Improved DNA methylation analysis via enrichment of demethylated cells expressing an X-inactivated transgene

Tammy W. Vallender and Bruce T. Lahn
University of Chicago, Chicago, IL, USA

BioTechniques 41:461-466 (October 2006)
doi 10.2144/000112246

The demethylating drug 5-aza-2′-deoxycytidine (5-aza-2dC) is frequently used to investigate the effect of global DNA demethylation on gene expression in cultured mammalian cells. Here, we describe a method that uses the reactivation of an X-inactivated green fluorescent protein (GFP) transgene as a marker to enrich for cells that have undergone drug-induced demethylation. By combining it with microarray gene expression profiling, we demonstrate the method's utility in identifying genes activated by global DNA demethylation.

INTRODUCTION

Epigenetic modifications of chromatin play essential roles in the regulation of gene expression (1–3). One key modification is DNA methylation, which is typically associated with stable gene silencing in mammals. Over the past 30 years, a myriad of cellular processes have been examined that might be under the regulation of DNA methylation. The best-studied examples include genomic imprinting, X chromosome inactivation, and silencing of repetitive elements, all associated with hypermethylation at silenced loci. It is increasingly recognized that DNA methylation is also important for the silencing of genes during normal cellular differentiation and during cancer progression (4–8). As such, there is growing interest in identifying genes whose silencing is associated with DNA methylation. To this end, researchers have devised several approaches, such as interfering with DNA methyltransferase activity by conditional knockout or RNA interference (RNAi)-induced knockdown of the DNA methyltransferase I (*Dnmt1*) gene, and using drugs that inhibit DNA methylation. However, technical issues persist with these methods, including the lack of global demethylation, possibly due to the presence of other DNA methyltransferases, heterogeneity of demethylated cells in RNAi-induced *Dnmt1* knockdown, and incomplete incorporation and toxicity of the methylation-inhibiting drugs (9–12).

One common approach to study DNA methylation is to treat cells with the demethylating drug 5-aza-2′-deoxycytidine (5-aza-2dC), a derivative of deoxycytidine, and identify genes activated by this treatment. 5-aza-2dC is incorporated into replicating DNA where it irreversibly binds to DNMT1, thus depleting maintenance methyltransferase activity in the cell. This results in global demethylation or hypomethylation. A problem with 5-aza-2dC treatment is that cells in culture do not respond equally to the drug. Given that only actively dividing cells incorporate 5-aza-2dC, nondividing or slow-dividing cells are not affected by the drug. Furthermore, cells in different stages of the cell cycle during drug treatment may undergo varying amounts of demethylation, and other factors, such as the rates of drug uptake and catabolism, may also differ among cells. Thus, when cultured cells are treated with 5-aza-2dC, the result is a heterogeneous population of cells with differing levels of demethylation.

Here, we report an approach that uses the reactivation of an X-inactivated green fluorescent protein (GFP) transgene in mice to enrich for cells that have undergone drug-induced demethylation. We show that this approach, when combined with microarray analysis, enhances the ability to identify genes activated by 5-aza-2dC treatment. This technique is widely applicable to dividing cells derived from the available X-linked GFP transgenic mouse line, including fibroblasts, osteoblasts, chondrocytes, myoblasts, hepatocytes, and many others. Additionally, the basic strategy can be extended to studies involving other chromatin-modifying drugs such as those that influence histones.

MATERIALS AND METHODS

Cell Culture and Drug Treatment

Mice used have an X-linked GFP transgene driven by the chicken β-actin promoter and cytomegalovirus (CMV) intermediate early enhancer in a 129/ICR mixed genetic background obtained from The Jackson Laboratory (Bar Harbor, ME, USA) (13). Male transgenic animals were bred to wild-type BL6 female mice to obtain females that were heterozygous for the transgene. Dermal fibroblasts were obtained from explant cultures of 1- to 3-day neonatal female mice from these crosses, using procedures as previously described (14). Cultures were maintained in Dulbecco's modified Eagle's medium (DMEM; Invitrogen; Carlsbad, CA, USA) with L-glutamine and high glucose plus 10% fetal bovine serum (FBS; Invitrogen) and 100 U/mL penicillin/streptomycin (Invitrogen) at 37°C and 5% CO_2. Cells were passaged at confluence by 0.25% trypsin EDTA (Invitrogen). To obtain sufficient cell number and fibroblast purity, cells were cultured for 12 days. Freshly prepared 5-aza-2dC (Sigma, St. Louis, MO, USA) stock solution was diluted in growth medium to concentrations of 1–10 μM. A new preparation of the drug in growth medium was added after 24 h for experiments with a 48-h drug exposure.

Cell Sorting by Flow Cytometry

MoFloHTS™ (DakoCytomation, Glostrup, Denmark) was used for fluorescence-activated cell sorting (FACS) by the University of Chicago Immunology Application Core Facility, Chicago, IL. Cells were sorted in growth medium supplemented with 1 mM EDTA to prevent cell clumping. BD FACScan™ (BD Biosciences,

San Jose, CA, USA) was used for cytometry with CellQuest™ software (BD Biosciences) for analysis.

Methylation-Sensitive Restriction Enzyme Digestion and Quantitative PCR

Genomic DNA was isolated from cells by standard phenol/chloroform extraction. Five micrograms genomic DNA was digested with 4-fold excess *Hpa*II methylation-specific restriction enzyme and *Msc*I restriction enzyme (New England Biolabs, Ipswich, MA, USA) for 3 h. *Msc*I cuts outside of the amplified region and was necessary for improved PCR efficiency. Amplification primers for intracisternal A particle (IAP; GenBank® accession no. M18251) were IAPQF, 5′-CCCAATGAGATGGTTAAATGGAA-3′, and IAPQR, 5′-CTGGATCTTATCAAGATAGGATCCG-3′. The PCR efficiency as determined by standard curve analysis for the product with iQ™ SYBR® Green supermix (Bio-Rad Laboratories, Hercules, CA, USA) was 95%. Quantitative PCR was performed in triplicate using the iCycler® iQ system (Bio-Rad Laboratories) with amplification of *Msc*I-digested genomic DNA used to normalize the sample concentrations.

RNA Preparation and Quantitative Reverse Transcription PCR

Total RNA was isolated from cells using RNeasy® mini columns (Qiagen, Valencia, CA, USA) and treated with RNase-free DNase I (Invitrogen). RNA integrity and purity were confirmed using a spectrophotometer and Model 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). For quantitative reverse transcription PCR (RT-PCR), cDNA was generated using SuperScript™ First-Strand cDNA synthesis kit (Invitrogen). iQ SYBR Green supermix was used to quantify transcript levels with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) amplification to normalize the samples. A melt curve at the end of the amplification was performed to check for product specificity. Samples were run in triplicate using the iCycler iQ system. Primers for RT-PCR were designed with Beacon Designer 2.0 software (Biosoft International, Palo Alto, CA, USA).

Gene Expression Analysis

Probes were prepared using standard Affymetrix protocols. Briefly, double-stranded cDNA was generated using 1.2 µg total RNA with the GeneChip® One-Cycle cDNA Synthesis kit (Affymetrix, Santa Clara, CA, USA). Biotin-labeled cRNA was synthesized with the GeneChip IVT Labeling kit (Affymetrix). Both cDNA and cRNA cleanup and cRNA fragmentation were performed with GeneChip Sample Cleanup Module (Affymetrix). Fragmented cRNA samples were hybridized to GeneChip Mouse Genome 430 2.0 Array (Affymetrix). Sample hybridization, labeling, and scanning were all performed by the Protein and Nucleic Acid (PAN) facility at Stanford University, Stanford, CA.

Statistical Analysis

Data analysis was performed with GeneSpring™ 6.1 software (Silicon Genetics, Redwood City, CA, USA). Raw data were normalized globally to median 50% chip data to minimize process variation between arrays. Expression level normalization was set to the untreated sample to gauge expression changes relative to the control or set to median. Cross gene error model was used to allow for additional statistical analysis. Data points with no present call in any sample were filtered out as unreliable. One sample Student's *t*-test was used when data was normalized to the control untreated GFP^- sample with a *P* value of 0.05. Analysis of variance (ANOVA) used parametric testing and a *P* value of 0.1 to analyze differential expression between treated GFP^+ and treated GFP^- data with data normalized to median.

RESULTS AND DISCUSSION

To determine the optimal drug conditions for demethylation and to enrich for demethylated cells, a method was devised that used the reactivation of an X-inactivated transgene as a proxy for demethylation. The major steps of the method are depicted in Figure 1A. Briefly, primary cells were obtained from female heterozygous mice harboring an X-linked GFP transgene driven by a ubiquitous promoter (13). The transgene is subject to silencing through X-inactivation (15). These cells were sorted by flow cytometry to isolate cells whose

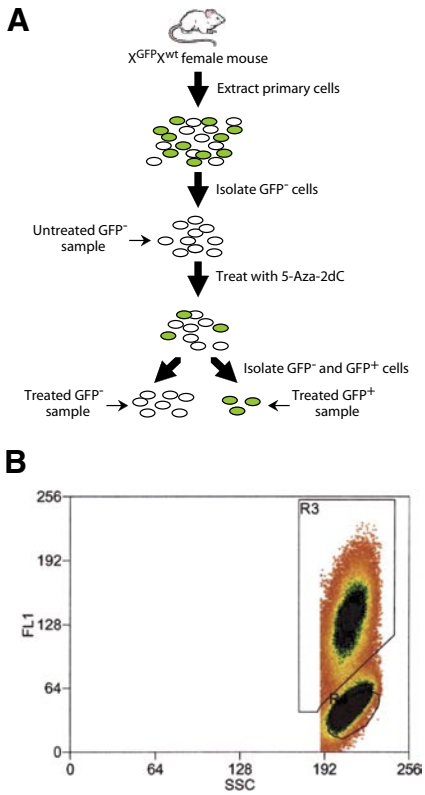

Figure 1. Procedure for obtaining the three experimental cell populations. (A) Depiction of the procedure. In female mice heterozygous for the X-linked green fluorescent protein (GFP) transgene, random X inactivation results in about half of the cells expressing GFP and the other half not expressing it. Primary dermal fibroblasts were extracted from these female mice. They were sorted by flow cytometry to obtain the GFP^- population (the untreated GFP^- cells). These GFP^- cells were then treated with the demethylating drug 5-aza-2′-deoxycytidine (5-aza-2dC). Flow cytometry was used again to separate cells that had undergone the reactivation of the X-linked GFP transgene (treated GFP^+ cells) from those that did not reactivate GFP (treated GFP^- cells). The three cell populations (untreated GFP^-, treated GFP^+, and treated GFP^-) were subjected to microarray expression analysis. (B) A representative flow cytometry plot from the first sort showing two distinct populations of cells based on green fluorescence. The population on the top (R3) is GFP^+, and population on the bottom (R2) is GFP^-. The two populations are roughly equal in number. FL1, fluorescence; SSC, side-scatter.

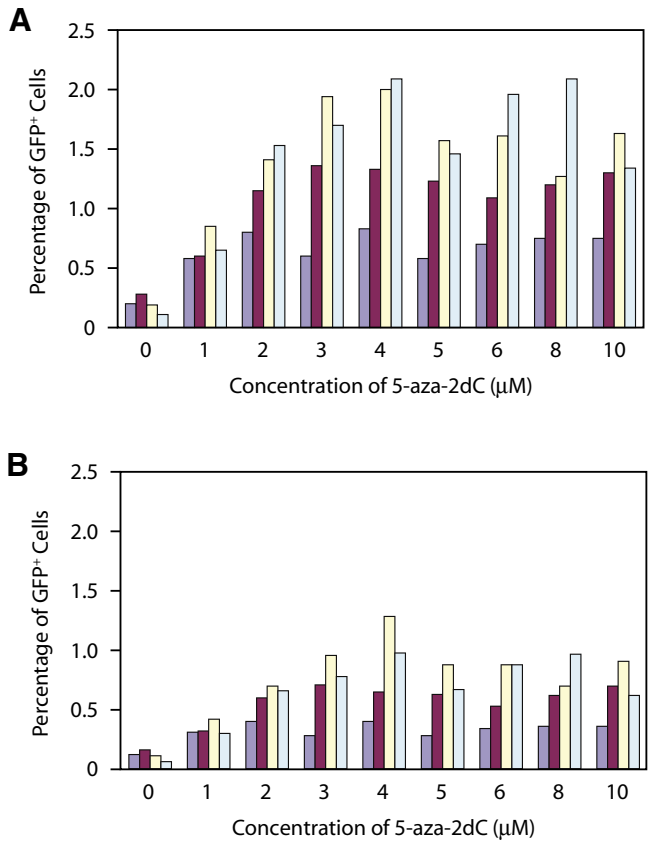

Figure 2. Identifying optimal conditions of drug treatment of primary dermal fibroblasts. (A) The percentages of cells that underwent reactivation of the X-linked green fluorescent protein (GFP) transgene among live cells as determined by flow cytometry. (B) The percentages of cells that underwent reactivation of the X-linked GFP transgene among all cells. Under each 5-aza-2′-deoxycytidine (5-aza-2dC) concentration, four different exposure/recovery time courses were tested, including (from left to right): 24-h exposure followed by 24-h recovery; 24-h exposure followed by 48-h recovery; 48-h exposure followed by 24-h recovery; or 48-h exposure followed 48-h recovery.

transgene resides on the inactive X (i.e., cells that do not express GFP due to X inactivation—these cells are referred to as the untreated GFP$^-$ sample) (Figure 1B). These cells were then treated with 5-aza-2dC, and flow cytometry was again used to separate cells that underwent reactivation of the GFP transgene (the treated GFP$^+$ sample) from cells that did not undergo GFP reactivation (the treated GFP$^-$ sample). Sort stringency was high in the second sort for separating the GFP$^+$ cells to prevent contamination from GFP$^-$ cells. Approximately 99.5% purity was achieved in the GFP$^+$ population, while the denoted GFP$^-$ population contained all other live cells not within the GFP$^+$ window. The GFP$^+$, GFP$^-$, and untreated samples were used for separate microarray expression analysis. In this way, potential gene expression changes incurred from FACS are eliminated.

Dermal fibroblasts from female pups heterozygous for the transgene were cultured for 12 days to obtain sufficient cell number and purity for experimentation. Approximately 50% of the cells did not express GFP and were sorted from GFP$^+$ cells and dead cells or debris. In the resulting GFP$^-$ cells, contamination from GFP$^+$ cells was no greater than 0.4% based on a second sort.

The GFP$^-$ cells, whose GFP transgene apparently resided on the inactive X chromosome, were used for assessing optimal treatment conditions (including 5-aza-2dC concentration, exposure time to the drug, and recovery time in normal growth medium after exposure) by measuring the percentage of cells that undergo GFP reactivation under each condition. Concentrations of 5-aza-2dC ranged from 1 to 10 μM, and both exposure time and recovery time ran either 24 or 48 h. Under each condition, the percentage of live cells that now expressed GFP was measured by flow cytometry (Figure 2A). Since 5-aza-2dC is toxic, this percentage was multiplied by the percentage of live cells determined by forward-versus-side scatter values to estimate the percentage of GFP$^+$ cells among all sorted cells (Figure 2B). Viability ranged from approximately 85% to 60% at higher 5-aza-2dC concentrations. Dermal fibroblasts in log growth phase exposed to 4 μM 5-aza-2dC for 48 h followed by 24 h recovery produced the highest number of live GFP$^+$ cells (Figure 2B). Experimentation with several other cells types, such as osteoblasts, showed that different cell types have different conditions that are optimal for GFP reactivation. While others studying the effect of 5-aza-2dC treatment have used a range of drug concentrations and exposure/recovery times (16–18), this is, to our knowledge, the first systematic attempt to quantitate the optimal conditions of 5-aza-2dC treatment in any cell type.

Once we determined the optimal drug condition (i.e., 4 μM 5-aza-2dC exposure for 48 h followed by 24 h recovery), a large amount of untreated GFP$^-$ cells in log growth phase were subjected to this condition. These cells were sorted to separate the cells that had undergone GFP reactivation from those cells that did not. The resulting GFP$^+$ cells showed a purity of 98% as judged from a second sort. The untreated GFP$^+$ cells were also subjected to flow cytometry to separate live cells from dead cells and debris, and to make them a more comparable control to the drug-treated cells.

The three populations of cells (untreated, GFP$^-$, and GFP$^+$) were assayed for methylation of a site within the highly repetitive IAP. IAP element sequences are dispersed throughout the genome, are known to be heavily methylated, and are thus commonly used to assess global methylation levels (19–21). We used methylation-sensitive *Hpa*II restriction enzyme digestion to assay methylation levels at the cut site. Primers flanking the

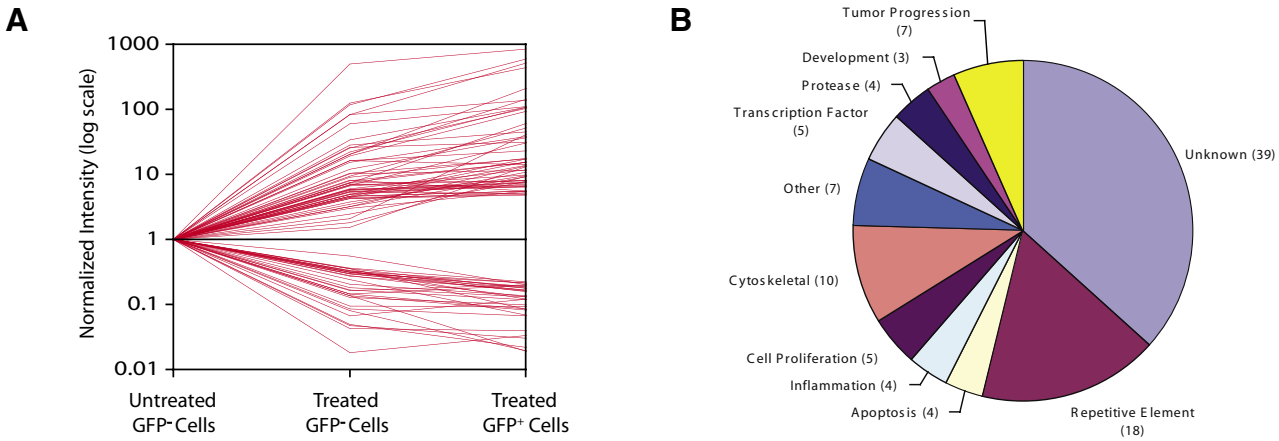

Figure 3. Genes in primary dermal fibroblasts showing significant up- or down-regulation following 5-aza-2′-deoxycytidine (5-aza-2dC) treatment as determined by microarrary analysis. (A) Graph of expression changes. Three cell populations were compared, including untreated green fluorescent protein (GFP)$^-$ cells, treated GFP$^+$ cells, and treated GFP$^-$ cells. Expression levels of all three populations were normalized to that of the untreated GFP$^-$ cells and plotted on a logarithmic scale. (B) Functional categorization of genes showing significant differential expression.

*Hpa*II site were used to amplify the region. Unmethylated sites were cut by *Hpa*II, and no PCR amplification could occur. At methylated sites, however, *Hpa*II digestion was blocked, and PCR amplification could occur. Differences in amplification were measured using quantitative PCR. The quantitative PCR results indicated a 34.2% decrease in methylation in the GFP$^-$ population compared with the control and a 51.0% decrease in methylation in the GFP$^+$ population. These results show a greater quantified demethylation throughout the genome through the sorting technique as compared with nonsorting following 5-aza-2dC treatment.

Two biological replicates, each with samples of untreated GFP$^-$ cells, treated GFP$^-$ cells, and treated GFP$^+$ cells, were used for microarray expression analysis. The microarray data for this experiment have been deposited in Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) and are accessible through GEO Series accession no. GSE4718. The quality of the microarray data were evaluated by visual examination and inspection of the normalization plot. After filtering out unreliable data points, ANOVA was used to identify differentially expressed genes between the treated GFP$^+$ sample and the control untreated GFP$^-$ sample, as well as between treated GFP$^-$ sample and the control untreated GFP$^-$ sample, using a cutoff P value of 0.05. Fifty-five genes were identified as differentially expressed between the treated GFP$^-$ sample and the control untreated sample. An additional 45 genes were identified that were not found in the list of genes differentially expressed in the treated GFP$^-$ sample, which is an 80% increase in the reported gene list. Furthermore, all but three genes differentially expressed between the treated GFP$^+$ sample and the control untreated GFP$^-$ sample had greater fold changes than the corresponding genes in the comparison between treated GFP$^-$ sample and the control untreated GFP$^-$ sample, whether it is up- or down-regulated (Figure 3A). Thus, differential gene expression resulting from 5-aza-2dC treatment is more dramatic in cells with reactivation of the X-linked GFP transgene than those cells that do not reactivate the transgene ($P < 0.0001$ by Fisher's exact test). Using the sorting technique increased the differentially expressed gene list by nearly 2-fold. Of the genes detected, there is also more confidence in the change of gene expression due to the generally greater fold difference in expression levels. Our technique thus clearly enhances results as compared with traditional 5-aza-2dC experiments that lack the enrichment step.

Genes were categorized by function through literature searches. Approximately half of the differentially expressed genes between the treated GFP$^+$ sample and the control untreated GFP$^-$ sample had unknown function or contained repetitive elements normally methylated (Figure 3B). There were also transcription factors and developmental regulators that were activated by demethylation along with tumor-related genes. A full list of differentially expressed genes is available in Supplementary Table S1 available online at www.BioTechniques.com. To confirm the microarray results, 14 genes were analyzed by quantitative RT-PCR (Supplementary Table S1). Strong correlation was seen between the microarray results and the quantitative RT-PCR results. Of the 45 genes differentially expressed in the GFP$^+$ sample only, 40 have a single, known chromosomal location. When the chromosomal distribution of the 40 genes was examined, there is a significant bias toward X-linked genes with 7 of the 40 on the X chromosome (χ^2 test, $P < 0.0002$). The results indicate an increased sensitivity for X-linked gene activation through this technique. Comparing the microarray results of all 100 genes identified in this study to the microarray results of *Dnmt1* deletion in fibroblasts by Jackson-Grusby et al., many genes overlap (10). Both have up-regulation of repetitive element containing genes, transcription factors, developmental regulators, and other genes. However, in the genes noted as differentially expressed between the control and GFP$^+$ population only, just one of the 45 genes was identified as differentially expressed following *Dnmt1* silencing. This shows the GFP activation technique detailed here

identifies additional genes with alterations in gene expression following demethylation.

For other experimental techniques such as conditional knockout or RNAi-induced knockdown of *Dnmt1*, the method can also be applied. In RNAi knockdown studies, researchers have noted that following transfection, a significant percentage of clones lacked any detectable change in methyltransferase activity such that DNA methylation is not affected in these cells (9). With the technique detailed in this report, results from these RNAi studies could be enhanced by eliminating cells that do not reactivate the GFP transgene following RNAi treatment. Conditional knockout of *Dnmt1* requires time-consuming antibiotic selection of resistant clones harboring desirable transgenes, which could significantly complicate the final results (10). With the GFP activation technique, researchers can monitor in real-time the cellular response to DNA demethylation and can subsequently sort out only cells showing GFP reactivation, perhaps even without the need of selection for resistant clones. Indeed, the GFP reactivation technique can probably be applied, with some modification, to all known methods of DNA demethylation to enhance results. The X-linked GFP mice are phenotypically and genotypically normal and can be bred into other established lines. In the transgenic line, GFP expression is ubiquitous so all cell types that are not male-specific can be examined with our technique. This method has been demonstrated to not only give enhanced results in the study of DNA demethylation, but also allow researchers real-time estimation of cellular response to drug treatment to better determine the optimal treatment regime. Finally, we argue that the basic strategy can be expanded to study the biological effects of other chromatin-modifying drugs such as the histone deacetylase inhibitor, trichostatin A (TSA).

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Fischle, W., Y. Wang, and C.D. Allis.** 2003. Histone and chromatin cross-talk. Curr. Opin. Cell Biol. *15*:172-183.
2. **Mello, C.C. and D. Conte, Jr.** 2004. Revealing the world of RNA interference. Nature *431*:338-342.
3. **Jaenisch, R. and A. Bird.** 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33(Suppl)*:245-254.
4. **Li, E., C. Beard, and R. Jaenisch.** 1993. Role for DNA methylation in genomic imprinting. Nature *366*:362-365.
5. **Schoenherr, C.J., J.M. Levorse, and S.M. Tilghman.** 2003. CTCF maintains differential methylation at the Igf2/H19 locus. Nat. Genet. *33*:66-69.
6. **Jahner, D., H. Stuhlmann, C.L. Stewart, K. Harbers, J. Lohler, I. Simon, and R. Jaenisch.** 1982. De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature *298*:623-628.
7. **Feinberg, A.P. and B. Vogelstein.** 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*:89-92.
8. **Ehrlich, M.** 2003. Expression of various genes is controlled by DNA methylation during mammalian development. J. Cell. Biochem. *88*:899-910.
9. **Ting, A.H., K.W. Jair, H. Suzuki, R.W. Yen, S.B. Baylin, and K.E. Schuebel.** 2004. CpG island hypermethylation is maintained in human colorectal cancer cells after RNAi-mediated depletion of DNMT1. Nat. Genet. *36*:582-584.
10. **Jackson-Grusby, L., C. Beard, R. Possemato, M. Tudor, D. Fambrough, G. Csankovszki, J. Dausman, P. Lee, et al.** 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. *27*:31-39.
11. **Christman, J.K.** 2002. 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*:5483-5495.
12. **Juttermann, R., E. Li, and R. Jaenisch.** 1994. Toxicity of 5-aza-2′-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. Proc. Natl. Acad. Sci. USA *91*:11797-11801.
13. **Hadjantonakis, A.K., M. Gertsenstein, M. Ikawa, M. Okabe, and A. Nagy.** 1998. Non-invasive sexing of preimplantation stage mammalian embryos. Nat. Genet. *19*:220-222.
14. **Goldring, K., G.E. Jones, and D.J. Watt.** 2000. A factor implicated in the myogenic conversion of nonmuscle cells derived from the mouse dermis. Cell Transplant. *9*:519-529.
15. **Eggan, K., H. Akutsu, K. Hochedlinger, W. Rideout 3rd, R. Yanagimachi, and R. Jaenisch.** 2000. X-Chromosome inactivation in cloned mouse embryos. Science *290*:1578-1581.
16. **Ng, Y., I. Tan, L. Lim, and T. Leung.** 2004. Expression of the human myotonic dystrophy kinase-related Cdc42-binding kinase gamma is regulated by promoter DNA methylation and Sp1 binding. J. Biol. Chem. *279*:34156-34164.
17. **Kulaeva, O.I., S. Draghici, L. Tang, J.M. Kraniak, S.J. Land, and M.A. Tainsky.** 2003. Epigenetic silencing of multiple interferon pathway genes after cellular immortalization. Oncogene *22*:4118-4127.
18. **Stimson, K.M. and P.M. Vertino.** 2002. Methylation-mediated silencing of TMS1/ASC is accompanied by histone hypoacetylation and CpG island-localized changes in chromatin architecture. J. Biol. Chem. *277*:4951-4958.
19. **Walsh, C.P., J.R. Chaillet, and T.H. Bestor.** 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. *20*:116-117.
20. **Mays-Hoopes, L.L., A. Brown, and R.C. Huang.** 1983. Methylation and rearrangement of mouse intracisternal a particle genes in development, aging, and myeloma. Mol. Cell. Biol. *3*:1371-1380.
21. **Csankovszki, G., A. Nagy, and R. Jaenisch.** 2001. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. *153*:773-784.

Received 10 January 2006; accepted 8 July 2006.

Address correspondence to Bruce T. Lahn, Howard Hughes Medical Institute, Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. e-mail: blahn@bsd.uchicago.edu