

Human embryonic stem cells have a unique epigenetic signature

Marina Bibikova,¹ Eugene Chudin,¹ Bonnie Wu,¹ Lixin Zhou,¹ Eliza Wickham Garcia,¹ Ying Liu,^{2,12} Soojung Shin,^{2,12} Todd W. Plaia,³ Jonathan M. Auerbach,³ Dan E. Arking,⁴ Rodolfo Gonzalez,⁵ Jeremy Crook,⁶ Bruce Davidson,⁶ Thomas C. Schulz,⁷ Allan Robins,⁷ Aparna Khanna,⁸ Peter Sartipy,⁹ Johan Hyllner,⁹ Padmavathy Vanguri,¹⁰ Smita Savant-Bhonsale,¹⁰ Alan K. Smith,¹¹ Aravinda Chakravarti,⁴ Anirban Maitra,⁴ Mahendra Rao,^{2,12} David L. Barker,¹ Jeanne F. Loring,⁵ and Jian-Bing Fan^{1,13}

¹llumina, Inc., San Diego, California 92121, USA; ²Laboratory of Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, Maryland 21224, USA; ³Stem Cell Center, American Type Culture Collection, Manassas, Virginia 20108, USA; ⁴McKusick—Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA; ⁵Stem Cell Center, Burnham Institute for Medical Research, La Jolla, California 92037, USA; ⁶ES Cell International, The Gemini 117610, Singapore; ⁷BresaGen, Inc., Athens, Georgia 30605, USA; ⁸Reliance Life Sciences Pvt. Ltd., Mumbai 400 701, India; ⁹Cellartis AB, 413 46 Goteborg, Sweden; ¹⁰Theradigm, Inc., Baltimore, Maryland 21227, USA; ¹¹Cognate Therapeutics, Inc., Baltimore, Maryland 21227, USA

Human embryonic stem (hES) cells originate during an embryonic period of active epigenetic remodeling. DNA methylation patterns are likely to be critical for their self-renewal and pluripotency. We compared the DNA methylation status of 1536 CpG sites (from 371 genes) in 14 independently isolated hES cell lines with five other cell types: 24 cancer cell lines, four adult stem cell populations, four lymphoblastoid cell lines, five normal human tissues, and an embryonal carcinoma cell line. We found that the DNA methylation profile clearly distinguished the hES cells from all of the other cell types. A subset of 49 CpG sites from 40 genes contributed most to the differences among cell types. Another set of 25 sites from 23 genes distinguished hES cells from normal differentiated cells and can be used as biomarkers to monitor differentiation. Our results indicate that hES cells have a unique epigenetic signature that may contribute to their developmental potential.

[Supplemental material is available online at www.genome.org.]

Human embryonic stem (hES) cells are unique in their abilities to maintain pluripotency and a normal diploid karyotype over long periods in culture. These properties make hES cells leading candidates for use in cell therapy and for studies of early human development. Human ES cells have been investigated by multiple techniques, including gene expression profiling, mitochondrial sequencing, immunocytochemistry, genotyping, and functional assays (Andrews et al. 2005; Loring and Rao 2006). These complementary approaches have been applied to a representative subset of the more than 200 hES lines now in existence, helping to construct a comprehensive molecular profile of an archetypal hES cell line. While recent studies have reported gene expression changes and variations in the DNA sequence of hES cells during long term culture, very little is known about epigenetic regulation in hES cells. Methylation and demethylation of regulatory sequences in the genome are known to have profound effects on cellular behavior and fate (Allegrucci et al. 2005). Massive demethylation is believed to underlie the global genomic reprogramming of gamete DNA that occurs after fertilization (Morgan

et al. 2005). Maintenance in gametes of methylation patterns in imprinted genes is responsible for parental-specific inheritance of human disorders such as Prader-Willi/Angelman syndrome (Kantor et al. 2004).

Human ES cell lines are derived from blastocyst-stage embryos that are excess after in vitro fertilization (IVF) procedures (Thomson et al. 1998; Keller 2005). The blastocyst stage, which occurs at about five days after fertilization in humans, is characterized by high levels of epigenetic activity, including DNA methylation, X chromosome inactivation, and dynamic chromatin remodeling. Recent reports from IVF clinics have suggested an unexpectedly high occurrence of imprinting and other epigenetic abnormalities in early-stage human embryos (Jacob and Moley 2005), raising the possibility that cultured embryonic stem cells may vary considerably in their epigenetic status and that these differences may underlie functional differences in differentiation ability. But an equally strong argument can be made for the existence of common epigenetic characteristics in different hES cell lines, perhaps as a property that allowed them to be expanded in culture as cell lines in the first place. In either case, it is not known whether the epigenetic profile of ES cells is stable during long-term culture, nor how it may change as the cells differentiate along different developmental pathways. Efforts have been initiated to assess the epigenetic status of a small number of specific genes in hES cell lines (Allegrucci et al. 2005; Keller

¹²Present address: Invitrogen, Inc., Carlsbad, CA 92008, USA.

¹³Corresponding author.

E-mail jfan@illumina.com; fax (858) 202-4680.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5319906>. Freely available online through the *Genome Research* Open Access option.

2005; Maitra et al. 2005), but so far there has been no global assessment of their overall methylation status.

Results

Methylation profiles of human ES cell lines

We applied a comprehensive DNA methylation profiling approach to assess the epigenetic state of 36 human ES cell cultures, derived from 14 independently isolated lines of hES cells (Thomson et al. 1998; Reubinoff et al. 2000; Amit and Itskovitz-Eldor 2002; Brimble et al. 2004; Cowan et al. 2004; Heins et al. 2004; Zeng et al. 2004; Maitra et al. 2005) at various times in culture, and the embryonal carcinoma cell line NTERA2 (Andrews et al. 1984) and asked whether these pluripotent cells differed from other types of cells, including other types of stem cells. We used an array-based method (Bibikova et al. 2006) to analyze the methylation status of 1536 CpG sites selected from the 5'-regulatory regions of 371 genes. These genes were chosen on the basis of their importance to cellular behavior and differentiation, and included known imprinted genes and genes previously reported to be differentially methylated, as well as tumor suppressor genes, oncogenes, and genes coding for factors involved in cell cycle checkpoint. We also included genes that are regulated by various signaling pathways and/or are responsible for altered cell growth, differentiation and apoptosis, and genes involved in DNA damage repair and oxidative metabolism. The methylation state of all 1536 specific CpG sites was quantitatively measured in a single reaction by multiplexed genotyping of bisulfite-treated genomic DNA (Bibikova et al. 2006).

We measured methylation in the 14 independently derived hES lines that originated in several different laboratories, an aneuploid derivative of the diploid line BG01 (BG01V) and a dip-

loid revertant of the aneuploid line SA02 (SA02.5), and NTERA2 (Table 1). The cells were all undifferentiated populations. Each cell line showed a slightly different methylation profile (epigenotype), but unsupervised hierarchical clustering based on the methylation status at each of the 1536 sites revealed that all hES cells (and NTERA2) could be easily separated from all differentiated cells, somatic stem cells, and cancer cell lines (Fig. 1A; Supplemental Table 1).

Within the ES cell type, the cell lines separated into subclusters according to their sex, since the analyzed loci include multiple X chromosome-linked genes. As shown in Figure 1A, all the male lines (BG01, BG02, WA01, TE06, SA01, and HUES7) were clustered together, while the female lines were clustered into three subgroups: group 1 (BG03, TE04, WA07, and WA09), group 2 (SA02/SA02.5), and group 3 (Relicell hES1, ES02, and ES03). The only exception was NTERA2, an aneuploid (hypotriploid) embryonal carcinoma cell line, which was most similar to the female group 3, even though it is a male cell line by origin. Furthermore, cluster analysis of methylation profiles of six X-linked housekeeping genes (*ELK1*, *EFNB1*, *FMR1*, *G6PD*, *GLA*, and *GPC3*) segregated all the hES cells, including various in vitro passages, NTERA2, somatic stem cells, differentiated lymphoblastoid cells, and human tissues accurately by their sex (Fig. 1B).

NTERA2 has many of the qualities of hES cells, including pluripotency. Although methylation patterns of NTERA2 cells were in general similar to those of hES cells, certain CpG sites were distinctly different. For example, four imprinted genes (*NDN*, *SGCE*, *PEG3*, and *SNURF*) were hypermethylated in NTERA2 but unmethylated in all hES cells, while other genes such as *RUNX3*, *POMC*, *PTPN6*, and *KCNK4* are hypomethylated in NTERA2 and hypermethylated in the hES cells (data not shown).

Using all 1536 CpG sites, we determined the overall meth-

Table 1. Human embryonic and somatic stem cells

Embryonic stem cell lines	Passage ^a	Gender (extra chromosomes)	Ethnicity	Reference
BG01	P11, P41	Male	Unknown	Brimble et al. (2004)
BG01V	P25	Male (+12, +17, +X)	Unknown	Zeng et al. (2004)
BG02	P13, P14, P40	Male	Unknown	Brimble et al. (2004)
BG03	P14, P47	Female	Unknown	Brimble et al. (2004)
ES02 (HES2)	P55, P146, P147	Female	Chinese	Reubinoff et al. (2000)
ES03 (HES3)	P59, P92, P93	Female	Chinese	Reubinoff et al. (2000)
HUES7	P14	Male ^c	Unknown	Cowan et al. (2004)
Relicell hES1	P30	Female	Indian	Mandal et al. (2006)
SA01	P14, P25, P32, P60	Male	Unknown	Heins et al. (2004)
SA02	P18, P29, P30, P59	Female (+13)	Unknown	Heins et al. (2004)
SA02.5 ^b	P12, P49, P50, P77	Female diploid	Unknown	Heins et al. (2006)
TE04 (14)	P29	Female	Unknown	Amit and Itskovitz-Eldor (2002)
TE06 (16)	P44	Male	Unknown	Amit and Itskovitz-Eldor (2002)
WA01 (H1)	P80	Male	Unknown	Thomson et al. (1998)
WA07 (H7)	P41, P63	Female	Unknown	Thomson et al. (1998)
WA09 (H9)	P33, P68, P78	Female	Unknown	Thomson et al. (1998)
Embryonal carcinoma				
NTERA-2		Male (aneuploid)	Unknown	Andrews et al. (1984)
Somatic stem cells				
THD-hWB-015 (NSC)		Male ^c	Unknown	Unknown
THD-hFB-017 (NSC)		Female ^c	Unknown	Unknown
BMSC-21		Male	African american	Unknown
BMSC-25		Male	African american	Unknown

^a"Passages" are relative to the beginning of the indicated line, and lines derived from other lines (BG01V and SA02.5) will have been cultured in vitro longer than the passage number suggests.

^bSA02.5 itself is a clonal derivation from SA02 after 155 passages, so its 12 passage is really 155 + 12 = 167 passages from initiation.

^cGender determined in this study.

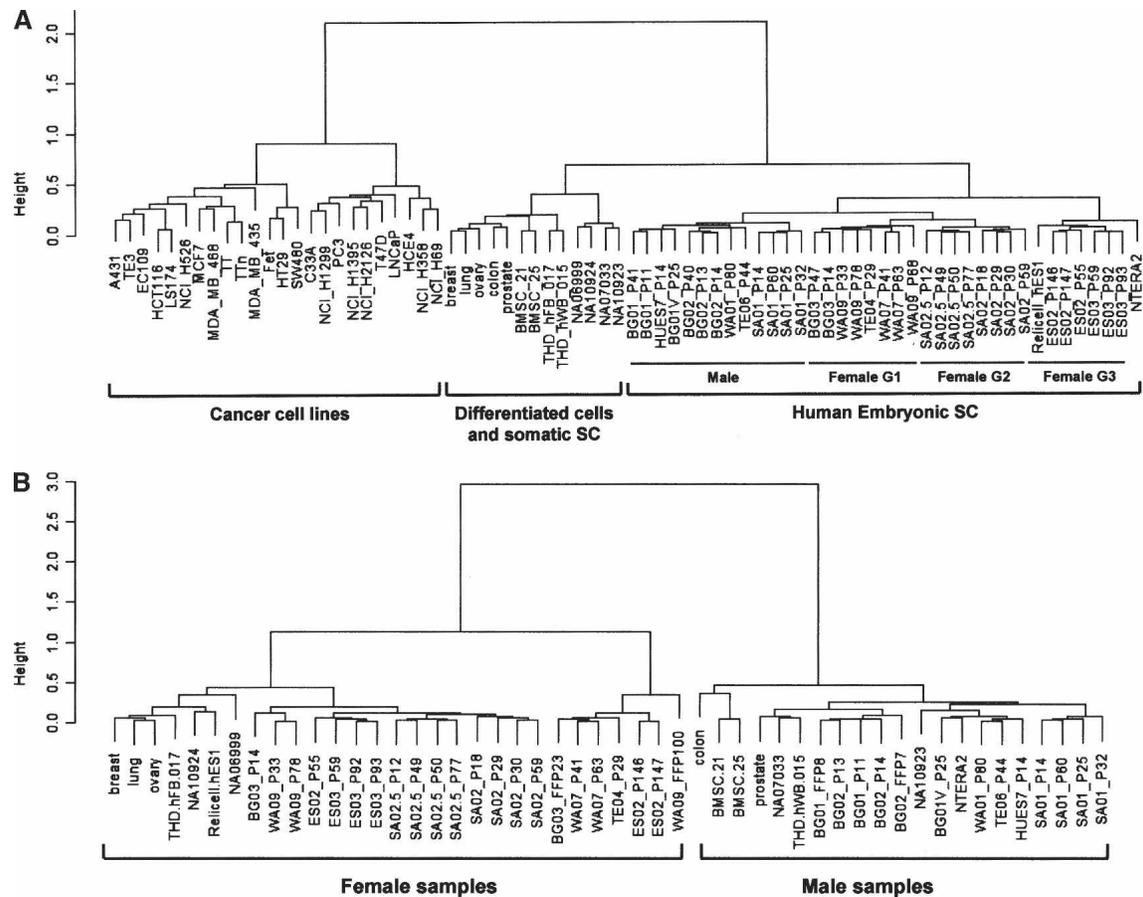


Figure 1. Cluster analysis of human embryonic stem cells, differentiated cells, somatic stem cells, and cancer cell lines based on correlation of methylation profiles of 1536 CpG sites (A) and 36 CpG sites from six X-linked housekeeping genes: *EFNB1*, *ELK1*, *FMR1*, *G6PD*, *GLA*, and *GPC3* (B).

ylation levels for three sample groups: the hES cell lines, cancer cell lines, and a group that comprised the differentiated lymphoblastoid cells and somatic stem cells. We found the average methylation level of analyzed genes in cancer cell lines (38%) to be greater than in hES cell lines (35%) which in turn was greater than in differentiated cells and somatic stem cells (32%) (Fig. 2A) at a high confidence level ($P = 4.78 \times 10^{-5}$). Since the 371 genes that we examined were not selected randomly (for example, many tumor suppressor and other cancer-related genes were included), the higher methylation level we obtained for the cancer cell lines was expected, because promoter hypermethylation has been established as the most frequent mechanism for gene inactivation in cancers (Esteller 2002; Herman and Baylin 2003) (Fig. 2C). However, the intermediate methylation level in hES cell lines is surprising, given that, in general, global hypomethylation has been reported in embryonic stem cells (Zvetkova et al. 2005). We note that Figure 2A should not give the impression that the overall distribution of methylation among the 1536 sites peaks between 30% and 40% methylation. Rather, the distribution is biased toward low methylation (Supplemental Fig. 1).

We further looked at the overall methylation levels of these cell populations according to different gene categories—imprinted genes, tumor suppressor genes, and genes in the MHC region (Supplemental Table 2). While the average methylation level of imprinted genes was quite similar in all sample groups (Fig. 2B), the variation was small in differentiated cells and hES

cells, while cancer cell lines exhibited a very broad distribution of methylation levels in imprinted genes. This observation indicates that patterns of imprinting are established very early in embryo development and, in general, they are tightly controlled during in vitro culturing. It also confirms that loss of imprinting is one of the common epigenetic changes in cancer (Lee 2003; Holm et al. 2005). Our finding of increased methylation levels in tumor suppressor genes in cancer cell lines compared with hES cells and differentiated cells (Fig. 2C) agrees well with the established mechanism of malignant transformation through silencing of tumor suppressor genes by CpG island promoter hypermethylation (Herman and Baylin 2003). In contrast, we found that genes in the MHC region had higher levels of methylation in hES cells and lower levels in differentiated tissues and cancer cell lines (Fig. 2D).

Methylation profiles differ between hES cells and differentiated cells

To discover which CpG sites contribute the most to the epigenetic signature of different cell types, we used a supervised analysis method, “Between Group Analysis” (Culhane et al. 2002). We compared the DNA methylation patterns among four groups of cell types: hES cells, somatic stem cells (two neural stem cell lines and two bone marrow stem cell preparations), differentiated cells, and cancer cell lines. Only the earliest passage of each hES

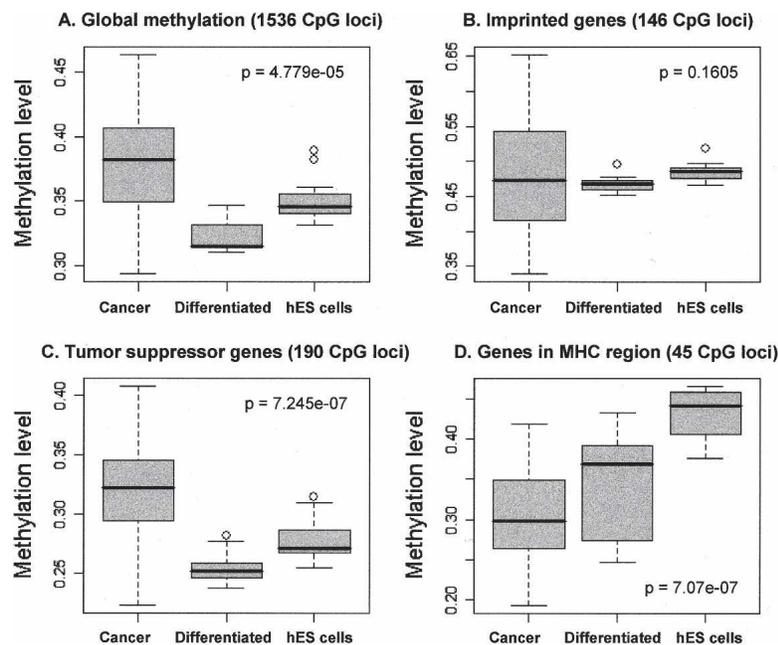


Figure 2. Boxplots of methylation levels in three sample groups—hES cells ($N = 16$), differentiated and somatic stem cells ($N = 13$), and cancer cell lines ($N = 24$). (A) Global methylation levels, calculated using all 1536 CpG sites; (B) methylation levels of 146 CpG sites in 33 imprinted genes; (C) methylation levels of 190 CpG sites in 45 tumor suppressor genes; (D) methylation levels of 45 CpG sites in 13 genes from the MHC region. The black bar represents the median methylation level for each cell type, calculated from the means of each individual cell line. The gray box defines quartiles (25% and 75%, respectively). The error bars are 1st and 99th percentiles of the distribution. Dots represent the outliers. P -values were computed using the Kruskal-Wallis test, which indicates the likelihood that all medians are the same.

cell line was used for this analysis and NTERA2 was excluded. This analysis produced a list of 49 CpG sites from 40 genes that contribute most to the separation of the four groups (see Methods). We then clustered all samples based on their relative methylation levels at these 49 CpG sites (Fig. 3). Despite the multiple locations of origin, different sample preparation, and karyotypes, all pluripotent cell lines (including early and later hES passages and NTERA2) were correctly aggregated into a single cluster, which is well separated from the two other major clusters, which comprise cancer cell lines and differentiated cells with somatic stem cell lines. This suggests that all human ES cells share a common epigenetic signature, which is likely to be linked to ES cell-specific properties such as self renewal and pluripotency.

Both unsupervised clustering based on all 1536 sites (Fig. 1A) and the methylation signatures shown in Figure 3 indicated that somatic stem cells had overall methylation profiles that were more similar to differentiated cells than to hES cells. When we use the methylation signature of the 49 CpG sites, somatic stem cells separate from normal tissues and lymphoblastoid cell lines, but this difference is much smaller than between other groups of samples (Fig. 3).

Unique methylation signature of pluripotent cells

The choice of the 49 CpG sites discussed above was heavily influenced by the methylation profiles of the cancer cell lines. To identify methylation markers that could be used to monitor stability of pluripotent, undifferentiated hES cells, we used a t -test to ask which CpG sites had a mean methylation level in hES cells that differed significantly from the mean level in the normal

samples (somatic stem cells and differentiated cells, with cancer cell lines excluded). This analysis yielded 168 CpG sites that had a highly significant difference ($P < 0.001$) in methylation between the hES cell group and the differentiated and somatic stem cell groups (Supplemental Table 3). A cluster diagram and heat map based on the 25 most significant sites derived from 23 genes are shown in Figure 4. Although only the earliest passage of each hES line was used to select the CpG sites, all pluripotent cells (hES cells and NTERA2) form one cluster that is widely separated from all normal differentiated cell samples. Therefore, the 25 CpG sites should be useful as biomarkers diagnostic for undifferentiated, pluripotent cells.

Methylation changes during prolonged passage

Methylation profiling of hES cells can serve to assess epigenetic stability during cell culture. To assess the amount of change during culture, we compared the methylation pattern at different passage numbers for nine hES cell lines. The passage interval of hES cells varies in different laboratories, owing both to planned schedules and to experimental conditions that affect growth rates, but is generally between four and seven days. We observed that different hES cell lines showed different changes with time in culture, and that the degree of overall change in methylation was roughly proportional to the number of passages separating compared preparations. For example, we observed few changes between p41 and p63 (22 additional passages) in the WA07 cell line, but relatively more changes between p33 and p78 (45 additional passages) in WA09 and between p55 and p146 (91 additional passages) in ES02. Some CpG sites lost methylation during time in culture, while some gained methylation (Supplemental Table 4). Among these sites, there are known imprinted genes (*ASCL2*, *PEG10*), nonimprinted autosomal genes such as *COL1A2*, *HLA-DQA1*, and *RASSF1*, a gene in which changed methylation level after prolonged passage was reported previously (Maitra et al. 2005). We cannot determine if the methylation changes are reflecting the changes in each individual cell or are the result of selection (during prolonged passage) of a cell population with different methylation levels. The changes could be a trend or could be simply fluctuations that occur during multiple passages.

Although methylation changes did occur with time in culture, we did not find a specific set of genes in the group studied that changed predictably. Complicating this analysis is the fact that passage number is only a rough measure of time in culture, and that some lines, such as SA02.5, were derived from previous lines (Heins et al. 2006). Overall, however, the methylation changes during time in culture are small compared with differences among cell types, as illustrated by the cluster analysis in Figure 1A, in which the same cell lines of different passages are still clustered together and separate from other cell lines.

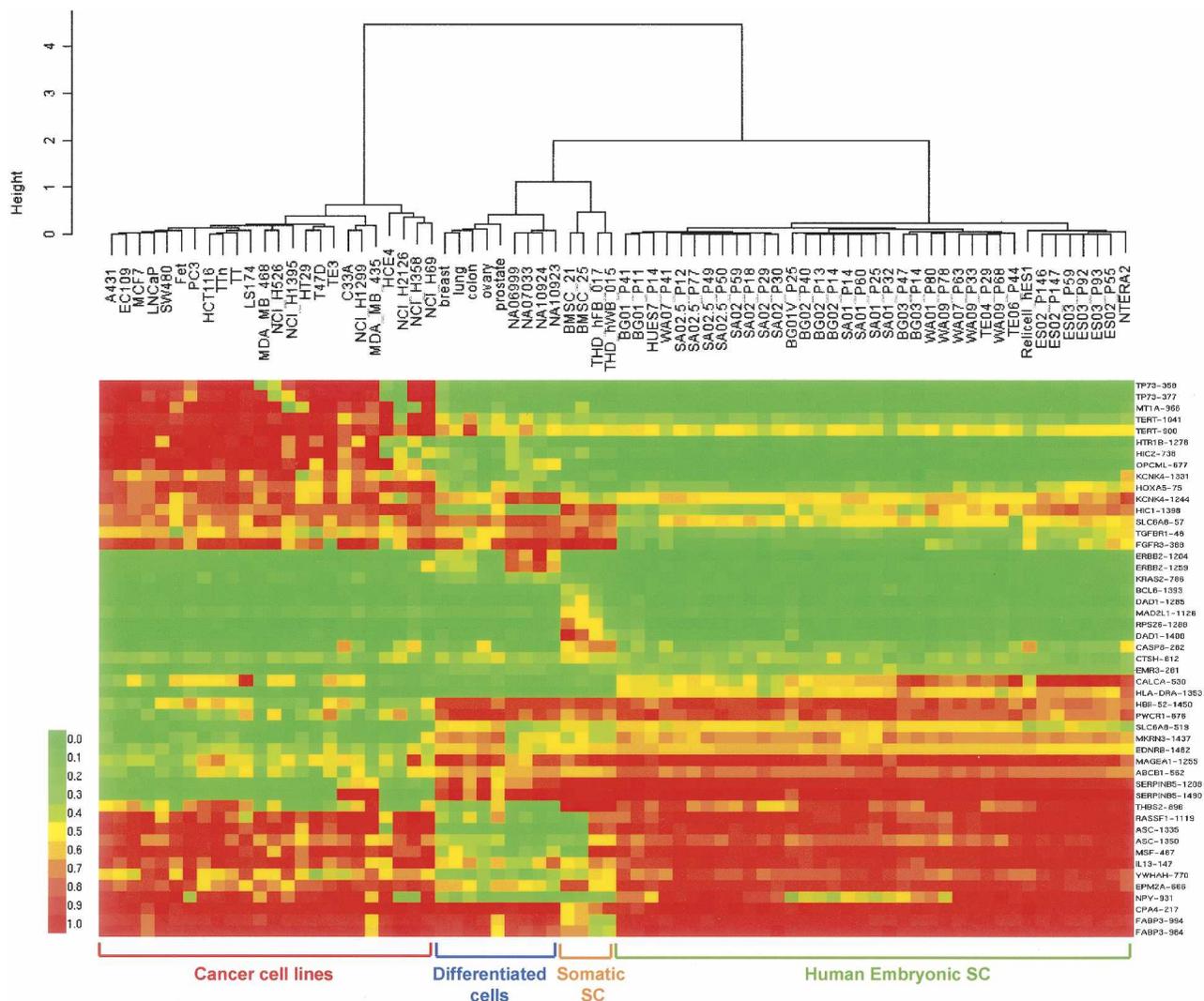


Figure 3. Methylation profiles of human embryonic stem cells, differentiated tissues, somatic stem cells, and cancer cell lines for 49 CpG sites from 40 genes that most strongly distinguish hES cells and the other three cell categories.

Validation of BeadArray methylation data

We selected four CpG sites (ASC_1350, HS3ST2_955, HTR1B_1278, and TP73_377) that showed differential methylation between hES cells and the other cells (Fig. 3) for MSP validation. MSP primers were developed (Supplemental Table 5) and used to measure the methylation levels at the corresponding sites in DNA isolated from 16 cell lines and tissues: normal lung and breast tissues, normal lymphoblastoid cell lines (NA07033 and NA06999), cancer cell lines (NCI-H358, T-47D, NCI-H1395, and MCF-7), hES cell lines (WA07_P41, BG01V_P25, WA07_P63, WA09_P33, BG02_P14, WA09_P78, BG02_P40), and NTERA2.

A good correlation between methylation patterns was obtained between real-time MSP and BeadArray data. The Spearman correlation coefficient between the two methods for each tested CpG site was 0.87 (ASC_1350), 0.82 (HS3ST2_955), 0.86 (HTR1B_1278), and 0.89 (TP73_377), respectively (Supplemental Table 6). This agrees well with the correlation previously obtained with five other sites using different samples (Bibikova et al. 2006). When real-time MSP and array data disagreed, there

was a tendency for array data to indicate a higher level of methylation (Supplemental Table 6). Without a third independent method, we cannot tell which values are more accurate. However, the power of array data is that a very large number of sites is determined simultaneously, allowing highly reproducible global patterns to be discovered.

Discussion

The ability to access the epigenomic information for a large number of genes or the entire genome (Rakyan et al. 2004; Heisler et al. 2005; Murrell et al. 2005; Weber et al. 2005) should greatly facilitate the understanding of the nature of pluripotency in embryonic stem cells. It should also have significance for studies of human epigenetic disorders and assisted reproduction (Allegrucci et al. 2004; Cowan et al. 2005; Jacob and Moley 2005). For example, there are indications that some hES cell lines undergo spontaneous differentiation more readily than others under standardized conditions (Bodnar et al. 2004; Clark et al. 2004), and it

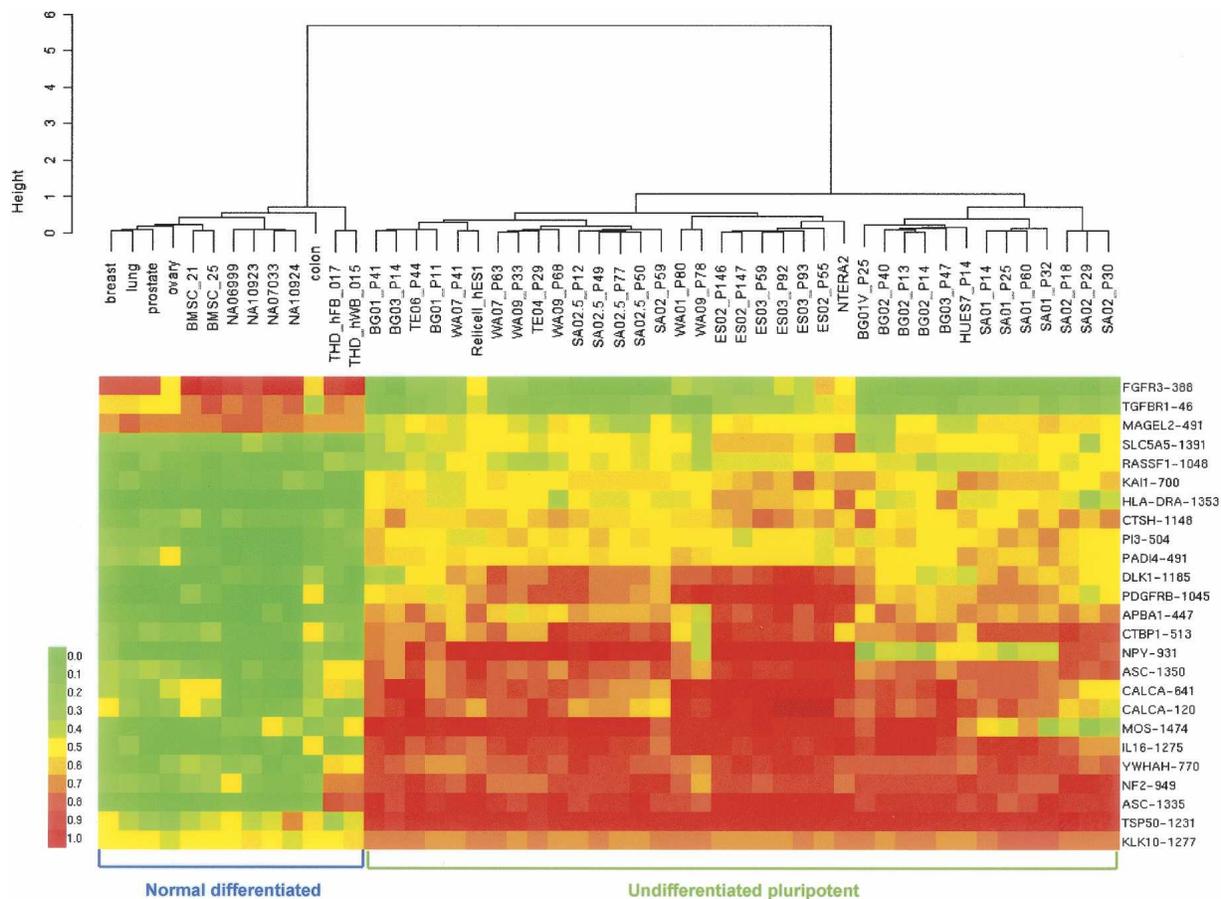


Figure 4. Methylation profiles of human embryonic stem cells, differentiated tissues, and somatic stem cells for 25 CpG sites (from 23 genes) selected for their distinct difference in methylation level between hES cells and normal differentiated tissues and somatic stem cells.

seems likely that genomic differences among diverse hES cell lines will affect their abilities to develop into different kinds of mature cell types. In this study, we took a comprehensive approach to analyze DNA methylation profiles of 371 genes in a diverse group of hES cell lines, somatic stem cell lines, differentiated cells, cancer cell lines, and an embryonal carcinoma cell line. Our study revealed unique epigenotypes (i.e., baseline DNA methylation profiles) for each of these cell types.

Our results may help resolve a concern that, because hES cells share certain characteristics with cancer cells (lack of senescence, regulation of cell cycle, and contact inhibition), they may be partially transformed cells that are prone to form tumors. Our results show that the methylation profile of hES cells was significantly different from that of cancer cells, and that imprinted genes are tightly controlled in hES cells and differentiated cells, but not cancer cells (Fig. 2B). This suggests that some qualities of hES cells are coincidentally, not functionally, similar to cancer cells.

A panel of 49 CpG sites from 40 genes was identified whose differential methylation was sufficient to define distinct epigenetic signatures among hES cells, all differentiated cells, somatic stem cells, and cancer cell lines. Profound methylation differences between hES cells and every other cell type extended well beyond the 49 CpG sites used for the analysis in Figure 3, but these sites were major contributors. These 49 sites will be useful markers for monitoring hES cell cultures for signs that they are

becoming differentiated or acquiring epigenetic similarities to cancer cells. We also selected a list of 25 methylation sites from 23 genes that distinguish undifferentiated, pluripotent cells from normal differentiated cells and somatic stem cells. These 25 CpG sites will be valuable for detecting more subtle changes that are indicative of hES cell differentiation. It is important to note that these 25 sites are useful biomarkers of developmental potential, but they are not the only markers of cellular differentiation. The methylation profiles were different for a much larger number of CpG sites (Supplemental Table 3), which is reflected by the fact that good separation was achieved when we clustered these samples using all 1536 CpG sites (Fig. 1A). Among the genes with elevated methylation levels in hES cells are genes encoding proteins involved in nuclear and extracellular signaling (e.g., *THBS2*, *IL13*, *IL16*, *TNF*, *MSF(SEPT9)*, *PI3*), stress response and apoptosis *ASC(PYCARD)*, *CASP8*, cell cycle control (*CDKN1B*, *RASSF1*), and genes from the HLA locus (*HLA-DQA1*, *HLA-DPA1*). Two growth factor receptor genes, *FGFR3* and *TGFBR1*, had a very low methylation level in hES cells as compared with that in differentiated cells and somatic stem cells. TGF β signaling has been shown to play a crucial role in the maintenance of the undifferentiated state of hES cells (James et al. 2005), while the basic fibroblast growth factor (FGF2) has been shown to promote long-term undifferentiated proliferation of hES cells (Xu et al. 2005). In order to address the key question of the relationship between methylation and expression of genes, we have begun experiments to

examine individual populations of hES cells for both methylation status and expression of the genes in this panel.

We monitored epigenomic stability through multiple passages of nine of the hES cell lines and were able to detect finer resolution methylation changes during extended passages (Supplemental Table 4). The changes, while useful for pinpointing differences in hES cell preparations, do not obscure the fundamental distinctive methylation profile of hES cells; in most cases, the early and late passage preparations still cluster closely together (Figs. 1A, 3), making it feasible to establish a correlation between epigenetic stability measured by methylation profiling and variations in culture conditions. Recently efforts have been directed toward development of new culture systems for the maintenance of hES cells *in vitro* (Ludwig et al. 2006), and such studies potentially represent important technical advances for the field. But without knowing how much particular cell lines differentiate under a given set of growth conditions, or how the properties of a given cell line vary when grown in different laboratories, it is difficult to assess the utility of any new protocol (Andrews et al. 2005; Loring and Rao 2006). A simple test using methylation profiling of signature genes such as those we identified should prove to be useful for comparing stem cell lines derived from different laboratories and for assessing the effects of new culture methodology on stem cell stability.

The methylation profiles also served to distinguish between male and female lines with high reliability. Detailed examination showed that this difference could be attributed to methylation patterns of X-linked genes included in the analysis (Fig. 1B). It is tempting to suggest that the methylation pattern we observe could be used as a surrogate for X inactivation. However, a careful study of X chromosome inactivation in hES cells found evidence of X inactivation in WA09 but not in WA07 lines (Hoffman et al. 2005), which suggested that hES cell lines can differ in this epigenetic characteristic. Our results suggest that these two lines have a similar DNA methylation pattern of X-linked genes (five of six X-linked genes), suggesting that the processes of establishment and maintenance of X inactivation in hES cells may be more complex than has been imagined.

Our results show that DNA methylation pattern distinguishes hES cells from other types of cells examined, which suggests that this epigenetic characteristic is involved in hES cell characteristics such as self-renewal and pluripotency. The genes were selected for the array because of their significance in cancer, not because of their association with stem cells. Yet the methylation pattern at even a small subset of these genes is sufficient to reliably distinguish hES cells from normal adult cells, somatic stem cells, and tumor cell populations. This suggests that global epigenetic changes may lie at the heart of features of hES cells such as pluripotency and self-renewal. Perturbation of the expression of key differentially methylated genes may reveal the functional significance of these epigenetic events and yield clues about how the genes function in stem cell maintenance at a molecular level.

Our results also highlight the importance of methylation in regulating cell differentiation. Compared with hES cells, somatic stem cells had overall methylation profiles that were more similar to differentiated cells, as is illustrated by their coclustering with normal tissues (Figs. 1A, 3, 4). This result may be a consequence of the more restricted developmental repertoire of somatic stem cells compared with embryonic stem cells.

The hES cell preparations used for this analysis differed in ethnic origin, time in culture, culture conditions, and laboratory-

specific methods. Despite this diversity, there is striking similarity in methylation pattern among all hES cell lines, demonstrating that there may be an hES cell-specific epigenetic signature. These observations should be considered in light of the efforts to create patient-specific hES cells by transplantation of somatic nuclei (Holm et al. 2005). If somatic cell nuclear transfer (SCNT) eventually succeeds for humans, it will be important to test the accuracy of epigenetic reprogramming by comparing the methylation profile of the SCNT-derived hES cells with a normal profile of hES cell lines such as the one reported here.

Methods

Stem cell lines for methylation analysis

Twelve hES cell lines listed on the NIH registry were analyzed for this study (Table 1). The cell pellets were obtained from five providers or collaborating laboratories, where the lines had been maintained under each laboratory's standard conditions. The characteristics of these twelve hES cell lines used are available at <http://stemcells.nih.gov/research/registry/eligibilityCriteria.asp>. Multiple passages were available on nine of the lines, as described (Maitra et al. 2005). In addition, two hES cell lines outside the NIH registry (HES1 [Reliance Life Sciences] and HUES7 [HUES Cell Facility]) were also analyzed, as was the teratocarcinoma cell line NTERA2 (Andrews et al. 1984). Two neural stem cells (THD-hWB-015 and THD-hFB-017) were obtained from Theradigm, Inc., and two bone marrow mesenchymal stromal cells (BMSC-21 and BMSC-25) were obtained from Cognate Therapeutics. No additional cultures were performed at Illumina, Inc., nor at Johns Hopkins University for any of the hES cell or adult stem cell lines. DNA was extracted from the cell pellets using the Qiagen DNeasy kit in preparation for analysis on the bead arrays.

DNA samples

Twenty-four human cancer cell lines, four B-lymphocyte non-cancer cell lines and five differentiated tissues were analyzed as a comparison with hES cells and adult stem cell lines (Supplemental Table 1). DNA from normal lung, ovary, breast, colon, and prostate tissues was purchased from Clinomics Biosciences. DNA samples NA06999, NA07033, NA10923, and NA10924 were purchased from the Coriell Institute for Medical Research. Breast cancer cell lines MCF-7, MDA-MB-435, MDA-MB-468, and T-47D, colon cancer cell lines Fet, HT29, HCT116, LS174, SW480, esophageal cancer cell lines EC109, T.T., T.Tn, TE3, HCE4, prostate cancer cell lines PC3 and LNCaP, cervical carcinoma C33A, and epidermal carcinoma A431 were a gift from Prof. Wei Jiang (The Burnham Institute). DNA from these cancer cell lines was extracted using a modified Trizol method according to the manufacturer's recommendations (Invitrogen). DNA from lung cancer cell lines NCI-H69, NCI-H526, NCI-H358, NCI-H1299, NCI-H1395, and NCI-H2126 was purchased from ATCC.

DNA methylation profiling using bead arrays

Methylation detection for 1536 CpG sites was performed as described previously (Bibikova et al. 2006). Briefly, for each CpG site, four probes were designed: two allele-specific oligos (ASO) and two locus-specific oligos (LSO). Each ASO-LSO oligo pair corresponded to either the methylated or unmethylated state of the CpG site. Bisulfite conversion of DNA samples was done using the EZ DNA methylation kit (Zymo Research). After bisulfite treatment, the remaining assay steps were identical to the GoldenGate genotyping assay (Fan et al. 2003), using Illumina-supplied reagents and conditions. Technical replicates of each

bisulfite-converted sample were run. The duplicates all agreed well with each other (average $R^2 = 0.98$) and were averaged together for further analysis. In several cases, essentially the same cell lines were obtained from two different laboratories. In these cases both samples were analyzed and showed good agreement with each other. The array hybridization was conducted under a temperature gradient program, and arrays were imaged using a BeadArray Reader (Barker et al. 2003). Image processing and intensity data extraction software were as described previously (Galinsky 2003). Each methylation data point is represented by fluorescent signals from the M (methylated) and U (unmethylated) alleles. Background intensity computed from a set of negative controls was subtracted from each analytical data point. The ratio of fluorescent signals was then computed from the two alleles: $\beta = (\max(M,0)/|U| + |M| + 100)$. The β -value reflects the fractional methylation level of each CpG site.

Cluster analysis and selection of differentially methylated CpG sites

Using β -values, samples were clustered using agglomerative nesting with the Ward method with dissimilarity between samples defined as $1 - r$, where r is the correlation coefficient (Fig. 1).

We used a supervised analysis method, "Between Group Analysis" (Culhane et al. 2002), to identify differences between four groups of samples: (1) hES cells, (2) somatic stem cells, (3) fully differentiated cells, and (4) cancer cells. For the hES cells, we used only the earliest passage of each line. Computations were carried out using the *bga* function from the MADE4 R package (Culhane et al. 2005). Briefly, samples were projected on a subspace formed by three eigenvectors (number of groups minus one) computed using principle component analysis. The first two eigenvectors cleanly separated three groups, hES samples, cancer samples, and somatic stem cells plus differentiated cells. The third eigenvector separated somatic stem cells from fully differentiated cells. The significance of separation among groups was assessed by randomly permuting sample labels and calculating the ratio of between-group inertia to total inertia. Out of 100 random permutations, the maximum ratio was 0.098, while with the original labels the ratio was 0.287. Sites contributing the most to separation among groups have the most extreme coordinates when projected in the coordinate system defined by the three eigenvectors. We picked the 20 most extreme sites (positive and negative) for each axis, and then filtered out those with very low differences of β -values between groups (maximum mean difference between groups < 0.1). This produced the 49 sites shown on the heat map of Figure 3.

In order to identify sites that provide the best discrimination between hES samples and all normal differentiated samples (including adult stem cells), we performed a *t*-test on the difference in mean methylation level between groups. We selected sites with $P < 0.001$ and with the additional filter of mean $|\Delta\beta| > 0.17$, the estimated error in β (Bibikova et al. 2006). This resulted in a list of 168 sites (Supplemental Table 3). In this test, we also used only the earliest passage for each hES line. The top 25 sites, based on *P*-value, were used to construct the cluster diagram and heat map in Figure 4.

Methylation-specific PCR

The bisulfite-converted genomic DNA was amplified by real-time methylation-specific PCR using two sets of primers for each CpG locus of interest. The MSP primers (Supplemental Table 5) were designed using CpGWare (Chemicon), software originally designed by Todd W. Plaia and Alexander Khripin. The methyl-

ation-specific PCR was performed as described previously (Bibikova et al. 2006).

Acknowledgments

We thank our colleagues at the BIMR, especially Evan Snyder, for contributing to the collegial culture of BIMR's Stem Cell Center. This work was supported in part by National Institutes of Health Grant R44-CA097851 (to J.-B.F.), and a grant from the Alzheimer's Disease Association (to J.F.L.). We wish to thank the private donors to the BIMR Stem Cell Center who make this work possible. Illumina, BeadArray, and GoldenGate are registered trademarks or trademarks of Illumina, Inc.

References

- Allegrucci, C., Denning, C., Priddle, H., and Young, L. 2004. Stem-cell consequences of embryo epigenetic defects. *Lancet* **364**: 206–208.
- Allegrucci, C., Thurston, A., Lucas, E., and Young, L. 2005. Epigenetics and the germline. *Reproduction* **129**: 137–149.
- Amit, M. and Itskovitz-Eldor, J. 2002. Derivation and spontaneous differentiation of human embryonic stem cells. *J. Anat.* **200**: 225–232.
- Andrews, P.W., Damjanov, I., Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C., and Fogh, J. 1984. Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation in vivo and in vitro. *Lab. Invest.* **50**: 147–162.
- Andrews, P.W., Benvenisty, N., McKay, R., Pera, M.F., Rossant, J., Semb, H., and Stacey, G.N. 2005. The international stem cell initiative: Toward benchmarks for human embryonic stem cell research. *Nat. Biotechnol.* **23**: 795–797.
- Barker, D.L., Theriault, D., Che, D., Dickinson, T., Shen, R., and Kain, R. 2003. Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. *Proceedings of SPIE* **4966**: 1–11.
- Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., et al. 2006. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**: 383–393.
- Bodnar, M.S., Meneses, J.J., Rodriguez, R.T., and Firpo, M.T. 2004. Propagation and maintenance of undifferentiated human embryonic stem cells. *Stem Cells Dev.* **13**: 243–253.
- Brimble, S.N., Zeng, X., Weiler, D.A., Luo, Y., Liu, Y., Lyons, I.G., Freed, W.J., Robins, A.J., Rao, M.S., and Schulz, T.C. 2004. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev.* **13**: 585–597.
- Clark, A.T., Bodnar, M.S., Fox, M., Rodriguez, R.T., Abeyta, M.J., Firpo, M.T., and Pera, R.A. 2004. Spontaneous differentiation of germ cells from human embryonic stem cells in vitro. *Hum. Mol. Genet.* **13**: 727–739.
- Cowan, C.A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J.P., Wang, S., Morton, C.C., McMahon, A.P., Powers, D., et al. 2004. Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**: 1353–1356.
- Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. 2005. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**: 1369–1373.
- Culhane, A.C., Perriere, G., Considine, E.C., Cotter, T.G., and Higgins, D.G. 2002. Between-group analysis of microarray data. *Bioinformatics* **18**: 1600–1608.
- Culhane, A.C., Thioulouse, J., Perriere, G., and Higgins, D.G. 2005. MADE4: An R package for multivariate analysis of gene expression data. *Bioinformatics* **21**: 2789–2790.
- Esteller, M. 2002. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene* **21**: 5427–5440.
- Fan, J.B., Oliphant, A., Shen, R., Kermani, B., Garcia, F., Gunderson, K., Hansen, M., Steemers, F., Butler, S.L., Deloukas, P., et al. 2003. Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68**: 69–78.
- Galinsky, V.L. 2003. Automatic registration of microarray images. II. Hexagonal grid. *Bioinformatics* **19**: 1832–1836.
- Heins, N., Englund, M.C., Sjoblom, C., Dahl, U., Tonning, A., Bergh, C., Lindahl, A., Hanson, C., and Semb, H. 2004. Derivation, characterization, and differentiation of human embryonic stem cells.

- Stem Cells* **22**: 367–376.
- Heins, N., Lindahl, A., Karlsson, U., Rehnstrom, M., Caisander, G., Emanuelsson, K., Hanson, C., Semb, H., Bjorquist, P., Sartipy, P., et al. 2006. Clonal derivation and characterization of human embryonic stem cell lines. *J. Biotechnol.* **122**: 511–520.
- Heisler, L.E., Torti, D., Boutros, P.C., Watson, J., Chan, C., Winegarden, N., Takahashi, M., Yau, P., Huang, T.H., Farnham, P.J., et al. 2005. CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. *Nucleic Acids Res.* **33**: 2952–2961.
- Herman, J.G. and Baylin, S.B. 2003. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* **349**: 2042–2054.
- Hoffman, L.M., Hall, L., Batten, J.L., Young, H., Pardasani, D., Baetge, E.E., Lawrence, J., and Carpenter, M.K. 2005. X-inactivation status varies in human embryonic stem cell lines. *Stem Cells* **23**: 1468–1478.
- Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout III, W.M., and Jaenisch, R. 2005. Global loss of imprinting leads to widespread tumorigenesis in adult mice. *Cancer Cell* **8**: 275–285.
- Jacob, S. and Moley, K.H. 2005. Gametes and embryo epigenetic reprogramming affect developmental outcome: Implication for assisted reproductive technologies. *Pediatr. Res.* **58**: 437–446.
- James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. 2005. TGF β /activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**: 1273–1282.
- Kantor, B., Kaufman, Y., Makedonski, K., Razin, A., and Shemer, R. 2004. Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. *Hum. Mol. Genet.* **13**: 2767–2779.
- Keller, G. 2005. Embryonic stem cell differentiation: Emergence of a new era in biology and medicine. *Genes & Dev.* **19**: 1129–1155.
- Lee, M.P. 2003. Genome-wide analysis of epigenetics in cancer. *Ann. N.Y. Acad. Sci.* **983**: 101–109.
- Loring, J.F. and Rao, M.S. 2006. Establishing standards for the characterization of human embryonic stem cell lines. *Stem Cells* **24**: 145–150.
- Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., et al. 2006. Derivation of human embryonic stem cells in defined conditions. *Nat. Biotechnol.* **24**: 185–187.
- Maitra, A., Arking, D.E., Shivapurkar, N., Ikeda, M., Stastny, V., Kassaei, K., Sui, G., Cutler, D.J., Liu, Y., Brimble, S.N., et al. 2005. Genomic alterations in cultured human embryonic stem cells. *Nat. Genet.* **37**: 1099–1103.
- Mandal, A., Tipnis, S., Pal, R., Ravindran, G., Bose, B., Patki, A., Rao, M.S., and Khanna, A. 2006. Characterization and in vitro differentiation potential of a new human embryonic stem cell line, ReliCellhES1. *Differentiation* **74**: 81–90.
- Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. 2005. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14**: R47–R58.
- Murrell, A., Rakyanc, V.K., and Beck, S. 2005. From genome to epigenome. *Hum. Mol. Genet.* **14**: R3–R10.
- Rakyanc, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., et al. 2004. DNA methylation profiling of the human major histocompatibility complex: A pilot study for the human epigenome project. *PLoS Biol.* **2**: e405.
- Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. 2000. Embryonic stem cell lines from human blastocysts: Somatic differentiation in vitro. *Nat. Biotechnol.* **18**: 399–404.
- Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* **282**: 1145–1147.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**: 853–862.
- Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. 2005. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat. Methods* **2**: 185–190.
- Zeng, X., Chen, J., Liu, Y., Luo, Y., Schulz, T.C., Robins, A.J., Rao, M.S., and Freed, W.J. 2004. BG01V: A variant human embryonic stem cell line which exhibits rapid growth after passaging and reliable dopaminergic differentiation. *Restor. Neurol. Neurosci.* **22**: 421–428.
- Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J.E., Crompton, L.A., John, R., Feil, R., and Brockdorff, N. 2005. Global hypomethylation of the genome in XX embryonic stem cells. *Nat. Genet.* **37**: 1274–1279.

Received March 24, 2006; accepted in revised form June 22, 2006.



Human embryonic stem cells have a unique epigenetic signature

Marina Bibikova, Eugene Chudin, Bonnie Wu, et al.

Genome Res. 2006 16: 1075-1083

Access the most recent version at doi:[10.1101/gr.5319906](https://doi.org/10.1101/gr.5319906)

Supplemental Material

<http://genome.cshlp.org/content/suppl/2006/08/09/gr.5319906.DC1>

References

This article cites 40 articles, 6 of which can be accessed free at:
<http://genome.cshlp.org/content/16/9/1075.full.html#ref-list-1>

Open Access

Freely available online through the *Genome Research* Open Access option.

License

Freely available online through the Genome Research Open Access option.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
