Methods

# DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression

Shintaro Yagi,[1] Keiji Hirabayashi,[1] Shinya Sato,[1] Wei Li,[2] Yoko Takahashi,[1] Tsutomu Hirakawa,[1] Guoying Wu,[1] Naoko Hattori,[1] Naka Hattori,[1] Jun Ohgane,[1] Satoshi Tanaka,[1] X. Shirley Liu,[3] and Kunio Shiota[1,4,5]

[1]*Laboratory of Cellular Biochemistry, Department of Animal Resource Sciences/Veterinary Medical Sciences, The University of Tokyo, Tokyo 113-8657, Japan;* [2]*Division of Biostatistics, Dan L. Duncan Cancer Center, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA;* [3]*Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, Massachusetts 02115, USA;* [4]*National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8561, Japan*

DNA methylation constitutes an important epigenetic regulation mechanism in many eukaryotes, although the extent of DNA methylation in the regulation of gene expression in the mammalian genome is poorly understood. We developed D-REAM, a genome-wide DNA methylation analysis method for tissue-dependent and differentially methylated region (T-DMR) profiling with restriction tag-mediated amplification in mouse tissues and cells. Using a mouse promoter tiling array covering a region from −6 to 2.5 kb (~30,000 transcription start sites), we found that over 3000 T-DMRs are hypomethylated in liver compared to cerebrum. The DNA methylation profile of liver was distinct from that of kidney and spleen. This hypomethylation profile marked genes that are specifically expressed in liver, including key transcription factors such as *Hnf1a* and *Hnf4a*. Genes with T-DMRs, especially those lacking CpG islands and those with HNF-1A binding motifs in their promoters, showed good correlation between their tissue-specific expression and liver hypomethylation status. T-DMRs located downstream from their transcription start sites also showed tissue-specific gene expression. These data indicate that multilayered regulation of tissue-specific gene function could be elucidated by DNA methylation tissue profiling.

[Supplemental material is available online at www.genome.org and at http://www.vm.a.u-tokyo.ac.jp/seika/D-REAM/. The array data from this study have been submitted to ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) under accession no. E-TABM-551.]

In multicellular organisms, cells, and tissues form as a result of differentiation of a single fertilized egg, and phenotypes are inherited over several cell generations without alteration in the DNA sequences. Epigenetic systems are recognized as memory systems for these inheritable gene functions and, in mammals, they comprise DNA methylation and histone modifications of chromatin. DNA methylation in tissue-dependent and differentially methylated regions (T-DMRs) is involved in expression of tissue-specific genes as well as expression of key transcription factors that constitute transcription networks governing tissue or cell specificity (Shen and Maniatis 1980; Cho et al. 2001; Imamura et al. 2001; Hattori et al. 2004b, 2007; Nishino et al. 2004). Abnormal methylation of T-DMRs has been implicated in the pathogenesis of certain diseases (Jones 2002; Ushijima 2005).

DNA methylation occurs at the cytosine residue of CpG dinucleotides, which are unevenly distributed within the mammalian genome (Bird 1980). CpG islands (CGIs) have been identified as CpG-rich regions that are associated with ~50% of the promoter regions in the mouse genome (Bird et al. 1985; Gardiner-Garden and Frommer 1987). Previous genome-wide DNA methylation analyses, focusing on CGIs, have indicated that every cell and tissue type has a unique DNA methylation profile, comprising at least hundreds of T-DMRs (Ohgane et al. 1998; Shiota et al. 2002; Strichman-Almashanu et al. 2002), and these data suggested that a methylation profile could be used to identify cell types (Shiota 2004).

To identify genes with differentially methylated regions, several microarray technologies have been developed (Lieb et al. 2006), and microarray technology has been applied to identify aberrantly methylated regions in cancer cells and characterize cell lines such as human embryonic stem cells (Hatada et al. 2006; Keshet et al. 2006; Ordway et al. 2006; Rauch et al. 2006; Shen et al. 2006). However, the DNA methylation profiles obtained have not been directly related to gene function (Ching et al. 2005; Eckhardt et al. 2006; Khulan et al. 2006). The limited number of loci or regions available for genome-wide analysis of normal cells or tissues and the existence of method biases can affect the implementation of methylated profiles (for review, see Khulan et al. 2006).

**[5]Corresponding author.**
**E-mail ashiota@mail.ecc.u-tokyo.ac.jp; fax 81-3-5841-8189.**
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.074070.107.

# DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression

Shintaro Yagi, Keiji Hirabayashi, Shinya Sato, et al.

*Genome Res.* 2008 18: 1969-1978 originally published online October 29, 2008
Access the most recent version at doi:10.1101/gr.074070.107

We developed a novel, low-bias method for genome-wide DNA methylation analysis, and examined the DNA methylation profile of the promoter regions in normal mouse liver by comparing them with those of cerebrum, kidney, and spleen. The results indicate that the resultant methylation profile was implicated in tissue-specific function.

## Results

### Features of T-DMR profiling by restriction tag-mediated amplification

To illustrate the genome-wide mouse DNA methylation profile, we developed a method involving T-DMR profiling with restriction tag-mediated amplification (D-REAM), which combined microarray technology and modified ligation-mediated polymerase chain reaction (LM-PCR) (Fig. 1A; Supplemental Fig. S1A). D-REAM recognizes various combinations of microarrays and restriction enzymes for LM-PCR. In this study, we used the GeneChip DNA microarray, tiled with ~4.4 million 25-nt oligomers corresponding to regions located from approximately −6 to 2.5 kb of transcription start sites (TSS) of 30,120 Ensembl mouse transcripts (Supplemental Table S1; Supplemental Fig. S2A). To screen unmethylated regions, we used HpyCH4IV, a methylation-sensitive restriction enzyme that recognizes ACGT residues. Such residues are distributed throughout the mouse genome in a less biased manner than HpaII sites, which are localized mainly at CGIs around the TSS on the promoter array (Supplemental Fig. S2B). The microarray probes covered regions comprising 10.1% of all the HpyCH4IV sites (1,751,098) in the mouse genome (Fig. 1C).

When the mouse genome was digested with HpyCH4IV in silico, it generated fragments with median and average sizes of 907 and 1468 bp, respectively (Supplemental Fig. S2B). Because of the hypermethylation status of mouse genome, actual mouse liver HpyCH4IV fragments were deemed too large for efficient PCR amplification. To address this issue, the D-REAM method uses TaqI, a methylation-insensitive restriction enzyme, to reduce the size of the single-digested fragments (Fig. 1B). The modified LM-PCR protocol facilitates the selective amplification of unmethylated HpyCH4IV–TaqI and HpyCH4IV–HpyCH4IV, but not TaqI–TaqI fragments (Supplemental Fig. S1A,B). The selective amplification by this process was confirmed by using fragments digested by rarely occurring restriction enzymes, such as NotI (Supplemental Fig. S1C).

To analyze the D-REAM microarray data, we applied model-based analysis of tiling arrays (MAT) (Johnson et al. 2006). MATscores represent the enrichment of unmethylated fragments and are indicative of the relative methylation status of the HpyCH4IV site. They were visualized using the Integrated Genome Browser (IGB). Intra-genomic comparison can be affected by PCR sequences or fragment length biases; however, these effects were minimized by comparing the same regions in different tissues or cells. In this study, we designated these differentially methylated HpyCH4IV sites as T-DMRtags that represent T-DMRs (Fig. 1A).

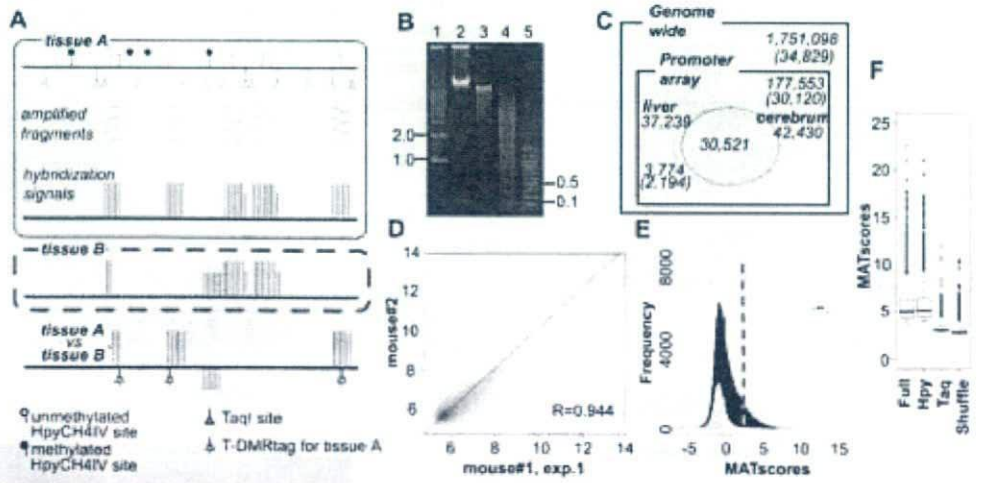



**Figure 1.** DNA methylation profiles were analyzed by D-REAM. (*A*) Illustration of the D-REAM method. Genomic DNA was digested with methylation-sensitive restriction enzyme HpyCH4IV and amplified by modified LM-PCR (Supplemental Fig. S1). Amplified fragments (gray bars) were hybridized with mouse promoter tiling array (*upper* panel). Array signal intensities (vertical bars) were analyzed to identify regions corresponding to fragments in unmethylated HpyCH4IV loci. Comparison of signals from different samples enabled identification of differentially methylated regions (*lower* panel). HpyCH4IV loci overlapping with regions yielding differential signals were defined as T-DMRtags. (*B*) Agarose gel electrophoresis of undigested (lane *2*), HpyCH4IV-digested (lane *3*), and HpyCH4IV–TaqI-digested (lane *4*) mouse liver DNA. Positions corresponding to 0.1, 0.5, 1.0, and 2.0 kbp (lanes *1*,*5*) are indicated on one side of the gel image. (*C*) Venn diagram of DNA methylation status at HpyCH4IV sites in mouse liver and cerebrum. Numbers without parentheses represent numbers of HpyCH4IV sites, while Ensembl transcripts IDs are in parentheses. *Outer* and *inner* rectangles represent whole mouse genome and regions covered by the promoter tiling array, respectively. Ovals indicate unmethylated HpyCH4IV sites of liver and cerebrum identified by D-REAM. (*D*) Correlation of microarray probe intensities in duplicate mouse liver experiments, plotted on logarithmic axes (base 2). (*E*) MATscore distribution of array regions corresponding to the TaqI–TaqI fragments (gray) and HpyCH4IV-digested fragments (black). The dotted line represents the MATscore cutoff value. (*F*) Reliability of comparative MAT analysis. Bar-plots of MATscores of the hypomethylated regions identified by MAT ($P < 10^{-3}$) using full .bpmap (Full) and subsets of .bpmap corresponding to HpyCH4IV fragments (Hpy) and TaqI–TaqI fragments (Taq). Shuffle column MATscores obtained by using both treatment and control samples containing both liver and cerebrum data from different mice. The boxes, and lines within the boxes, represent the interquartile ranges and medians of the ratios, respectively.

### T-DMRs hypomethylated in mouse liver compared to cerebrum

D-REAM assays of liver DNA obtained from two male mice were performed, and reproducible microarray results were obtained (Fig. 1D; Supplemental Fig. S1D). The number of fragments above the MATscore cutoff value of 2.44 was larger at regions corresponding to HpyCH4IV fragments than at those corresponding to TaqI–TaqI fragments (Fig. 1E). This indicated selective amplification of fragments with unmethylated HpyCH4IV sites. At least 20.9% of HpyCH4IV sites were estimated to be unmethylated within the regions covered by the microarray (Fig. 1C). In mouse cerebra (correlation coefficient of 0.955 of the biological duplicate), MAT analysis predicted 42,430 sites of unmethylated HpyCH4IV sites ($P < 10^{-3}$), and 71.9% (30,521 loci) of these sites overlapped those in the liver tissue.

In a comparative analysis of liver and cerebrum, the differences in MATscores were significantly large at regions of HpyCH4IV fragments compared to those of HpyCH4IV fragments with a shuffled combination, and to those of TaqI–TaqI fragments (Fig. 1F). A total of 3774 of differentially hypomethylated T-DMRtags were identified in liver ($P < 10^{-3}$). These tags were located in neighboring 10-kb regions of the TSS for 2194 (7.28% of all IDs) Ensembl transcripts (Fig. 1C).

### Positions of T-DMRs relative to the TSS

We classified the genes for these 2194 transcripts into two types, CGI and non-CGI genes, according to TSS position. CGI gene TSS are located within 1 kb from the CGI, and non-CGI genes are located further than 1 kb from the CGI. The positions of T-DMRtags were distinct using this criterion (Fig. 2A). T-DMRtags neighboring non-CGI genes were observed up to −6 kb distal to and 2.5 kb downstream from TSS, and exhibited distribution patterns similar to the probes. No correlations were observed between T-DMR distribution and CpG density (observed/expected), GC percentage, or localization of HpyCH4IV sites (Fig. 2A; Supplemental Fig. S2A). Among the CGI genes, the distribution of T-DMRs seemed to negatively correlate with CpG density and the sequence conservation score among animals, suggesting that CGIs and first exons would be T-DMR-poor regions (Fig. 2A,B).

T-DMRtags in eight liver-specific non-CGI genes were confirmed using combined bisulfite restriction analysis (COBRA) (Supplemental Fig. S3A,B; Xiong and Laird 1997). The degrees of methylation were not proportional to the differences in the MATscores that were observed in three sites of the *Fga* loci, but all 15 HpyCH4IV sites were confirmed to be differentially hypomethylated in liver. The T-DMRs were distributed from the 5′-upstream to the 3′-downstream region of TSS, similar to the distribution of all T-DMRtags in the non-CGI genes (Fig. 2A).

In liver-specific *Gnmt* gene, T-DMRtags were observed from −5.9 to −4.3 kb 5′ upstream and from 2.2 to 2.4 kb 3′ downstream from TSS. COBRA showed an unmethylated status of HpyCH4IV in CGIs of both liver and cerebrum, and a hypomethylated status of each T-DMRtags in liver (Fig. 2D). Similar meth-

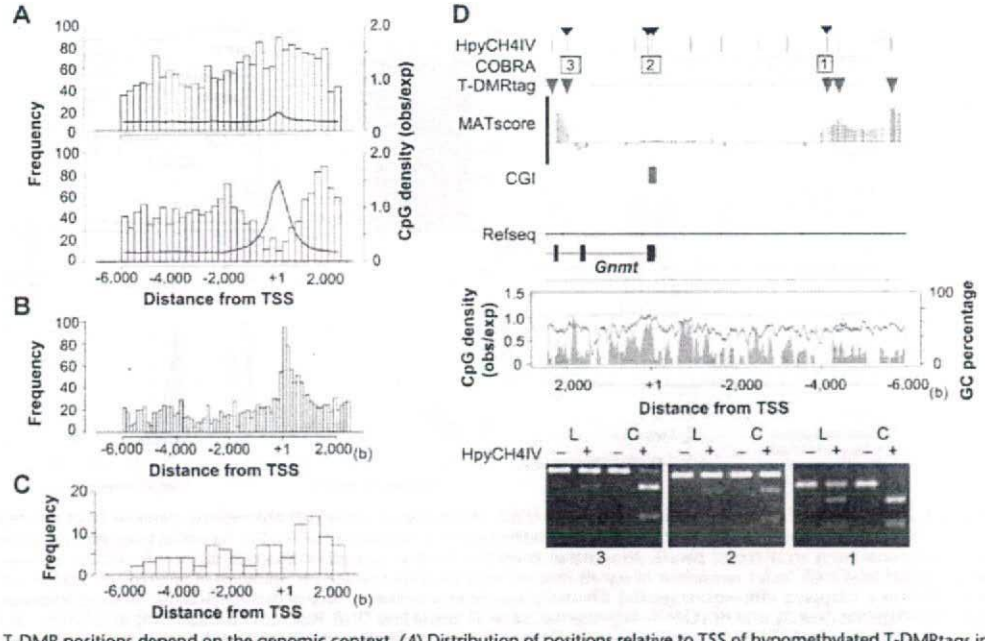



**Figure 2.** T-DMR positions depend on the genomic context. (*A*) Distribution of positions relative to TSS of hypomethylated T-DMRtags in liver. *Upper* and *lower* panels display distributions in non-CGI and CGI genes, respectively. The width of histogram units is 250 bp. CpG densities are indicated by blue lines. (*B*) Center of phastCons track regions in all CGI genes on chromosomes 5, 12, and 15 (obtained from UCSC genome browser database) plotted with a histogram unit width of 125 bp. (*C*) Positions of T-DMRtags on liver-specific non-CGI genes with HNF1 motifs with expression levels in liver >2-fold those in cerebrum, plotted with a 500-bp histogram category width. (*D*) T-DMRs neighboring the *Gnmt* genes analyzed by COBRA. *Upper* panel displays the position of HpyCH4IV, the regions of restriction mapping with the analyzed HpyCH4IV site, indicated by small arrowheads on the *top*, and positions of T-DMRs plotted over the comparative MATscores on IGB browser from 6000 bp upstream to 2500 bp downstream from the TSS. *Middle* panel shows CpG density (blue) and GC percentage (gray line) in this region. *Bottom* panels show agarose-gel electrophoresis images of COBRA. Hypomethylated fragments converted by bisulfite treatment were resistant to HpyCH4IV digestion (+). L and C indicate liver and cerebrum samples, respectively.

ylation patterns of HpyCH4IV sites in five other CGI genes were confirmed by COBRA (Supplemental Fig. S3C,D).

### Biased ontology annotations of genes with T-DMRs with regard to CGIs

To determine features represented by genes with T-DMRs hypomethylated in liver, we applied ontology analysis, using g:GOSt in the g:profiler web database (Table 1; Supplemental Table S2) (Reimand et al. 2007). Significantly enriched ontology terms for biological processes (BPs) ($P < 10^{-5}$) were observed in 1817 g:profiler IDs that had been converted from 2194 Ensembl transcripts with T-DMRs for analysis. These BPs code for genes responsible for metabolism of organic acids and lipids and responses to defense and stress. Among these overrepresented terms, were found liver dominant tissue-specific functions present in the coagulation cascade (described in the Kyoto Encyclopedia of Genes and Genomes [KEGG] pathway database) (Kanehisa et al. 2006), and folate and methyl group metabolism (Fig. 3A,B) (Williams and Schalinske 2007).

The ratios of CGI genes to non-CGI genes were significantly biased among some ontology terms ($P < 5 \times 10^{-2}$, $\chi^2$ test), with non-CGI genes overrepresented among the genes encoding proteins exported to extracellular regions (Table 1). To avoid preexisting bias in the ontology terms in our criteria, we analyzed them using functional annotation tools in the DAVID Bioinformatics Resources 2007 website (Huang da et al. 2007) using lists of CGI or non-CGI genes as background. Most terms observed in the g:GOSt analysis were significantly overrepresented in the non-CGI genes with T-DMRs compared to CGI genes with T-DMRs. Among the CGI genes, only mitochondrial genes were significantly overrepresented, and one of the identified mitochondrial CGI genes with T-DMRs was *Cpt2* (Supplemental Fig. S3C).

### Genes encoding transcription factors responsible for expression of liver genes with T-DMRs

Transcription factor motifs were analyzed by referring to the MAPPER database in the 1-kb 5′-upstream region of TSS (Marinescu et al. 2005). When non-CGI genes with T-DMRs were analyzed, we found significantly enriched motifs, including HNF1A, HNF4, and RXRa (Table 2). In contrast, there were no overrepresented motifs among CGI genes with T-DMRs.

*Hnf1a* and *Hnf4a* are expressed in a tissue-specific manner and are involved in regulating liver-specific gene expression (Schrem et al. 2002). The MATscores and, hence, the methylation status of these regions were significantly different between liver and cerebrum (Fig. 4A,B). Bisulfite sequencing identified four CpGs hypomethylated in liver, corresponding to the T-DMRtags at −523 bp upstream of *Hnf1a* TSS (Fig. 4A). Among the six T-DMRtags of *Hnf4a*, two located at downstream regions of TSS and two at 5′ distal regions were present in the first exon of *0610008F07Rik*, a gene in which transcription initiates from the 5-kb upstream region of *Hnf4a* TSS in the opposite direction. Bisulfite DNA sequencing revealed that 27 CpGs in these 5-kb regions were hypomethylated in liver but hypermethylated in cerebrum (Fig. 4B).

We further investigated T-DMRs in other liver-enriched transcription factors that support liver-specific gene expression (Giguere 1999; Handschin and Meyer 2005). In the cases of *Nr1h3* (*Lxr*) and *Nr1i2* (Pregnane X receptor, *Pxr*), the T-DMRs in

**Table 1.** Annotation analysis of genes with liver T-DMRtags

| Ontology type[a] | Term | DAVID[b] |
|---|---|---|
| GO:BP | Lipid metabolic process | NC |
| | Cellular lipid metabolic process | NC |
| | Response to stress | |
| | Generation of precursor metabolites and energy | NC |
| | Organic acid metabolic process | NC |
| | Carboxylic acid metabolic process | NC |
| | Monocarboxylic acid metabolic process | |
| GO:MF | Vitamin binding | NC |
| | Cofactor binding | NC |
| | FAD binding | NC |
| Biased to non-CGI genes | | |
| GO:BP | Defense response | |
| | Response to wounding | NC |
| | Inflammatory response | |
| | Acute inflammatory response | |
| GO:CC | Extracellular region | NC |
| | Extracellular region part | |
| | Extracellular space | NC |
| KEGG | Complement and coagulation cascades | NC |
| Biased to CGI genes | | |
| GO:CC | Cytoplasm | NC |
| | Cytoplasmic part | NC |
| | Mitochondrion | NC, CGI |
| | Mitochondrial part | |
| GO:MF | Catalytic activity | NC |

Detailed data are shown in Supplemental Table S3. $\chi^2$ tests were applied to examine the difference in the proportions of CGI and non-CGI genes for each criterion among all the genes (1817 genes) containing T-DMRtags. Percentage of non-CGI genes among all 1817 genes is 53.6%.
[a]Ontology types of GO:BP, GO:CC, and GO:MF indicate biological process, cellular component, and molecular function in Gene Ontology criteria, respectively. KEGG represents KEGG pathway database.
[b]The DAVID column indicates the overrepresentation of the terms in DAVID 2007 analysis among all genes classified into the same criterion according to the position of CGIs. NC, non-CGI genes; CGI, CGI genes.

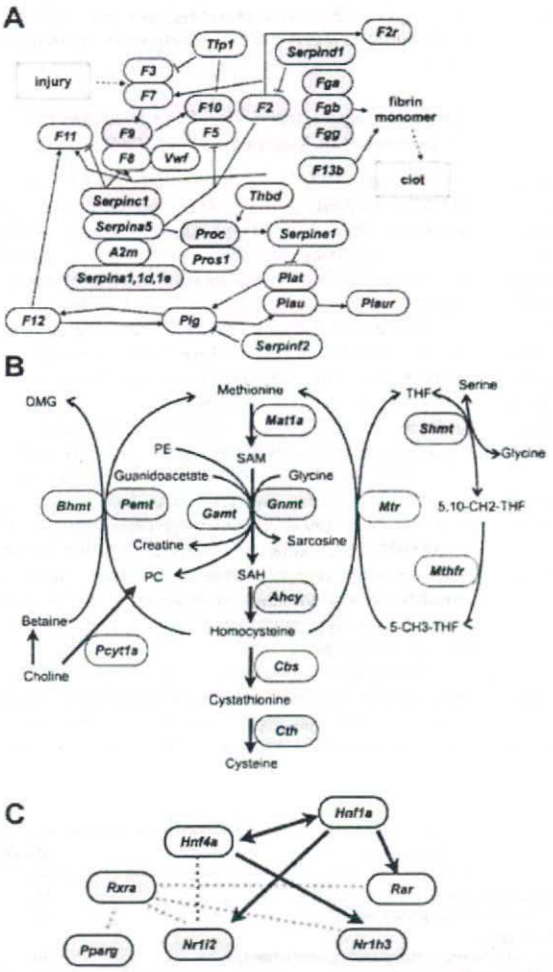



**Figure 3.** Genes with T-DMRtags (gray) in the complement and coagulation cascade in the modified KEGG pathway map (ID 04610) (*A*), and in the folate and methyl group metabolism pathway (Williams and Schalinske 2007) (*B*). (*C*) Transcription factor network for liver-specific gene expression. Arrows indicate that the gene expression is controlled by transcription factors. Dotted lines represent molecular interaction between factors.

liver were localized downstream from their TSS (Fig. 4C,D). T-DMRs in the *Rxra* gene were localized between exons 1 and 2 (Fig. 4E), where an alternative TSS for the testis-specific transcript, which is detected in liver, is located (Brocard et al. 1996). These data suggested that T-DMRs could be involved in gene regulation of transcription factors, aiding liver-specific gene expression (Fig. 3C).

## Expression of genes with HNF1 motifs and T-DMRs

Previously, 222 HNF1-binding promoters were identified in human hepatocytes by chromatin immunoprecipitation combined with DNA microarray analysis (ChIP-chip) experiments (Odom et al. 2004). Assuming that mouse orthologs of these genes would be bound to HNF1 in mouse liver, we searched for these orthologs using g:Orth in the g:Profiler web database, and observed that 43 genes of 174 orthologs (unique Entrez gene IDs) were accompanied with T-DMRs (Supplemental Fig. S4A). Tissue specificity of expression was investigated in these 43 orthologs and in 180 Entrez gene IDs with HNF-1A motifs (TRANSFAC ID M00790) and T-DMRs (Supplemental Fig. S4B), which were selected by referring to the MAPPER database. The gene expression levels in liver were more than twofold of those in cerebrum in 32 of 43 orthologs and in 68 of 180 genes with HNF-1A motifs (Fig. 5A). In cerebrum, 75% of genes exhibited expression levels <229.6 (gcRMA preprocessed data), which indicated that gene hypermethylation might repress these genes in cerebrum (Fig. 5B). Among CGI genes with T-DMRs, liver-specific expressed genes were significantly overrepresented in genes with HNF-1A motifs, but less significantly in those with HNF4 motifs (Fig. 5C). These data supported the conclusion that HNF1 and T-DMRs are involved in the regulation of these genes in mice.

## Correlations between expression levels of genes and T-DMRs in somatic tissues

We examined correlations between T-DMRs and gene expression in liver (Fig. 5C). The expression levels of genes with T-DMRs in liver were significantly higher than those in cerebrum, especially with regard to non-CGI genes. In 77 genes (Ensembl gene IDs) with HNF1-binding motifs that were expressed in liver (Fig. 2C), the T-DMR distribution profile was similar to that of all non-CGI genes (Fig. 2A).

In liver-specific CGI genes, the positions of T-DMRs were biased toward the regions 0.5–2.5 kb, which are 3′ downstream from the TSS (Fig. 5C,D). These regions corresponded to the first introns, judged by the distribution of the gene conservation index and by CpG density (Fig. 2B). These data suggested that T-DMRs in noncore promoter regions could be involved in the regulation of gene plurality.

We analyzed the DNA methylation status at T-DMRs in mouse kidney and spleen in genes with expression levels higher in liver than in cerebrum. The regions corresponding to these T-DMRs were clustered into four groups according to their

**Table 2. Transcription factor motifs overrepresented in genes with T-DMRs among non-CGI genes**

| MAPPER factor name[a] | All | T-DMR | $\chi^2$ test (*P*-value)[b] |
|---|---|---|---|
| ARP-1 (NR2F2, HNF4A) | 1580 | 160 | $8.76 \times 10^{-4}$ |
| COUP direct repeat 1 (NR2F2) | 1198 | 131 | $1.27 \times 10^{-4}$ |
| COUP-TF:HNF-4 (NR2F2, HNF4A) | 1292 | 136 | $5.17 \times 10^{-4}$ |
| ER-alpha (ESR1) | 2250 | 232 | $6.03 \times 10^{-6}$ |
| FOXD3 (FOXD3) | 4111 | 400 | $2.73 \times 10^{-8}$ |
| GCR1[c] | 1268 | 133 | $7.11 \times 10^{-4}$ |
| HNF-1A (HNF1A) | 1661 | 196 | $2.24 \times 10^{-9}$ |
| HNF-4 direct repeat 1 (HNF4A) | 1051 | 116 | $2.41 \times 10^{-4}$ |
| HNF-4alpha (HNF4A) | 1232 | 131 | $4.45 \times 10^{-4}$ |
| HNF-4alpha1 (HNF4A) | 3113 | 348 | $6.45 \times 10^{-15}$ |
| NF-kappaB (NFKB) | 4484 | 430 | $2.08 \times 10^{-8}$ |
| RAP1[c] | 4201 | 379 | $5.03 \times 10^{-4}$ |
| RXR-alpha (RXRA) | 1230 | 142 | $3.11 \times 10^{-6}$ |
| Zic3 (ZIC3) | 386 | 54 | $4.42 \times 10^{-5}$ |

[a]Detailed information is available from MAPPER factors table (http://bio.chip.org/mapper/factors-table, free registration required). Symbols in parentheses correspond to the factors described in the models.
[b]$\chi$-Square tests were applied to examine the difference in the proportions of each transcription factor motif between all non-CGI and non-CGI carrying T-DMRs.
[c]Models are based on the yeast transcription factors.

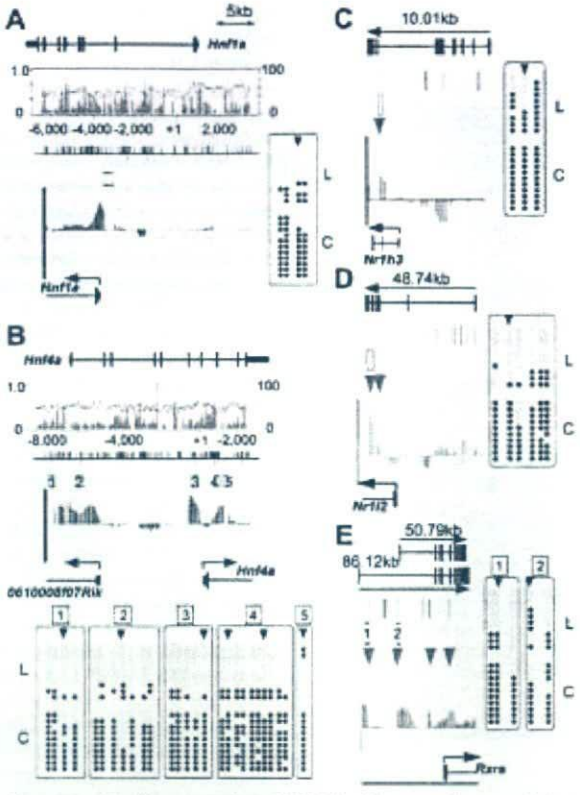



**Figure 4.** Bisulfite sequencing of T-DMRs of liver-specific transcription factors, as *Hnf1a* (*A*), *Hnf4a* (*B*), *Nr1h3* (*C*), *Nr1i2* (*D*), and *Rxra* (*E*). Genomic structures are presented at the *top* of each figure section. The graphs in boxes toward the center in *A* and *B* represent CpG density (blue) and GC percentages (gray). The bars visible along the *top* of the center lines in *A* and *B* represent CpG dinucleotide positions; bars *below* represent HpyCH4IV sites. Boxes and arrowheads represent T-DMRs and T-DMRtags, respectively. IGB plots of comparative microarray signals corresponding to the regions in the abovementioned figures are displayed toward the *bottom* of the middle sections. Bisulfite sequencing data obtained for 10 isolates from liver (L) and cerebrum (C) are summarized at the *bottom* or *side* of the figure section. Open and closed circles represent unmethylated and methylated CpG, respectively.

MATscores of comparative D-REAM analysis, representing relative methylation status (Fig. 5E). Distinct DNA methylation profiles were observed among tissues, with spleen hypomethylated T-DMRs classified into clusters 1 and 2. Cluster 3 contained genes that hypomethylated in both liver and kidney, while cluster 4 contained genes concerned with liver-specific hypomethylation. Although the expression of these genes was repressed in both spleen and cerebrum (Supplemental Fig. S5A), the expression levels in kidney were similar to those in liver in clusters 1 and 2, exhibited a greater distribution toward kidney in cluster 3, and toward liver in cluster 4 (Fig. 5F; Supplemental Fig. S5). In the genes with HNF-1A motifs, *Hnf4a* was classified into cluster 3, while *Nr1h3*, *Nr1i2*, and *Serpina1e* were placed in cluster 4 (Supplemental Figs. S4, S5B). These data indicated that tissue specificity of gene expression in these regions would reflect patterns of T-DMRs of the genes containing tissue-specific transcription factors and their targets.

## Discussion

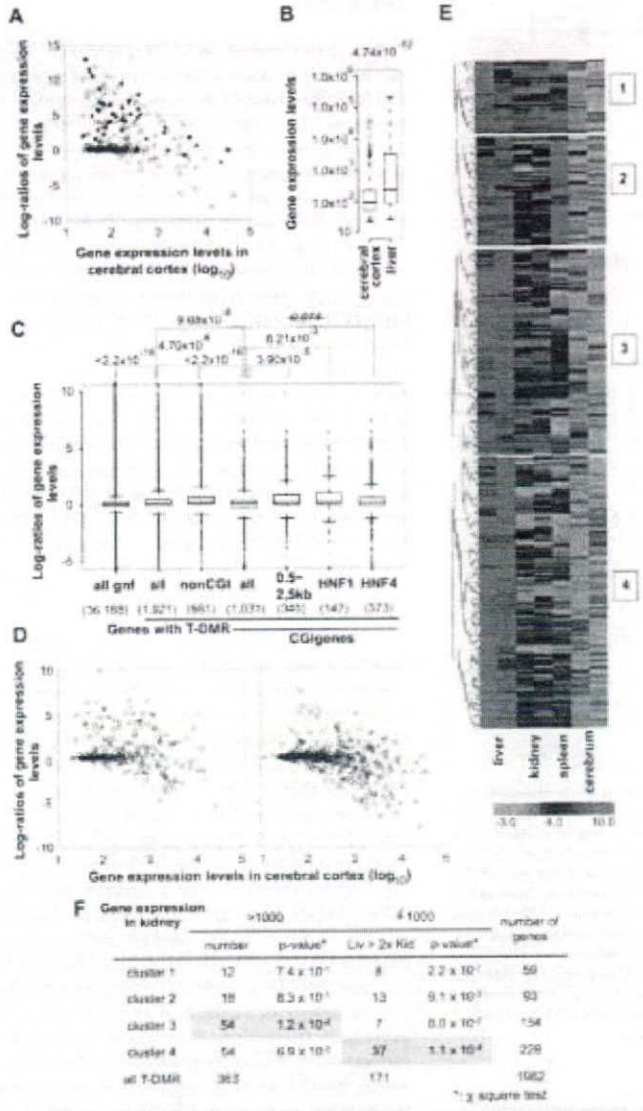
D-REAM, performed using a high-density genome tiling array for the mouse promoter regions, revealed thousands of T-DMRs that are hypomethylated in liver and indicated that these T-DMRs represent the profiles of genes specifically expressed in liver, including those responsible for the liver phenotype and for transcription factors that regulate liver-specific gene expression.

A combination of transcription factors (e.g., HNF1 and HNF4a) is reported to confer liver-specific gene expression (Schrem et al. 2002). HNF1 affects the gene expression of HNF4a, and vice versa (Ktistaki and Talianidis 1997), and these factors, along with other liver-enriched transcription factors such as NR1H3 (LXR), NR1I2 (PXR), RXR, and PPARγ, affect each other's expressions and functions (Geier et al. 2007). D-REAM indicated that T-DMR is involved in the expression of these transcription factors, and that many genes with HNF-1A motifs and T-DMRs are specifically expressed in liver and kidney (Figs. 3, 5F). D-REAM revealed coordinated DNA methylation in transcription factors and the target genes in somatic tissues including liver, kidney, spleen, and cerebrum, and that the combinations of them were distinct among tissues. These results indicated that the DNA methylation profile, comprising T-DMRs of transcription factors and their target genes, is responsible for tissue-specific gene expression in somatic tissues and, consequently, for their function.

Correlations between T-DMRs and transcriptional regulation have been revealed in several CGI and non-CGI genes by the previous investigations on T-DMRs focusing primarily on short core promoter regions (Imamura et al. 2001; Hattori et al. 2004a; Nishino et al. 2004). Genome-wide DNA methylation analysis of the human promoter regions (−700 to 200 bp from the TSS) suggested that the DNA methylation status of a limited number of genes correlated with transcriptional activity (Weber et al. 2005, 2007). A HELP assay, another genome-wide study, revealed plurality of T-DMRs within 1 kb in promoter regions, suggesting their involvement in tissue-specific expression (Khulan et al. 2006). In our study, D-REAM revealed that T-DMRs are localized at a few kilobases both upstream and downstream of the TSS and that the DNA methylation status of these T-DMRs correlates with the transcriptional activities of the neighboring genes. Preferential tissue-specific expression of non-CGI genes has been reported previously (Yamashita et al. 2005), but in the present study, hypomethylation and gene expression were also observed in CGI genes. Thus, the DNA methylation profile is expected to reflect the tissue-specific gene expression profile.

It has become feasible to identify the DNA methylation status of every CpG dinucleotide in plant (Cokus et al. 2008; Lister et al. 2008) and mammalian (Meissner et al. 2008) genomes; however, high redundancy in mammalian genome sequences narrows the window of analysis. All known analysis methods have bias windows, and different windows illustrate different results. In the mouse genome, the HpaII/MspI sites are concentrated in the region between −1 kb and +1 kb from the TSS. In the case of meDIP, the precipitation efficiency of methylated DNA depended on the density of CpG, whose distribution is biased (Keshet et al. 2006; Weber et al. 2007). We suggest that the different T-DMR profiles observed in this study were due to the relatively low bias of HpyCH4IV.

Microarray technology for the analysis of DNA methylation has advantages and disadvantages (Khulan et al. 2006). When compared with systems using isoschizomers, D-REAM is flexible




F

| Gene expression in kidney | >1000 | | <1000 | | number of genes |
|---|---|---|---|---|---|
| | number | p-value* | Liv > 2x Kid | p-value* | |
| cluster 1 | 12 | $7.4 \times 10^{-1}$ | 8 | $2.2 \times 10^{-1}$ | 59 |
| cluster 2 | 18 | $8.3 \times 10^{-1}$ | 13 | $9.1 \times 10^{-1}$ | 93 |
| cluster 3 | 54 | $1.2 \times 10^{-4}$ | 7 | $8.0 \times 10^{-1}$ | 154 |
| cluster 4 | 54 | $6.9 \times 10^{-1}$ | 37 | $1.1 \times 10^{-4}$ | 228 |
| all T-DMR | 363 | | 171 | | 1982 |

*: $\chi$ square test

**Figure 5.** The tissue-specific DNA methylation profiles and gene expression in mouse tissues. (*A*) Expression levels of human gene orthologs identified by the ChIP-Chip experiment using anti-HNF1 (closed rectangles) and of genes with HNF-1A motifs (model IDs T01211 and T00368) classified by the MAPPER database (open circles). (*B*) The distribution of expression levels of the genes listed in *A*, in cerebrum and liver represented by box plots. Expression levels were box-plotted with logarithmic scale (base = 10). (*C*) Box plots of log ratios (base = 2) of liver and cerebrum gene expression indicate factors affecting liver-specific expression of genes with T-DMRs. *P*-values obtained from Wilcoxon's matched-pair signed rank test are indicated on the *top* of the plot. (*D*) Correlation of expression between two tissues. CGI genes containing T-DMRs are divided into two groups by the position of T-DMRs: T-DMRs within 0.5 to 2.5 kb downstream from the TSS (*left* panel) and those outside of this region (*right* panel). Gene expression levels and ratios of gene expression levels are expressed with logarithmic scale of base 10 and base 2, respectively. Numbers of liver-specific genes expressed, with expression levels <1000 in cerebral cortex and at a liver:cerebral cortex level ratio of >2, in the *left* and *right* panels are 60 out of 346 and 45 out of 685, respectively. (*E*) K-means clustering of regions corresponding to T-DMRs by Pearson's correlations of their MATscores. The ranges of the MATscores represented in the plot are shown at the *bottom* of the panels. The MATscores were obtained by MAT analysis of D-REAM data from liver, kidney, and spleen using cerebrum data as the control. (*F*) $\chi^2$ test for distributions of genes, classified by the expression levels, first in kidney (>1000 or not) and those in the later set divided by their expression in liver (>2-fold of those in kidney or not), in each cluster. Statistically significant distributions are shadowed in pink.

because it enables selective amplification of DNA fragments digested with any restriction enzyme, e.g., HpyCH4IV and NotI, a methylation-sensitive enzyme with recognition sites without isoschizomers. One of the disadvantages in D-REAM is that intra-genomic comparison of MATscores between different genomic loci does not always represent differences in DNA methylation. However, a panel of profiles comprising several samples allowed us to conduct a broad intragenomic comparison due in part to the presence of numerous T-DMRs that could provide standard sample data.

A recent study revealed that the T-DMR distribution in randomly selected NotI sites was disproportionate in non-CGI loci, which were located both upstream of the TSS and in the intronic regions (Sakamoto et al. 2007). Present study demonstrates that the T-DMR distribution pattern appears to be independent of CpG density and GC content, except in regions around the TSS in CGI genes. In addition, numerous T-DMRs are localized at 3′ regions downstream from the TSS. T-DMRs were also identified in distal regions (up to 6 kb from the TSS). Furthermore, the distribution patterns of hypomethylated T-DMRs in cerebral and liver genes are similar (S. Yagi, K. Hirabayachi, T. Hirakawa, S. Sato, C. Maeda, J. Ohgane, S. Tanaka, and K. Shiota, unpubl.). These findings suggest that the function of T-DMRs might differ with respect to their positions in the genes.

In our analyses, especially those involving CGI genes, hypomethylation of T-DMRs in the 3′-regions downstream from the TSS was correlated with high gene expression levels in liver. The first exon and the first intron are hot spots of antisense RNA transcription, and antisense RNA TSS is preferentially observed in relatively long CGIs covering exon 1 and extending into intron 1 (Finocchiaro et al. 2007). Both sense and antisense RNA play a role in gene expression by affecting DNA methylation status (Sleutels et al. 2002; Pickford and Cogoni 2003; Imamura et al. 2004). ChIP-chip experiments on regions adjacent to the TSS have indicated that H3K4 methylation occurs in a 2-kb region downstream from the TSS and that the peaks in the region are located 1 kb downstream (Barski et al. 2007). In addition, histone modification has been reported to affect DNA methylation in a locus-specific manner (Ikegami et al. 2007). These data suggest that T-DMRs located

in the 3′-region downstream from the TSS are involved in gene regulation.

In this study, we demonstrated T-DMR plurality to be involved in tissue-specific gene expression. DNA methylation regulates not only in gene expression, but also in other gene functions; therefore, T-DMRs identified by D-REAM could provide investigative insight into the roles of genome-wide DNA methylation. We conclude that T-DMR profiles are tissue specific and facilitate tissue identification by reflecting tissue-specific gene functions.

## Methods

### Mice and genomic DNA extraction

Male mice (C57BL/6NCrj, 12- to 13-wk-old mice for liver, cerebrum, and kidney; 6-wk-old for spleen) were euthanized after fasting for 16 h. Tissue samples were collected and frozen at −80°C until use. The samples (<20 mg) were thawed, homogenized, and incubated with 300 µL of lysis solution (10 mM Tris-HCl at pH 8.0, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, and 200 µg/mL proteinase K) at 55°C for 30 min. Samples were extracted with a phenol/chloroform/isoamyl alcohol (PCI) mixture, incubated with RNase for 30 min, and re-extracted with PCI. DNA was precipitated with ethanol and dissolved in 20 µL of Tris-EDTA (TE) buffer (pH 8.0).

### Combined bisulfite restriction analysis (COBRA) and bisulfite sequencing

Genomic DNA was digested with PstI. Digested DNA (3 µg) was denatured with 0.3 N NaOH. Sodium metabisulfite (pH 5.0) and hydroquinone were added to a final concentration of 2.0 M and 0.5 mM, respectively. The reaction mixture was incubated in the dark at 55°C for 16 h. DNA was purified using the Wizard DNA Clean-up System (Promega KK), treated with 0.3 M NaOH at 37°C for 15 min, and precipitated with ethanol. It was then dissolved in 20 µL of TE buffer (pH 8.0) and used in a concentration range of 1/100 to 1/20 for PCR analysis with Immolase Taq DNA polymerase (Bioline). During the bisulfite reaction, unmethylated CpGs are converted to TpGs, while methylated CpGs remain intact. For restriction mapping, 10% of the PCR product was digested with HpyCH4IV at 37°C overnight and electrophoresed with the undigested product (control) on a 1% agarose gel. The CpG methylation status within the HpyCH4IV restriction sites was assessed according to the proportion of cleaved fragments. For bisulfite sequencing, 50% of the PCR product was gel-extracted and subcloned into the pGEM-T easy vector (Promega KK). A minimum of 10 clones was sequenced, and the methylation status of individual CpGs was determined.

### D-REAM

Genomic DNA (5 µg) was digested with HpyCH4IV (New England BioLabs) overnight. The digestion was monitored by gel electrophoresis. Digested DNA was recovered by ethanol precipitation following extraction with PCI and chloroform, and was dissolved in TE buffer (pH 8.0). Fifty nanograms of the DNA sample were ligated to the R-adaptor pair (Supplemental Table S2) using T4 DNA ligase (New England BioLabs). Following treatment with the Klenow fragment, the DNA was digested with TaqI at 65°C for at least 1 h and purified using a Microspin S-300 HR column (GE Healthcare UK Ltd.). DNA samples were then ligated to the N-adaptor pair (Supplemental Table S2) and purified using the Wizard SV Gel and PCR Clean-up System (Promega KK). PCR was performed using Immolase Taq DNA polymerase and the R18 and N18 primers in the presence of dUTP under the following conditions: denaturation at 95°C for 7 min and 20 cycles, each cycle comprising 95°C for 30 sec, 62°C for 30 sec, and 72°C for 2 min. A total of 10 µg of amplified DNA was used for microarray analysis. When NotI was used as the first methylation sensitive restriction enzyme, we used the R-adaptor pair for NotI instead of that for HpyCH4IV.

Microarray analysis was conducted using the GeneChip System (Affymetrix), and all procedures were performed according to the Affymetrix chromatin immunoprecipitation assay protocol provided by the manufacturer. DNA samples were labeled using the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix) and hybridized with Affymetrix GeneChip mouse promoter 1.0R arrays. The arrays were stained and washed with GeneChip Fluidics Station 450 and scanned with the GeneChip 3000 7G Scanner to obtain a ".CEL" file describing the probe intensities. The instruments were operated using the GeneChip operating software version 1.4.

### Bioinformatics

Data flow is summarized in Supplemental Fig. S6. To satisfy gene ID requirements of the bioinformatics analysis, we converted gene IDs under certain circumstances. MAT (Johnson et al. 2006) (bandwidth, 300 bp) was used to analyze the tiling array .CEL files and identify the hypomethylated regions based on tiling probe signals, probe sequences, and copy numbers. xMAN (Li et al. 2008) was used to remap the original tiling probes according to the mouse genome assembly of version mm8 (March 2006 build) from the UCSC genome database (Kuhn et al. 2007). A separate ".bpmap" file, containing a subset of probes for the HpyCH4IV–HpyCH4IV and HpyCH4IV–TaqI fragments, was used to verify the MAT analysis. The data were visualized using the Integrated Genome Browser (http://www.affymetrix.com/support/-developer/tools/download_igb-.affx).

Statistical analysis was performed using the R software package and BioConductor package (Gentleman et al. 2004). The tiling array package in BioConductor was used to examine the reproducibility of the microarray data. MultiExperiment Viewer (MeV in TM4 Microarray Software Suite) (http://www.tm4.org/mev.html) was used for K-means clustering of MATscores (Saeed et al. 2003). Genomic annotations, including Ensembl gene assignments (Birney et al. 2004), were obtained from the Galaxy website (http://g2.bx.psu.edu; Giardine et al. 2005). Transcriptome data were obtained from the GNF SymAtlas website (http://symatlas.gnf.org/SymAtlas/; Su et al. 2002), and annotation and ontology analyses were conducted using g:profiler (http://biit.cs.ut.ee/gprofiler/; Reimand et al. 2007), DAVID 2007 (http://niaid.abcc.ncifcrf.gov/; Huang da et al. 2007), and KEGG pathway database (http://www.genome.jp/kegg/kegg2.html; Kanehisa et al. 2006). EMBOSS (Rice et al. 2000) was applied for DNA sequence analysis, and the BIQ analyzer (Bock et al. 2005) was used to analyze the bisulfite sequencing data. Mouse gene symbols were confirmed by referring to the MGI database (http://www.informatics.jax.org/). Transcription factor motifs 1 kb upstream of TSS were analyzed on the MAPPER database website (http://bio.chip.org/mapper/; Marinescu et al. 2005).

## Acknowledgments

We thank Dr. Bruce Murphy (University of Montreal) for critically reviewing the manuscript. Our work was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN), Japan, and a Grant-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan 20062003 (S.T.) and 15080202 (K.S.)

## References


Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823–837.

Bird, A. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8:** 1499–1504.

Bird, A., Taggart, M., Frommer, M., Miller, O., and Macleod, D. 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40:** 91–99.

Birney, E., Andrews, T., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., et al. 2004. An overview of Ensembl. *Genome Res.* **14:** 925–928.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. 2005. BiQ Analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21:** 4067–4068.

Brocard, J., Kastner, P., and Chambon, P. 1996. Two novel RXR alpha isoforms from mouse testis. *Biochem. Biophys. Res. Commun.* **229:** 211–218.

Ching, T.T., Maunakea, A.K., Jun, P., Hong, C., Zardo, G., Pinkel, D., Albertson, D.G., Fridlyand, J., Mao, J.H., Shchors, K., et al. 2005. Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. *Nat. Genet.* **37:** 645–651.

Cho, J., Kimura, H., Minami, T., Ohgane, J., Hattori, N., Tanaka, S., and Shiota, K. 2001. DNA methylation regulates placental lactogen I gene expression. *Endocrinology* **142:** 3389–3396.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452:** 215–219.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V., Attwood, J., Burger, M., Burton, J., Cox, T., Davies, R., Down, T., et al. 2006. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38:** 1378–1385.

Finocchiaro, G., Carro, M., Francois, S., Parise, P., DiNinni, V., and Muller, H. 2007. Localizing hotspots of antisense transcription. *Nucleic Acids Res.* **35:** 1488–1500.

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

Geier, A., Wagner, M., Dietrich, C.G., and Trauner, M. 2007. Principles of hepatic organic anion transporter regulation during cholestasis, inflammation and liver regeneration. *Biochim. Biophys. Acta* **1773:** 283–308.

Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5:** R80. doi: 10.1186/gb-2004-5-10-r80.

Giardine, B., Riemer, C., Hardison, R., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **15:** 1451–1455.

Giguere, V. 1999. Orphan nuclear receptors: From gene to function. *Endocr. Rev.* **20:** 689–725.

Handschin, C. and Meyer, U.A. 2005. Regulatory network of lipid-sensing nuclear receptors: Roles for CAR, PXR, LXR, and FXR. *Arch. Biochem. Biophys.* **433:** 387–396.

Hatada, I., Fukasawa, M., Kimura, M., Morita, S., Yamada, K., Yoshikawa, T., Yamanaka, S., Endo, C., Sakurada, A., Sato, M., et al. 2006. Genome-wide profiling of promoter methylation in human. *Oncogene* **25:** 3059–3064.

Hattori, N., Abe, T., Hattori, N., Suzuki, M., Matsuyama, T., Yoshida, S., Li, E., and Shiota, K. 2004a. Preference of DNA methyltransferases for CpG islands in mouse embryonic stem cells. *Genome Res.* **14:** 1733–1740.

Hattori, N., Nishino, K., Ko, Y., Hattori, N., Ohgane, J., Tanaka, S., and Shiota, K. 2004b. Epigenetic control of mouse *Oct-4* gene expression in embryonic stem cells and trophoblast stem cells. *J. Biol. Chem.* **279:** 17063–17069.

Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., Tanaka, S., and Shiota, K. 2007. Epigenetic regulation of *Nanog* gene in embryonic stem and trophoblast stem cells. *Genes Cells* **12:** 387–396.

Huang da, W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., et al. 2007. DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35:** W169–W175.

Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., Greally, J., Yagi, S., Hattori, N., and Shiota, K. 2007. Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells* **12:** 1–11.

Imamura, T., Ohgane, J., Ito, S., Ogawa, T., Hattori, N., Tanaka, S., and Shiota, K. 2001. CpG island of rat sphingosine kinase-1 gene: Tissue-dependent DNA methylation status and multiple alternative first exons. *Genomics* **76:** 117–125.

Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S., and Shiota, K. 2004. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem. Biophys. Res. Commun.* **322:** 593–600.

Johnson, W., Li, W., Meyer, C., Gottardo, R., Carroll, J., Brown, M., and Liu, X. 2006. Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci.* **103:** 12457–12462.

Jones, P. 2002. DNA methylation and cancer. *Oncogene* **21:** 5358–5360.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. 2006. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **34:** D354–D357.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R., Niveleau, A., Cedar, H., et al. 2006. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38:** 149–153.

Khulan, B., Thompson, R., Ye, K., Fazzari, M., Suzuki, M., Stasiek, E., Figueroa, M., Glass, J., Chen, Q., Montagna, C., et al. 2006. Comparative isoschizomer profiling of cytosine methylation: The HELP assay. *Genome Res.* **16:** 1046–1055.

Ktistaki, E. and Talianidis, I. 1997. Modulation of hepatic gene expression by hepatocyte nuclear factor 1. *Science* **277:** 109–112.

Kuhn, R., Karolchik, D., Zweig, A., Trumbower, H., Thomas, D., Thakkapallayil, A., Sugnet, C., Stanke, M., Smith, K., Siepel, A., et al. 2007. The UCSC genome browser database: Update 2007. *Nucleic Acids Res.* **35:** D668–D673.

Li, W., Carroll, J., Brown, M., and Liu, X. 2008. xMAN: Extreme mapping of oligonucleotides. *BMC Genomics* **9:** S20. doi: 10.1186/1471-2164-9-S1-S20.

Lieb, J., Beck, S., Bulyk, M., Farnham, P., Hattori, N., Henikoff, S., Liu, X., Okumura, K., Shiota, K., Ushijima, T., et al. 2006. Applying whole-genome studies of epigenetic regulation to study human disease. *Cytogenet. Genome Res.* **114:** 1–15.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133:** 523–536.

Marinescu, V.D., Kohane, I.S., and Riva, A. 2005. The MAPPER database: A multi-genome catalog of putative transcription factor binding sites. *Nucleic Acids Res.* **33:** D91–D97.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454:** 766–770.

Nishino, K., Hattori, N., Tanaka, S., and Shiota, K. 2004. DNA methylation-mediated control of *Sry* gene expression in mouse gonadal development. *J. Biol. Chem.* **279:** 22306–22313.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. 2004. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303:** 1378–1381.

Ohgane, J., Aikawa, J., Ogura, A., Hattori, N., Ogawa, T., and Shiota, K. 1998. Analysis of CpG islands of trophoblast giant cells by restriction landmark genomic scanning. *Dev. Genet.* **22:** 132–140.

Ordway, J., Bedell, J., Citek, R., Nunberg, A., Garrido, A., Kendall, R., Stevens, J., Cao, D., Doerge, R., Korshunova, Y., et al. 2006. Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* **27:** 2409–2423.

Pickford, A.S. and Cogoni, C. 2003. RNA-mediated gene silencing. *Cell. Mol. Life Sci.* **60:** 871–882.

Rauch, T., Li, H., Wu, X., and Pfeifer, G. 2006. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res.* **66:** 7939–7947.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. 2007. g:Profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35:** W193–W200.

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16:** 276–277.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., et al. 2003. TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **34:** 374–378.

Sakamoto, H., Suzuki, M., Abe, T., Hosoyama, T., Himeno, E., Tanaka,

S., Greally, J.M., Hattori, N., Yagi, S., and Shiota, K. 2007. Cell-type specific methylation profiles occuring disproportionately in CpG-less regions that delineate developmetal similarity. *Genes Cells* **12:** 1123–1132.

Schrem, H., Klempnauer, J., and Borlak, J. 2002. Liver-enriched transcription factors in liver function and development. Part I: The hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol. Rev.* **54:** 129–158.

Shen, C. and Maniatis, T. 1980. Tissue-specific DNA methylation in a cluster of rabbit beta-like globin genes. *Proc. Natl. Acad. Sci.* **77:** 6634–6638.

Shen, Y., Chow, J., Wang, Z., and Fan, G. 2006. Abnormal CpG island methylation occurs during in vitro differentiation of human embryonic stem cells. *Hum. Mol. Genet.* **15:** 2623–2635.

Shiota, K. 2004. DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. *Cytogenet. Genome Res.* **105:** 325–334.

Shiota, K., Kogo, Y., Ohgane, J., Imamura, T., Urano, A., Nishino, K., Tanaka, S., and Hattori, N. 2002. Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells* **7:** 961–969.

Sleutels, F., Zwart, R., and Barlow, D.P. 2002. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415:** 810–813.

Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., and Feinberg, A.P. 2002. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res.* **12:** 543–554.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci.* **99:** 4465–4470.

Ushijima, T. 2005. Detection and interpretation of altered methylation patterns in cancer cells. *Nat. Rev. Cancer* **5:** 223–231.

Weber, M., Davies, J., Wittig, D., Oakeley, E., Haase, M., Lam, W., and Schübeler, 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.

Weber, M., Hellmann, I., Stadler, M., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39:** 457–466.

Williams, K.T. and Schalinske, K.L. 2007. New insights into the regulation of methyl group and homocysteine metabolism. *J. Nutr.* **137:** 311–314.

Xiong, Z. and Laird, P.W. 1997. COBRA: A sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25:** 2532–2534.

Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. 2005. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* **350:** 129–136.

*Received November 8, 2007; accepted in revised form August 13, 2008.*

Molecular
Gastrointestinal
Medicine



# 再生医療のための エピジェネティクスとエピゲノム

前田千晶* 塩田邦郎*

KEY WORDS

エピジェネティクス，エピゲノム，DNA メチル化，ヒストン修飾，リプログラミング，幹細胞，iPS 細胞，再生医療

SUMMARY

再生医療に用いられる細胞として，組織幹細胞，胚性幹細胞（embryonic stem cell：ES 細胞）につづいて，新たに人工多能性幹細胞（induced pluripotent stem cell：iPS 細胞）が作出された．少数遺伝子導入により多分化能が獲得できることが判明し，"細胞とは？細胞の評価法は？細胞の安定性とフレキシビリティーとは？" などが改めて問われている．エピジェネティクス制御系は，DNA 塩基配列は一定のままで，遺伝子発現を固定・記憶する機構である．"初期化" や "リプログラミング" の分子機構にはエピジェネティクスが深くかかわっている．ゲノム全体のエピジェネティクス情報をエピゲノムという．本稿では，医療への応用が期待される細胞や技術におけるエピジェネティクスとエピゲノム研究の重要性と緊急性を記す．

## はじめに

幹細胞は自らを複製，再生する能力（自己複製能）と，さまざまな細胞へと分化する能力（多分化能）を有する細胞である．幹細胞のうち最も多くの細胞に変わることができる細胞が胚性幹細胞（embryonic stem cell：ES 細胞）である．ES 細胞は胚発生初期の着床前の受精卵，胚盤胞の細胞（内部細胞塊）よりつくり出される．一方，組織幹細胞（成体幹細胞，体性幹細胞）は，分化した組織中に存在する未分化な細胞（群）で，通常，すべての種類の細胞には分化できないが，特定系列の複数種類の細胞へ分化が可能な幹細胞を指す．間葉系幹細胞，造血幹細胞，神経幹細胞などが含まれ，成体組織の他，臍帯など多数の組織から発見されている．近年，間葉系幹細胞が，骨組織，脂肪，骨格筋の他，神経細胞，肝細胞，インスリン分泌細胞など，多様な細胞系列に分化転換しうることが報告されてきた．さらに，2006 年に，マウス線維芽細胞から人工多能性幹細胞（induced pluripotent stem cell：iPS 細胞）が作製され[1]，翌年にはヒト iPS 細胞が作出されたことで[2,3]，再生医療は実現化に向けて大

* MAEDA Chiaki, SHIOTA Kunio/東京大学大学院農学生命科学研究科細胞生化学研究室

きく前進した.

iPS細胞, ES細胞を実際に応用するときに考慮するべきことは, どちらも特定の培養条件下で選択することにより人工的に創出された細胞だという点である. そのため, これらの標準細胞を生体内に求めることはできない. いったん分化した体細胞が多能性を再獲得するメカニズムと, 人工細胞をどのように評価するかに関心が集まっている.

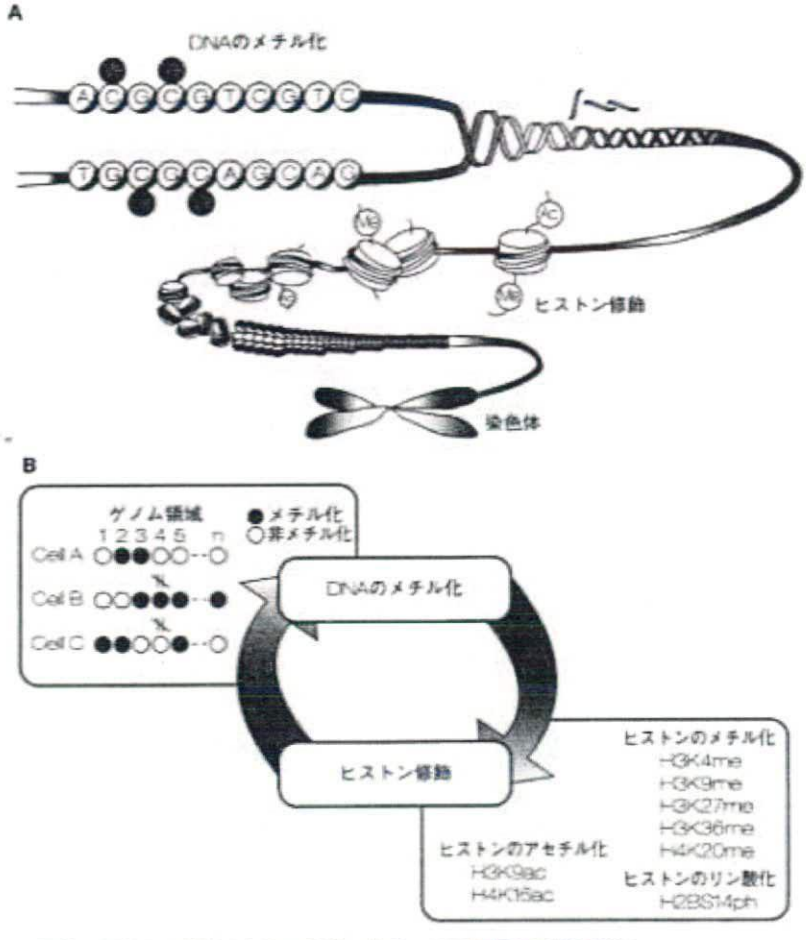
## エピジェネティクスとエピゲノムとは

哺乳類の細胞には, 約$3\times10^9$塩基対からなるDNAが含まれており, この配列情報は一部の例外を除いてすべての細胞で共通である. 一方, 細胞はその種類や分化段階に応じて発現する遺伝子を使い分けている. 分化した細胞では, 遺伝子の使い分け機構は細胞分裂をくり返しても維持される. エピジェネティクスとは, 膨大なDNAの塩基配列のカタログから, 必要な情報を選択して利用し, 記憶する機構といえる. エピジェネティクス制御系には, DNAのメチル化, ヒストン修飾(メチル化, アセチル化など), ヒストンバリアントの使い分け, non-coding RNAなどが含まれる(**図❶**). DNAのメチル化は, おもにシトシン, グアニンと連続するCpG配列のシトシンにメチル基が付加される現象を指す. DNAメチル化とヒストン修飾は互いに密接に関連しており, 個体の正常な発生に必須である. ゲノム全域におけるDNAメチル化やヒストン修飾などのエピジェネティック情報の総体をエピゲノムとよぶ.

移植治療のために特定の細胞集団を作出した場合, 得られた細胞が目的の細胞の形質を示し, それ以外の細胞の形質は示さないことをどのように担保するべきであろうか. 細胞の形態や遺伝子発現, および*in vitro*での分化能に加え, エピジェネティクス評価は威力を発揮する. ゲノム上には, 細胞の種類によってDNAメチル化状態の異なる領域, つまり組織・細胞種特異的なメチル化領域(tissue-dependent and differentially methylated region:T-DMR)が多数存在し[4][5], その細胞のエピゲノムを特徴づけている. 複数のT-DMRのメチル化・非メチル化のパターンの組み合わせを細胞の"DNAメチル化プロフィール"とよぶ. DNAメチル化プロフィールは, 細胞の種類ごとに異なる安定したゲノム上の情報である. このため, DNAメチル化プロフィールを細胞の同定に利用できると考えられる.

## エピゲノムはどのように形成されるのか

シトシンへのメチル基転移反応は, DNAメチル基転移酵素(DNA methyltransferase:Dnmt)によって触媒される. ところが, メチル化される領域とされない領域といったDNAメチル化の領域選択性は, Dnmtの酵素活性や発現量のみで決定されるのではない. なぜなら, Dnmtはいずれも DNAに対する配列特異性をもたず, *in vitro*ではすべてのCpG配列に作用するからだ. 領域選択的なDNAメチル化機構には, ヒストン修飾酵素が関係している. クロマチンが弛緩したユークロマチン領域に作用するヒストンメチル基転移酵素であるG9aは, ヒストンH3K9およびH3K27のメチル化を触媒する[6][7]. G9a欠損ES細胞においてゲノムワイドなDNAメチル化解析をおこなったところ, 解析可能な約1,300座位のうち, 32座位でDNAの脱メチル化が検出された[8]. これらの領域ではヒストンH3K9あるいはH3K27の脱メチル化も観察され, 確かにG9aの標的領域であると考えられた. G9a自体はDNAメチル基転移活性をもたないことから, このようなG9aの標的領域の一部では, ヒストン修飾状況がDNAメチル化を誘導するような環境をつくっていると考えられる. また, G9aがDnmt1と直接相互作用すること[9]や, 凝集したクロマチンに結合するヘテロクロマチン蛋白質HP1とDnmt1が協調してはたらくことも報告されてきた[10]. このように, DNAメチル化とヒストン修飾は相互作用を通して各領域のクロマチン構造を形成している(**図❶**). そこには, HP1をはじめとしたクロマチン結合因子や, メチル化されたDNAに結合するMBDファミリー蛋白質(MeCP2, MBD1-MBD4), RNA分子なども関与する.




**図❶ 細胞のエピゲノムとエピジェネティクス機構間の相互依存**

A）エピジェネティクス制御系には、DNAのメチル化、ヒストン修飾（メチル化、アセチル化など）、ヒストンバリアントの使い分け、non-coding RNAなどの多数の機構が含まれる。ゲノム領域ごとに、付加されているエピジェネティック修飾の組み合わせは異なっており、全体として細胞の種類に特異的なエピゲノムをつくりあげている。

B）DNAのメチル化とヒストン修飾は、相互作用しながらエピゲノム形成にかかわっている。

## リプログラミングとはエピゲノムの変化である

iPS細胞作製や核移植クローニングで盛んに用いられてきた、"初期化"あるいは"リプログラミング"とはエピゲノムの変化のことである。多くの場合は定義があいまいなまま、うまく細胞が変化し発生が進行したことをいいかえて、"初期化"あるいは"リプログラミング"がうまくいったといいかえているにすぎない。しかし、細胞分化の基盤としてエピジェネティクス制御系があることは間違いない。細胞の種類によってエピゲノムは異なることから、"初期化"あるいは"リプログラミング"をゲノムレベルで定義すれば、"初期胚や幹細胞と同等のエピゲノムへの書き換え"ということになる。

骨髄由来の間葉系幹細胞（mesenchymal stem cell：MSC）および脂肪組織由来の幹細胞（adipose tissue-derived stem cell：ASC）にDNAメチル化阻害剤である5-アザシチジンを添加すると、筋細胞、心筋細胞への分

化が促進される[11][12]．一方，MSCに，ヒストン脱アセチル化阻害剤のトリコスタチンA（TSA）を添加し，高グルコース下で培養すると，膵臓β細胞のマーカー遺伝子を発現するようになる[13]．また，FGF-4を含む複数の因子とTSAとの組み合わせで，ヒトMSCから肝細胞様の細胞もつくられている[14]．神経幹細胞においては，ヒストン脱アセチル化阻害剤のバルプロ酸（VPA）が，神経細胞への分化を誘導する[15]．これらの薬剤の作用機序は完全には明らかになっていないが，エピゲノムの変化が組織特異的な遺伝子発現を調節し，ひいては細胞系譜の運命決定にかかわっていると考えられる．

クローン技術を応用してつくられた幹細胞として，核移植胚由来のES細胞（nuclear transfer embryonic stem cell：ntES細胞）がある．ntES細胞は，体細胞から核移植をおこなってクローン胚盤胞を作製し，その内部細胞塊を培養して樹立されたES細胞である[16]．マウスntES細胞は種々の解析において自然交配の胚盤胞から樹立されたES細胞と区別がつかず[17][18]，ntES細胞から分化誘導した血球系細胞の移植が治療効果をあげている[19]．興味深いことに，クローン胚盤胞をあらかじめTSAで処理してからntES細胞株を樹立すると，処理しない場合にくらべて約2～3倍効率が上昇する[20]．また，核移植した卵をTSA処理することで，クローン胚盤胞の形成率，およびクローン個体の出生率も上昇する[20][21]．これらの結果は，エピジェネティクスが個体発生の基礎であるとする考えと矛盾しない．

iPS細胞は，体細胞に2～4つの遺伝子を導入することにより，幹細胞の形質を獲得させた細胞である．形態や分化能はES細胞に類似し，網羅的な遺伝子発現解析においてもES細胞と強い相関を示す．また，その樹立にヒトの胚を利用する必要がなく，患者本人の細胞をもとにつくることができるという利点をもつ．第2世代のマウスiPS細胞（Nanog-iPS）およびヒトiPS細胞では，*Oct3/4*遺伝子をはじめとするいくつかのES細胞マーカー遺伝子が，ES細胞と同様に低メチル化状態であることが確認された．また，一部のヒストン修飾についてもES細胞と近いパターンを示すことが報告されている[22]．

線維芽細胞からのiPS細胞の樹立率は，マウスでもヒトでも0.001～0.5%ときわめて低い[2][23]．多くはES細胞とは似つかない顆粒状のコロニーを形成し，多分化能は獲得しない．iPS細胞樹立の効率化に，DNAメチル化阻害剤やヒストン脱アセチル化阻害剤は有効であるのか，種々の化合物の影響が調べられた[24]．*Oct3/4*，*Klf4*，*Sox2*，*c-Myc*の4つの遺伝子を導入した線維芽細胞において，5-アザシチジン添加は単独で10倍，またデキサメタゾンとの組み合わせにより，約26倍ものiPS出現効率の上昇をもたらした．また，いずれもヒストン脱アセチル化阻害作用をもつスベロイラニリド・ハイドロザミック酸（SAHA），TSA，およびVPAのなかでは，VPAの効果がとびぬけて高く，未処理にくらべて100倍以上のiPS細胞が得られた．さらに，VPA処理は，癌遺伝子である*c-Myc*を除く3遺伝子導入の線維芽細胞においても，約50倍の効率上昇効果を示した．遺伝子導入していない通常の線維芽細胞へのVPA添加実験から，VPA添加によって，線維芽細胞の網羅的な遺伝子発現のパターンがES細胞のそれに近づくことが示されている[24]．

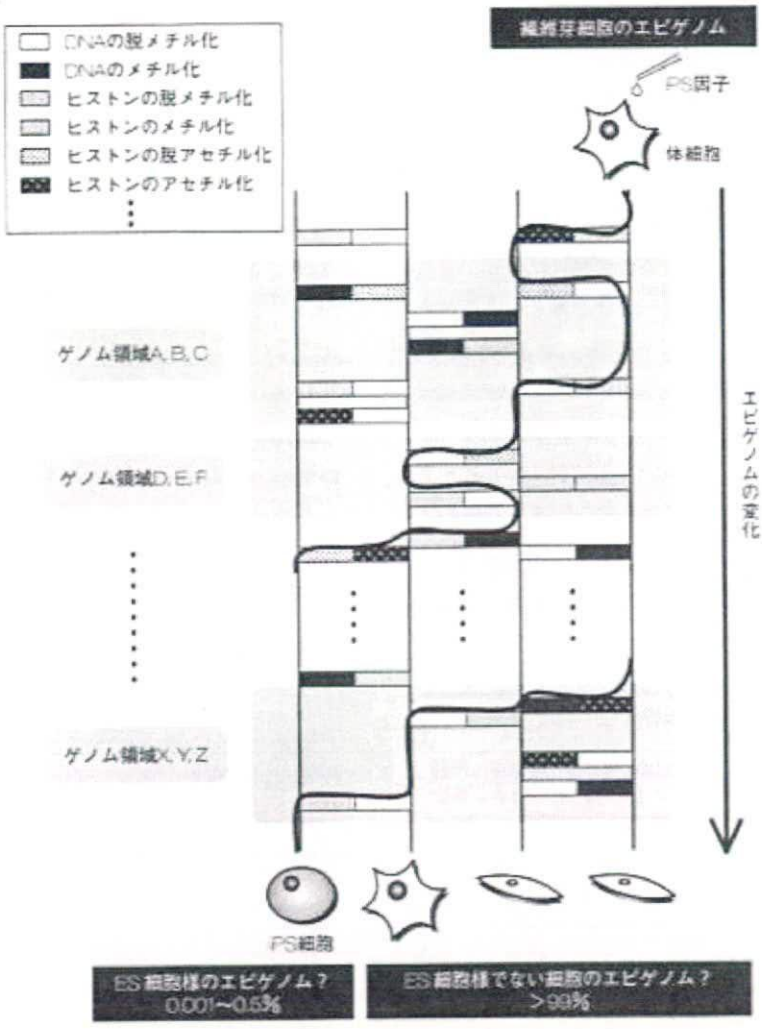
ntES細胞，iPS細胞において注意すべきなのは，ゲノム全体をヒストン高アセチル化，あるいはDNA脱メチル化することが，分化多能性の獲得に有利であるとは考えにくい点だ．先に記したように，エピゲノムは細胞ごとに異なり，ES細胞であってもゲノム全域のDNAメチル化状況は決して非メチル化ではない．ES細胞は他の細胞とくらべて異なったDNAメチル化プロフィールを有しているのであって，未分化＝DNA低メチル化ではないといえる[4]．そして，細胞分化やiPS化の過程には，必ず細胞選択過程が含まれる．DNAメチル化やヒストン修飾に影響を与える化合物処理は，ゲノムにエピジェネティクスショックを与えているとすれば理解しやすい．これらの化合物でエピジェネティクス変化が起きた場合，各細胞は生存をかけてエピゲノムをもとに戻すか，あるいは，さらに変化させ，もととは異なった別の状態に落ち着くはずである．どちらにも失敗した細胞は死滅する．薬剤処理をきっかけに引き起こされるエピゲノムの複合的な状況変化が，細胞の選択に偏りを生じさせ，分化多能性細胞の樹立促進を可能にしたと考えるべきであろう．

重要なことは，細胞が他の細胞に変化するとき，そこ

には mRNA や蛋白質の発現変化のみならず，そのもととなるエピゲノムの変化があるということである．体細胞が ES 細胞様の形質を獲得するためには，ゲノム領域ごとにエピジェネティック修飾が書き換えられ，ES 細胞のエピゲノムに近い状態に変化する必要があると考えられる（図❷）．

## エピゲノム解析技術

ゲノムワイドな DNA メチル化状況，およびヒストン修飾状況を明らかにするため，網羅的な解析法の開発が進んでいる．これらはおもにマイクロアレイを使うものと使わないものに大別できる．マイクロアレイでは，高密度プローブ実装技術の進歩に伴い，ゲノム配列をそのままタイリングした高解像度のタイリングアレイが使われるようになった．マイクロアレイに供するサンプルの




**図❷ リプログラミングとはエピゲノムの変化である**
細胞が新たな細胞に変化するとき，そこにはゲノム領域ごとに異なるエピジェネティック修飾の書き換えが起きている．線維芽細胞のエピゲノムが ES 細胞様のエピゲノムに変化することで，iPS 細胞がつくられると考えられる．

調整法としては，DNA メチル化抗体や特定のヒストン修飾の抗体で免疫沈降したゲノム分画を用いる me-DIP 法や ChIP-Chip 法の他，DNA メチル化感受性制限酵素 *Hpa* Ⅱと非感受性制限酵素 *Msp* Ⅰで切断したゲノムを用いる HELP 法，MIAMI 法などがある．

筆者らのグループは最近，新たなゲノムワイド DNA メチル化解析法として，D-REAM 法を確立した[25]．メチル化感受性制限酵素の認識部位のメチル化状況をタイリングアレイのシグナルに反映させることができる系であり，原理的にはメチル化感受性制限酵素の種類に依存しない解析が可能である．一方，高速シークエンサーの登場により，ハイスループットシークエンスが可能となり，ゲノムワイドな解析において勢いを増している．ゲノムサイズが比較的小さいシロイナズナでは，ゲノムの大部分の DNA メチル化状態が明らかにされた[26]．また，マウスにおいても，分化・未分化細胞の網羅的 DNA メチル化解析などに応用されている[27]．

このようなエピゲノム解読技術の発展に伴って，現在世界ではヒトのエピゲノム解明へ向けた動きが加速している．ヨーロッパを中心に進行中の Human Epigenome Project（HEP）や Epigenome Network of Excellence（NoE）に加え，アメリカの先導による Alliance for the Human Epigenome and Disease（AHEAD）がエピゲノムデータベースの基盤づくりに国際的な協調をよびかけている[28]．このプロジェクトでは，統一された抗体や手法を用い，複数種類の組織・細胞についてエピジェネティック情報を網羅的に記述したエピゲノムマップをつくっていこうとするものである．

## おわりに

幹細胞や前駆細胞を用いた再生医療は，現実のものとなりつつある．骨髄由来の幹細胞を用いたクローン病の治療は，臨床試験第Ⅲ相にある他，骨や心臓再生をめざした組織幹細胞治療も臨床試験第Ⅱ相に進んでいる．こうした進展に加え，今後は iPS 細胞の応用研究も新領域を切り拓くことになるだろう．今年 7 月，筋萎縮性側索硬化症（amyotrophic lateral sclerosis：ALS）の患者の皮膚細胞から樹立された iPS 細胞が，病気の進行で失われる運動神経細胞に分化することができたと報告された[29]．得られた細胞において，形態および神経細胞マーカーの発現は確認された．つぎのステップとしては，エピゲノムの正常性を調べる必要があるだろう．しかし，生体から分離した運動神経細胞と，培養下で樹立したそれ様のもので，ゲノム全域のエピジェネティック状態が寸分違わず一致することは考えにくい．樹立した細胞が神経細胞の表現型を示すには，神経細胞に特徴的に発現する，あるいは発現しない，遺伝子については，生体細胞と同じエピジェネティック状態を示す必要がある．しかし，それ以外の遺伝子領域について，また，反復配列などを含む遺伝子以外のゲノム領域について，どれほどのエピジェネティックな差が許容されるのか，今後の焦点となる重要なトピックである．幹細胞を評価する段階，幹細胞から目的の細胞を樹立する過程，およびできあがった細胞を評価する段階，のいずれにおいてもエピジェネティクスの観点は必要である．メカニズムの解明と臨床応用は，緊密に連携しながら進められていくべきであろう．

### 文　献


1) Takahashi K, Yamanaka S：Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**：663-676, 2006
2) Takahashi K, Tanabe K, Ohnuki M *et al*：Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**：861-872, 2007
3) Yu J, Vodyanik MA, Smuga-Otto K *et al*：Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**：1917-1920, 2007
4) Shiota K, Kogo Y, Ohgane J *et al*：Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes Cells* **7**：961-969, 2002
5) Lieb JD, Beck S, Bulyk ML *et al*：Applying whole-genome studies of epigenetic regulation to study human disease. *Cytogenet Genome Res* **114**：1-15, 2006
6) Tachibana M, Sugimoto K, Fukushima T *et al*：Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**：25309-25317, 2001

7) Tachibana M, Sugimoto K, Nozaki M *et al* : G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16** : 1779-1791, 2002
8) Ikegami K, Iwatani M, Suzuki M *et al* : Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells* **12** : 1-11, 2007
9) Estève PO, Chin HG, Smallwood A *et al* : Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev* **20** : 3089-3103, 2006
10) Smallwood A, Estève PO, Pradhan S *et al* : Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev* **21** : 1169-1178, 2007
11) Fukuda K : Use of adult marrow mesenchymal stem cells for regeneration of cardiomyocytes. *Bone Marrow Transplant* **32** (suppl 1) : S25-S27, 2003
12) Rangappa S, Fen C, Lee EH *et al* : Transformation of adult mesenchymal stem cells isolated from the fatty tissue into cardiomyocytes. *Ann Thorac Surg* **75** : 775-779, 2003
13) Tayaramma T, Ma B, Rohde M *et al* : Chromatin-remodeling factors allow differentiation of bone marrow cells into insulin-producing cells. *Stem Cells* **24** : 2858-2867, 2006
14) Snykers S, Vanhaecke T, De Becker A *et al* : Chromatin remodeling agent trichostatin A : a key-factor in the hepatic differentiation of human mesenchymal stem cells derived of adult bone marrow. *BMC Dev Biol* **7** : 24, 2007
15) Hsieh J, Nakashima K, Kuwabara T *et al* : Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells. *Proc Natl Acad Sci U S A* **101** : 16659-16664, 2004
16) Wakayama T, Tabar V, Rodriguez I *et al* : Differentiation of embryonic stem cell lines generated from adult somatic cells by nuclear transfer. *Science* **292** : 740-743, 2001
17) Wakayama S, Jakt ML, Suzuki M *et al* : Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. *Stem Cells* **24** : 2023-2033, 2006
18) Brambrink T, Hochedlinger K, Bell G *et al* : ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc Natl Acad Sci U S A* **103** : 933-938, 2006
19) Rideout WM 3rd, Hochedlinger K, Kyba M *et al* : Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109** : 17-27, 2002
20) Kishigami S, Mizutani E, Ohta H *et al* : Significant improvement of mouse cloning technique by treatment with trichostatin A after somatic nuclear transfer. *Biochem Biophys Res Commun* **340** : 183-189, 2006
21) Rybouchkin A, Kato Y, Tsunoda Y : Role of histone acetylation in reprogramming of somatic nuclei following nuclear transfer. *Biol Reprod* **74** : 1083-1089, 2006
22) Maherali N, Sridharan R, Xie W *et al* : Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1** : 55-70, 2007
23) Okita K, Ichisaka T, Yamanaka S : Generation of germline-competent induced pluripotent stem cells. *Nature* **448** : 313-317, 2007
24) Huangfu D, Maehr R, Guo W *et al* : Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol* **26** : 795-797, 2008
25) Yagi S, Hirabayashi K, Sato S *et al* : DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res*, 2008 [Epub ahead of print]
26) Cokus SJ, Feng S, Zhang X *et al* : Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452** : 215-219, 2008
27) Meissner A, Mikkelsen TS, Gu H *et al* : Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454** : 766-770, 2008
28) Moving AHEAD with an international human epigenome project. *Nature* **454** : 711-715, 2008
29) Dimos JT, Rodolfa KT, Niakan KK *et al* : Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321** : 1218-1221, 2008

## まえだ・ちあき

前田千晶　東京大学大学院農学生命科学研究科細胞生化学研究室

現在の研究テーマは脳のエピジェネティクス.
好きな言葉は「現状維持とは, 急速に世界から遅れていくこと」.

## しおた・くにお

塩田邦郎　東京大学大学院農学生命科学研究科細胞生化学研究室

1979年東京大学大学院修了（農学博士）. 武田薬品中央研究所を経て, 1987年より東京大学助教授, 1998年より教授.
専門はエピジェネティクス.
趣味は写真とソフトボール.
愛読書は山本七平.

# 幹細胞をエピジェネティクスで評価する

新井良和
八木慎太郎
塩田邦郎

ES細胞（embryonic stem cell）は，哺乳類の体を構成する生来の200種類の細胞とは異なる，人工的に作出された細胞である．ES細胞同様，人工的に作出されたiPS細胞（induced pluripotent stem cell）を評価するにあたり，比較すべきものの拠り所がないまま，再生医療への期待のみが先行しているのが現状である．再生医療時代に相応しい細胞の評価法を確立しなければ，この先の世界で取り残されてしまう．ゲノム全域のエピジェネティクス解析は，このような細胞評価，標準化の有力な手段となるだろう．

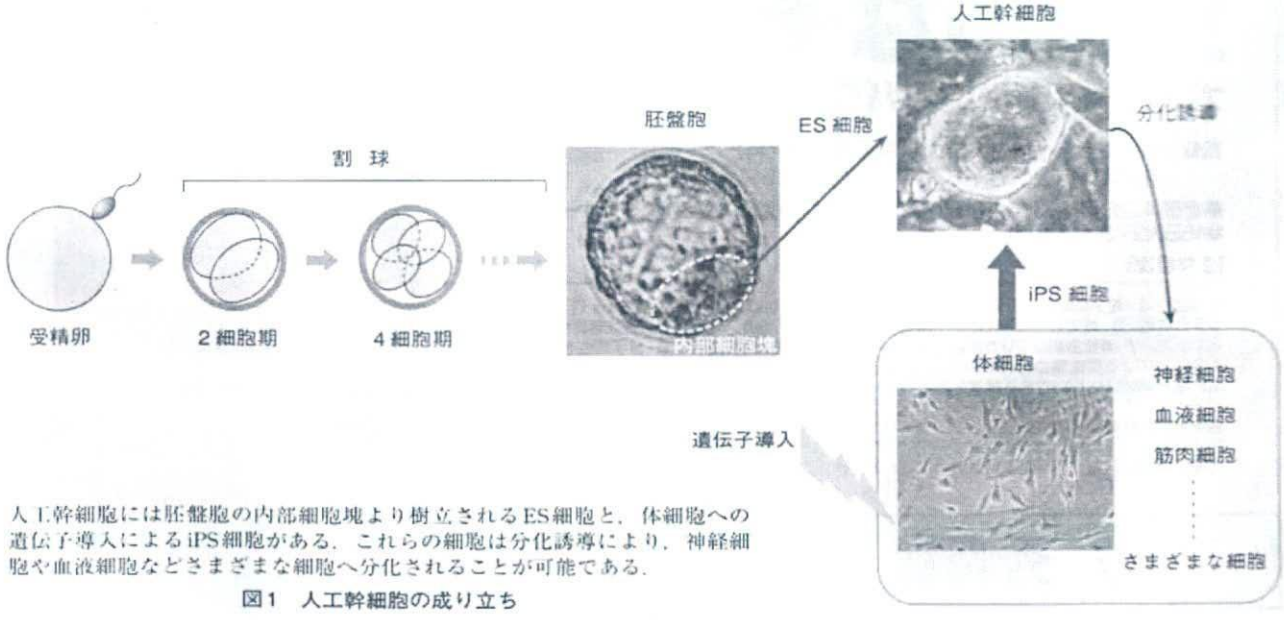
## 細胞評価にもパラダイムシフトが必要

幹細胞は自らを複製，再生する能力（自己複製能）と，別の違ったさまざまな細胞へと変身（分化）する能力（多分化能）をもつ細胞である．幹細胞のうち最も多くの細胞に変わることができる細胞がES細胞であり，胚発生初期の着床前の受精卵，胚盤胞の細胞（内部細胞塊）よりつくり出される（図1）．

カエルやイモリのように高い組織再生能を示す動物と比べて，哺乳類の場合は，いったん分化した細胞は元に戻ることは難しく，分化した細胞から異なる種類の細胞に変わる能力をもつ細胞を作製することは困難であった．しかし完全に分化した細胞の核を卵に導入することにより，頻度は低いがクローン動物をつくることができること，さらにクローン動物からES細胞も樹立できることから，再分化能を人工的に獲得することは不可能ではないことが証明された．

そして2006年に京都大学の山中伸弥らにより世界で初めて，ES細胞と同様にさまざまな細胞へ分化することができるiPS細胞が樹立された（文献1）．iPS細胞は，患者自身の細胞を基に作製することが可能であるため，ES細胞では避けて通れない倫理的問題や免疫拒絶の問題が回避できる．再生医療の夢が現実に大きく近づいた．

iPS細胞，ES細胞を実際に応用するときに，これから考慮するべきことは，iPS細胞はもとより，ES細胞は，特定の培養条件下で選択することにより人工的に創出された細胞であ




人工幹細胞には胚盤胞の内部細胞塊より樹立されるES細胞と，体細胞への遺伝子導入によるiPS細胞がある．これらの細胞は分化誘導により，神経細胞や血液細胞などさまざまな細胞へ分化されることが可能である．

図1　人工幹細胞の成り立ち

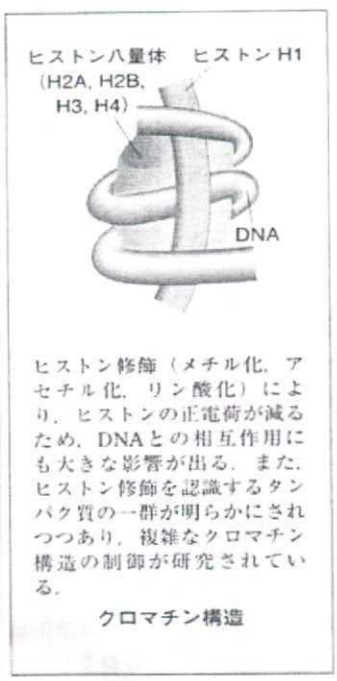
るため，ES細胞の標準細胞を生体内に求めることはできない点である．動物実験に限定して使われていた時代には，細胞の形態と体外培養下での分化能に加え，数種類のマーカー遺伝子の発現および，細胞移植によるテラトーマ（奇形腫）の形成とキメラ形成能などの生体内評価をすれば事足りた．しかし，iPS細胞の創出により，再生医療への応用が現実味を帯びてきた今，動物で検証されてきた方法，たとえばキメラ形成実験などがヒトでは不可能であること，再生医療は個々人を対象とするテーラーメード化された医療になったことを考えると，iPS細胞同様にパラダイムシフトした評価方法が必要である．

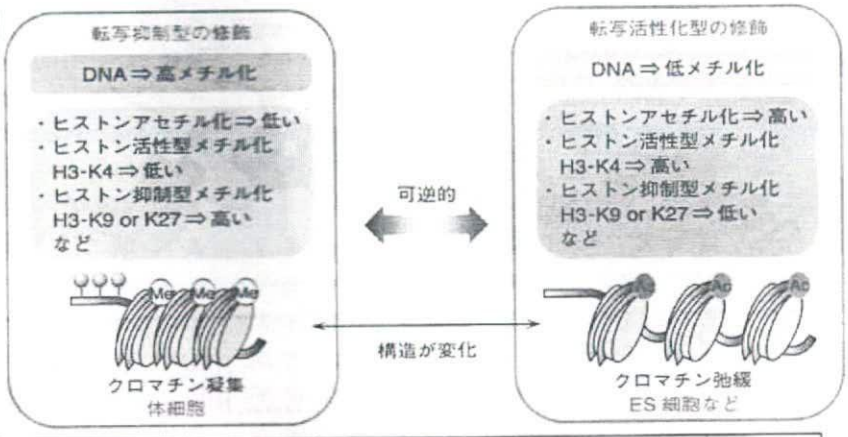
## 幹細胞のマスター遺伝子：*Oct3/4* と *Nanog*

iPS細胞を創出する際の重要な因子である *Oct3/4* や *Nanog* は，ES細胞のマーカー遺伝子として重要であることに加え，ES細胞をES細胞たらしめているマスター遺伝子，つまりES細胞特有の遺伝子の発現量を制御する転写因子であり，これらの因子，さらにこれらによって制御される転写因子がES細胞に特有の遺伝子発現を支えている．たとえば *Oct3/4* 遺伝子を欠損した初期胚（胚盤胞）では，ES細胞の基になる内部細胞塊を形成できず，ES細胞を樹立することもできない（文献2）．また，*Oct3/4* の発現量が適当でないとES細胞は多分化能を維持することができない（文献3），など *Oct3/4* の発現は，ES細胞の樹立と維持に必須である（本誌33ページも参照），そしてこれらのマスター遺伝子の発現を制御している重要な機構はエピジェネティック機構である．

## 幹細胞のエピジェネティクス

エピジェネティクスとは，「DNAの塩基配列の変化を伴わず細胞切裂後も継承される遺伝子機能の変化を研究する学問領域」を意味する．すなわち受精卵から胎児発生を経て個体が誕生するまで，それぞれの細胞に必要な遺伝子発現をオンにし，不必要な遺伝子をオフにする．一方で，いったん分化決定した細胞では，その発生・分化過程で確立した遺伝子のオン・オフ機構は，細胞分裂後も記憶される必要がある．この遺伝子制御とその記憶機構がエピジェネティック機構であ




b

| | | 発現 | DNAメチル化 | ヒストン修飾 アセチル化 | ヒストン修飾 メチル化 H3-K4（活性型） | ヒストン修飾 メチル化 H3-K9（抑制型） | クロマチン構造 |
|---|---|---|---|---|---|---|---|
| Oct3/4 | ES細胞 | ＋ | ▽ | ▲ | ▲ | ▽ | 弛緩 |
| | 体細胞 | － | ▲ | ▽ | ▽ | ▲ | 凝集 |
| Nanog | ES細胞 | ＋ | ▽ | ▲ | ▲ | ▽ | 弛緩 |
| | 体細胞 | － | ▲ | ▽ | ▽ | ▲ | 凝集 |

a）遺伝子発現はDNAメチル化やヒストン修飾状況により制御されている．DNAの高いメチル化状況や抑制型ヒストン修飾（ヒストンH3リシン9あるいは27番目（H3-K9 or 27）のメチル化など）によりクロマチンは凝集し，遺伝子の転写は抑制される．一方，DNAの低いメチル化状況や活性型ヒストン修飾（アセチル化やヒストンH3リシン4番目（H3-K4）のメチル化など）によりクロマチンは弛緩し，遺伝子の転写は活性化される．
b）*Oct3/4* や *Nanog* 遺伝子が発現するES細胞では，クロマチンが弛緩した活性型の修飾がみられる．一方，発現の認められない体細胞では，クロマチンが凝集した抑制型の修飾がみられる．このようなDNAメチル化とヒストン修飾により，遺伝子は細胞種に固有の発現パターンを示すのである．▲は修飾状況が高いこと，反対に▽は低いことを示す．

図2　*Oct3/4,Nanog* 遺伝子のエピジェネティクス制御