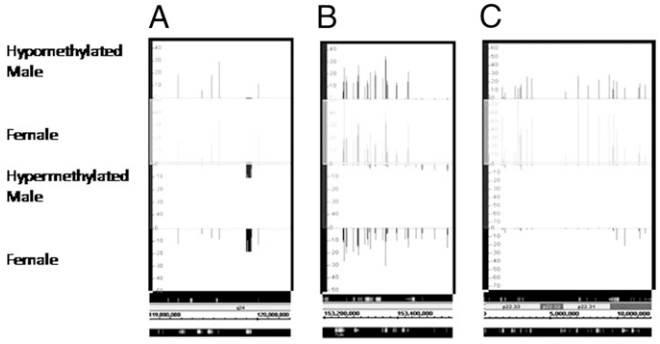

Fig. 1. Hypermethylated, hypomethylated, and heterogeneously methylated promoters on the female and male X chromosomes. (*A*) Hypermethylation of the promoters of the cancer/testis CT47 gene family. Heterogeneously methylated promoters are displayed on the left and right sides of the CT47 family. (*B*) Heterogeneous methylation at the promoters in an area containing G6PD. The hypermethylated promoters of CTAG1 are displayed on the right side. (*C*) `Hypomethylation of the promoters at the tip of the short arm. Heterogeneously methylated promoters are displayed on the right side. The height of the lines on the *x* and *y* axes indicates the number of sequences and their positions either from the unmethylated (HpaII) or methylated (MspI) sites, respectively. Hypomethylated state: male (*Upper*) and female (*Lower*). Hypermethylated state: male (*Upper*) and female (*Lower*). The numbers in the white bar below the display indicate the position of the sequences on the X chromosome (Human March 2006 Assembly, hg18). The white lines above and below the white bar indicate the position of genes, with solid boxes indicating exons and narrow lines indicating introns.

Relation Between Gene Expression and DNA Methylation of Promoter Regions. Gene expression in neutrophils was analyzed with Human HT-12 Expression BeadChips (Illumina) using total neutrophil RNA from a female subject and a male subject. After fluorescence signals of all chromosomes were normalized by a quantile method, $\log_2$ gene expression signals of the female X chromosome were plotted against those of the male X chromosome (Fig. 3). There were no obvious differences in relative gene expression levels between female and male X chromosomes. Median female and male gene expression levels of the X chromosomes were the seventh highest among the gene expression levels of all chromosomes.

Next, we examined the relation between gene expression and DNA methylation states at the promoter. Ninety-three percent (51 of 55) of genes with category (C2, C2) at the promoter did not express either female or male neutrophils. The two genes expressed in both male and female cells with category (C2, C2) promoters were SLC35A2 and ASB11. However, they were evaluated at only one site in promoter regions A and B, respectively. The genes with category (C1, C1) at more than two sites or with category (C3, C1) were equally likely to be expressed, 48% (26 of 54 genes) and 45% (23 of 51 genes), respectively, in both female and male neutrophils. Fifty-one percent (98 of 192) of genes with a mixture of categories (C3, C1) and (C1, C1) were expressed in both female and male cells.

Discussion

Previous cytological studies of the methylation profile over the entire inactive female X chromosome have been carried out in dividing cells in metaphase (11, 12), which does not take into account cell cycle-dependent dynamics of methylation patterns and does not provide high resolution. Additionally, the pattern of DNA methylation can change relatively rapidly when cells are placed in culture (9). We selected peripheral blood neutrophils for genome-wide methylation analysis because they can be obtained easily as a 95–99% pure nonreplicating cell population that is sufficient for genome-wide analysis of their methylome. Also, all the isolated cells are of approximately the same age, because their half-life in the blood is only about 6–10 h.

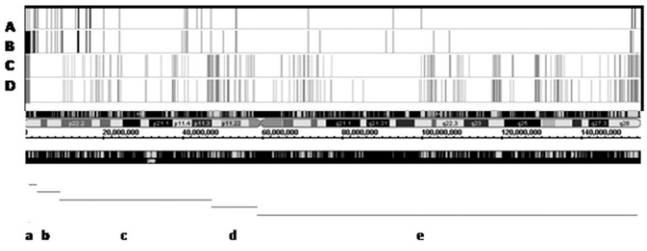

Fig. 2. Distributions of X-inactivated and escapee genes estimated by DNA methylation status at the promoters and determined by Ξ-expression in the hybrid cells (2). For distribution in Ξ-expression, 57 escapee and 229 X-inactivated genes were used from the 311 genes whose methylation status could be estimated. For distribution in the methylation status, 59 genes with (C1, C1) at three or more sites in the promoter regions were used for the escapee genes, whereas 210 genes with (C3, C1) and with a mixture of (C3, C1) and (C1, C1) were used for the X-inactivated genes. Distribution of escapee and X-inactivated genes was determined by Ξ-expression in cell lines: escapee genes (*A*) and X-inactivated genes (*C*). Distribution of escapee and X-inactivated genes was estimated by their methylation status: escapee genes (*B*) and X-inactivated genes (*D*). Traverse line, a, indicates pseudoautosomal region 1 and stratum 5. Traverse lines b, c, d, and e, indicate strata 4, 3, 2, and 1, respectively. Other annotations are the same as in Fig. 1.

MEDICAL SCIENCES

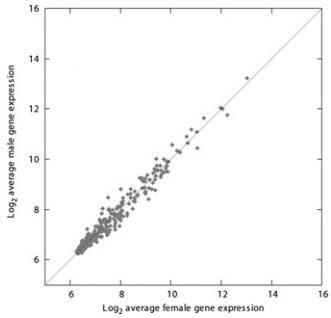

Fig. 3. Scatterplot of average female and male gene expressions in X chromosomes. Gene expression was analyzed with Human HT-12 Expression BeadChips using total RNA from female and male neutrophils. After fluorescence signals of all chromosomes were normalized by a quantile method, $\log_2$ gene expression signals of the female X chromosome were plotted against those of the male X chromosome.

Methylation Status of Female X Chromosome Promoters. X chromosome inactivation is initiated by a cis-acting noncoding RNA, XIST, and final DNA methylation is established at the promoters of genes on the inactive X chromosome after several steps of epigenetic marking during cell differentiation (13). CpG islands, often associated with the 5′ end of genes, are reported to be heavily methylated on the inactive X chromosome in somatic cells (6, 9). However, our findings demonstrate that 82% of identified potential methylation sites in the promoter regions are in CpG islands, and many sites on the inactive X chromosome were not methylated in females. Genes unmethylated at more than two sites on the female promoter were generally among the genes that were found to be expressed from the inactive X chromosome in cultured cells. A high correlation was also observed between partial methylation of the CpG sites in promoter regions and the lack of Ξ-expression in cultured cells. We also identified 15 genes of category (C3, C1) and 66 genes with a mixture of (C3, C1) and (C1, C1) not previously reported as stably X-inactivated genes (2). Our results also suggest that lack of methylation at more than two sites on the promoter is a signature for the escapee genes. This signature pattern was found at the previously reported escapee gene promoters as well as at the following uniquely identified potential escapee genes in this study: *SHROOM2*, *EGFL6*, *ZRSR2*, *DGKK*, *UPRT*, *PCDH11X*, *FAM133A*, and *LOC389904*. Among these, *ZRSR2* is reported to be more highly expressed in females than in males (3). These genes, with the exception of *DGKK* and *UPRT*, are located in the midst or at the boundary of the escapee gene clusters.

The distribution of escapee genes determined by Ξ-expression was similar to that of the genes classified as potential escapee genes based on their methylation status (Fig. 2). When we examined the methylation states of X-linked genes with Y homology, 81% (21 of 26) of the genes, excluding pseudoautosomal genes, were in category (C1, C1) with more than two sites at their promoters and most had been found to be escapee genes. Among these genes, *BCOR* was not considered to be an escapee because it was expressed in only two of nine hybrids. The methylation states of the 5 remaining genes were a mixture of (C1, C1) and (C3, C1) in *MAP3K7IP3*, *CASK*, and *TSPYL2* and (C2, C1) in *RBMX*; all were inactive in the Ξ-expression studies, with the exception of (C2, C2) in *FAM9C*, which was described as an escapee gene. The escapee genes tended to cluster, and the clustered genes were mapped mainly to the evolutionarily younger strata 3, 4, and 5 in addition to pseudoautosomal region 1. There was a difference in the escapee gene clustering in the oldest stratum 1 as determined by Ξ-expression in the hybrids and estimated by the methylation states at the promoter. One cluster of the escapee genes by Ξ-expression was located in a gene-rich region of Xq28. This is consistent with the suggestion that the region actually belongs to stratum 2, based on a comparison with the chicken genomic sequence (2, 14). However, these promoters were composed mainly of mixtures of categories (C3, C1) and (C1, C1). Although this distribution might be cell type-specific, sporadic distribution of escapee genes through the region may simply reflect the distance from the XIST gene.

Relation Between the Methylation States at the Promoters and Gene Expression. We did not find any difference in the expression level either between the female X chromosome and the autosomes or between the female and male X chromosomes at the expression level, confirming previous reports (3). There were no genes in females that were expressed greater than 1.6 times more than in males. Accordingly, compared with autosomal genes, 2-fold upregulation of expression was observed from male X chromosome genes and from the active alleles of female X chromosome genes subjected to X inactivation (Fig. 2). If any genes escaped X inactivation in neutrophils, this did not cause a detectable increase in the level of their expression above that seen from the single male X chromosome. This could occur if escape from inactivation was incomplete, as has been observed previously, or because the overall level of expression was regulated so that the expression from the two copies in the female is less than twice that of the single copy in the male.

It is generally accepted that hypomethylation at the promoter is permissive for gene expression, whereas hypermethylation of promoters in CpG islands causes repression of gene expression. Our results showed high correlation between hypermethylation at the promoter and transcriptionally silent genes. However, the hypomethylation state (C1, C1) at more than two positions was not correlated with gene expression, and only about 48% of hypomethylated genes were expressed at the mRNA level. Forty-five percent and 51%, of genes with category (C3, C1) or with a mixture of categories (C3, C1) and (C1, C1), respectively, were expressed, which approximates the expression of genes with category (C1, C1) at more than two sites in both females and males. There were almost no differences in gene expression between the genes with categories (C1, C1) and (C3, C1). These results are largely compatible with the RNA polymerase II recruitment status at the promoters analyzed by ChIP-ChIP experiments, where the frequencies of active promoters in high-CpG and intermediate-CpG content were 66% and 41%, respectively (6). Overall, there appear to be two mechanisms of inactivation of genes on the X chromosome, only one of which involves methylation of CpG islands in the promoter region. There is a strong correlation between the lack of methylation of promoters of a gene and the tendency of the gene to escape X inactivation in cultured cells.

Histone modifications are involved in X chromosome inactivation and gene expression. Gilbert and Sharp (15) investigated the histone H4 acetylation status of active and inactive human chromosomes using human/hamster somatic cell hybrids. They found that escapee genes had elevated levels of acetylation at their promoters on both chromosomes, whereas promoters of inactivated genes were markedly hypoacetylated. In gene expression studies, monomethylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 were all related to gene activation, whereas trimethylations of H3K27, H3K9, and H3K79 all correlated with gene repression. Also, acetylation of H3K27 was highly linked to promoter regions of transcriptionally active genes (16). Accordingly, we examined the relation between the histone modifications H3K27ac and H3K27me3 in GM12878 and K562 cell lines and DNA methylation status at promoters in neutrophils. Data on H3K27ac and H3K27me3 used were derived from the Encyclopedia of DNA Elements (ENCODE) project (17). Histone modifications were determined by the Broad Institute ChIP-sequence recorded in the UCSC Genome Browser. Analysis of genes lacking methylation at more than two sites at the promoter showed that 69% had sig-

nificant signals of H3K27ac but no significant signals of H3K27me3 in both cell lines, whereas 22% and 12% in GM12878 and K562, respectively, had no significant signals of H3K27ac but significant signals of H3K27me3. In contrast, among genes partially methylated at the promoter, 60% had significant signals of H3K27ac but no significant signals of H3K27me3, and 25% and 15% in GM12878 and K562, respectively, had no significant signals of H3K27ac but significant signals of H3K27me3 (Table S3). These analyses demonstrated almost no differences between the potential escapee and stably X-inactivated genes in histone modifications. Therefore, histone modifications tended to reflect the actual state of activity of the gene rather than its potential for escape from inactivation under other conditions. Thus, DNA methylation did not correlate closely with the histone modifications examined here and was a much better predictor of the potential for genes to escape X inactivation.

The differences between genes that escape X chromosome inactivation, estimated from the methylation states at the promoter in neutrophils, and those determined to escape from the microarray analysis in the lymphoblastoid cell lines may be explained by the different methods employed, as suggested for those in rodent/human somatic cell hybrids and nonrandomly inactivated primary fibroblasts (3). Alternatively, the methylation states at the promoter may provide a fundamental control for gene expression, but many other factors would be needed for the elevated levels of gene expression. In rodent/human somatic cell hybrids and nonrandomly X-inactivated primary fibroblasts, the normal control mechanism may be impaired and the methylation state at the promoter might become a primary determinant for the gene expression. To distinguish between these mechanisms better, further experiments such as analysis of the methylation state and allelic gene expression in normal clonal-derived cell populations (18) and individual cells from normal tissues will be necessary. Escape of individual X-chromosome genes from inactivation has been documented for the Jarid1C gene of the mouse (19, 20) and found to be an intrinsic property of the locus (20). Whether or where escape of other genes from X inactivation occurs in normal human tissues and, if so, whether this leads to gender differences in levels of gene expression in these tissues are important topics for further investigation.

Materials and Methods

Preparation of Genomic DNA and Total RNA. Neutrophils were isolated from venous blood of two male and two female volunteers using dextran sedimentation, centrifugation through Ficoll-Paque PLUS (Amersham Biosciences), and hypotonic lysis of erythrocytes (21). Genomic DNAs from male and female neutrophils were prepared using the DNeasy Tissue Kit (Qiagen). Total RNAs from a healthy male and female were prepared using the RNeasy Plus Mini Kit (Qiagen).

Preparation of HpaII and MspI Digested DNA Fragments for Identifying Hypo- and Hypermethylated States by Solexa Sequencing. The male and female genomic DNAs (10 μg) were digested with 80 units of Tsp509I (New England Biolabs) at 65 °C for 10 h. After purification of the digests with the QIAquick PCR purification kit (Qiagen), they were size-selected once in an E-Gel (1.2% agarose wt/vol) with SYBR safe (Invitrogen) and once in an E-gel (2% wt/vol agarose) with SYBR safe (Invitrogen) to obtain fragments from 500–4,000 bp. The fragments were digested with 20 units of HpaII (Roche), a CpG methylation-sensitive enzyme, in a final volume of 40 μL at 37 °C for 10 h. The digest was size-selected in an E-gel (2% wt/vol agarose) with SYBR safe to obtain the fragments from 70–350 bp that were used for identifying the hypomethylated state by Solexa sequencing. Also, the fragments from 500–4,000 bp were saved and digested with 40 units of MspI (New England Biolabs), a CpG methylation-insensitive enzyme, in a final volume of 40 μL at 37 °C for 10 h. The fragments from 70–350 bp were size-selected in an E-gel (2% wt/vol agarose) with SYBR safe, and were used for determining the hypermethylated state by Solexa sequencing.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants R01DK054369 and 1P50HG02357-01. O.M. was partially supported by a grant from the Kyushu University Global Centers of Excellence Program, "Center for Math-for-Industry," from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

1. Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* 190:372–373.
2. Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434:400–404.
3. Johnston CM, et al. (2008) Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet* 4:0088–0098.
4. Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. *Science* 211:393–396.
5. Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR (1984) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: Implications for X-chromosome inactivation. *Proc Natl Acad Sci USA* 81: 2806–2810.
6. Weber M, et al. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39:457–466.
7. Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16(1): 6–21.
8. Eckhardt F, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38:1378–1385.
9. Meissner A, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766–770.
10. D'Esposito M, et al. (1996) A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. *Nat Genet* 13:227–229.
11. Viegas-Pequignot E, Dutrillaux B, Thomas G (1988) Inactive X chromosome has the highest concentration of unmethylated Hha I sites. *Proc Natl Acad Sci USA* 85:7657–7660.
12. Prantera G, Ferraro M (1990) Analysis of methylation and distribution of CpG sequences on human active and inactive X chromosomes by in situ nick translation. *Chromosoma* 99:18–23.
13. Heard E, Disteche CM (2006) Dosage compensation in mammals: Fine-tuning the expression of the X chromosome. *Genes Dev* 20(14):1848–1867.
14. Kohn M, Kehrer-Sawatzki H, Vogel W, Graves JA, Hameister H (2004) Wide genome comparisons reveal the origins of the human X chromosome. *Trends Genet* 20(14): 598–603.
15. Gilbert SL, Sharp PA (1999) Promoter-specific hypoacetylation of X-inactivated genes. *Proc Natl Acad Sci USA* 96:13825–13830.
16. Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.
17. Birney E, et al., ENCODE Project Consortium NISC Comparative Sequencing Program Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center Broad Institute; Children's Hospital Oakland Research Institute (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
18. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318:1136–1140.
19. Xu J, Deng X, Disteche CM (2008) Sex-specific expression of the X-linked histone demethylase gene Jarid1c in brain. *PLoS One* 3:e2553.
20. Li N, Carrel L (2008) Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. *Proc Natl Acad Sci USA* 105:17055–17060.
21. Subrahmanyam YV, Baskaran N, Newburger PE, Weissman SM (1999) A modified method for the display of 3′-end restriction fragments of cDNAs: Molecular profiling of gene expression in neutrophils. *Methods Enzymol* 303:272–297.