Epigenome mapping of human normal purified hepatocytes: personal epigenome variation and genome–epigenome correlation

Eri Arai^{1,2,3}, Fumihito Miura^{3,4}, Yasushi Totoki^{3,5}, Satoshi Yamashita⁶, Ying Tian^{1,3}, Masahiro Gotoh^{2,3}, Hidenori Ojima^{1,3}, Hiroyuki Nakagawa⁷, Yoriko Takahashi⁷, Hiromi Nakamura⁵, Natsuko Hama⁵, Mamoru Kato⁸, Hiroshi Kimura^{3,9}, Yutaka Suzuki^{3,10}, Takashi Ito^{3,4}, Tatsuhiko Shibata^{3,5,11} & Yae Kanai^{*,1,2,3}

¹Department of Pathology, Keio University School of Medicine, Tokyo 160-8582, Japan

²Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo 104-0045, Japan

³AMED-CREST, Japan Agency for Medical Research & Development, Tokyo 100-0004, Japan

⁴Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka 812-8582, Japan

⁵Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo 104-0045, Japan

⁶Division of Epigenomics, National Cancer Center Research Institute, Tokyo 104-0045, Japan

⁷Biomedical Department, Cloud Service Division, IT Infrastructure Services Unit, Mitsui Knowledge Industry Co., Ltd, Tokyo 105-6215, Japan

⁸Department of Bioinformatics, National Cancer Center Research Institute, Tokyo 104-0045, Japan

⁹Graduate School of Bioscience & Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan

¹⁰Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan

¹¹Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Sciences, The University of Tokyo, Tokyo 108-8639, Japan

*Author for correspondence: ykanai@keio.jp

Aim: The aim of this study was to reveal the epigenome landscape of human normal hepatocytes. **Materials & methods:** Cells purified from partial hepatectomy specimens of Japanese patients were subjected to whole-genome bisulfite sequencing using postbisulfite adaptor tagging, chromatin immunoprecipitation sequencing, RNA sequencing and whole-genome sequencing. **Results:** CHG and CHH methylations were inversely associated with gene expression. Histone modification profiles of personal differentially methylated regions (pDMRs) differed considerably among samples. pDMRs were observed around the transcription start sites of genes whose expression is reportedly regulated by CpG methylation. pDMRs were frequently observed in the vicinity of single-nucleotide variations and insertions/deletions. **Conclusion:** Genetic variations may induce epigenetic variations, generating individual differences in the phenotypes of normal hepatocytes through variations in expression.

First draft submitted: 5 September 2017; Accepted for publication: 6 March 2018; Published online: 4 July 2018

Keywords: International Human Epigenome Consortium • personal differentially methylated region • postbisulfite adaptor tagging

Epigenomic processes are defined as heritable alterations of biological information unaccompanied by changes in the DNA sequence. Such modulation of epigenome profiles contributes to embryonic development and differentiation, and underlies responses to environmental signals such as nutrients and inflammation. Histone modification and DNA methylation are both key elements of epigenetic mechanisms, cooperatively determining chromatin configuration and regulating the levels of gene expression [1]. Epigenome alterations have recently attracted a great

deal of attention as the molecular basis of cancers [2] and metabolic disorders [3], such as hepatocarcinogenesis associated with hepatitis virus infection and/or nonalcoholic steatohepatitis (NASH) [4]. On the basis of disease-specific epigenome profiles, attempts are now being made to explore possible biomarkers for carcinogenetic risk estimation and prognostication [5,6], and druggable targets [7].

In order to accurately identify such disease-specific epigenome profiles, strict comparison with standard epigenome profiles of normal cells is indispensable. However, as epigenomic mechanisms show heterogeneity among tissues and cell lineages, it is not easy to obtain a comprehensive picture of standard epigenome profiles for normal cells. The International Human Epigenome Consortium (IHEC) [8] has set an ambitious goal to decipher at least 1000 epigenomes for various cell lineages and states [9]. We are one of the Japan's IHEC teams [10] and are now focusing on standard epigenome profiling of human normal hepatocytes.

The aim of this study was to reveal the epigenetic landscape based on whole-genome bisulfite sequencing (WGBS), chromatin immunoprecipitation (ChIP) sequencing (seq), RNA-seq and whole-genome sequencing (WGS) data obtained simultaneously from the same batch of purified hepatocytes, and to contribute the results to the IHEC data portal. The obtained data were used to examine correlations between non-CpG methylation and gene expression, the histone modification profiles of personal differentially methylated regions (pDMRs), correlations between personal variation of DNA methylation status and personal variation of gene expression, and genome–epigenome interaction (i.e., correlations between genomic variation and variation of DNA methylation). Notably, this study has revealed that pDMRs are frequent in the vicinity of single-nucleotide variations and insertions/deletions, suggesting that genetic variations may induce epigenetic variations, generating individual differences in the phenotypes of normal hepatocytes through variations in expression.

Materials & methods

Purification of human normal hepatocytes

Six samples of human normal hepatocytes (HPC6, 20, 25, 27, 28 and 35) were purified from partial hepatectomy specimens immediately after surgery as described in Supplementary Methods (page 1) from six Japanese patients without hepatocellular carcinoma who were negative for both hepatitis B virus surface antigen and anti-hepatitis C virus antibody. Clinicopathological information for these patients is summarized in Supplementary Table 1 (page 8).

To confirm the purity of the normal hepatocytes, a proportion of the isolated cells were prepared into formalin-fixed and paraffin-embedded blocks, and subjected to immunocytochemical analysis using Hep Par 1 antibody [11] (clone OCH1E5.2.10, 1:80 dilution, DAKO, CA, USA). All six samples were confirmed to be more than 95% pure (Supplementary Figure 1 [page 303]). The liver consists of various cell lineages, such as hepatocytes, liver sinusoidal endothelial cells, Kupffer cells and satellite cells. Maslak *et al.* showed that liver sinusoidal endothelial cells comprise about 20% of the total number of liver cells [12]. In addition, in rats, Marcos *et al.* showed that the average number of Kupffer cells per animal ranged from $1.54 \pm 0.15 \times 10^8$ (females) to $2.09 \pm 0.14 \times 10^8$ (males) and that of satellite cells was $2.86 \pm 0.58 \times 10^8$, whereas the average number of hepatocytes per animal was $2.0 \pm 0.3 \times 10^9$ [13]. Therefore, a purity of more than 95% can be considered to represent a significant concentration of hepatocytes. Genomic DNA and total RNA were isolated from purified normal hepatocytes using phenol-chloroform followed by dialysis and TRIzol reagent (Thermo Fisher Scientific, MA, USA), respectively.

Polyploidization of hepatocytes in the course of the human life cycle has been reported [14], and the possibility that epigenetic regulation of hepatocyte-specific genes may be affected by multiple nuclei and polyploid nuclei should not be totally neglected. Therefore, we microscopically evaluated the content of hepatocytes showing multiple (two or more) nuclei in hematoxylin-eosin-stained slides. In addition, the content of hepatocytes with polyploid (triploid or more) nuclei was evaluated by fluorescence *in situ* hybridization using a probe designed for the centromeric region of chromosome 12 (Jokoh Co., Ltd, Tokyo, Japan), as described previously [15]. About 1.85–12.40% and 11.51–23.03% of hepatocytes harbored multiple nuclei and polyploid nuclei, respectively. These data are summarized in Supplementary Figure 2 (page 304).

WGBS using postbisulfite adaptor tagging

Genomic DNA was, if necessary, further treated with RNase ONE (Promega, WI, USA) followed by Proteinase K (Qiagen, Hilden, Germany) in Ex Taq Buffer (Takara, Ohtsu, Japan). The enzyme-treated DNA was then purified using AMPure XP SPRI beads (Beckman, CA, USA). Genomic DNA (100 ng) spiked with unmethylated λ phage DNA (1 ng, Promega) was treated with EZ DNA Methylation-Gold kit (Zymo Research, CA, USA). A WGBS

library was constructed according to the postbisulfite adaptor tagging (PBAT) protocol revision 12 [16,17]. The concentration of the library was quantified using real-time PCR using an ABI7000 SDS system (Life Technologies) and KAPA Library Quantification Kits (Kapa Biosystems, MA, USA). Single-end 100-nt reads were generated by HiSeq2000 (Illumina, CA, USA) in accordance with the manufacturer's instructions except for the denaturation step prior to cluster generation [16] (also see Supplementary Discussion [page 4]). The obtained reads were simultaneously mapped to the human reference genome (University of California, Santa Cruz [UCSC] Genome Browser hg19 assembly) and the bacteriophage λ genome (GenBank accession number J02459) using BMap pipeline [18] and visualized as described previously [16]. Quality control (QC) metrics of WGBS using PBAT in each of six samples are summarized in Supplementary Table 2 (page 9): an average read depth of more than 30 was obtained in all samples, thus satisfying the recommendation of the IHEC [19]. All the results of WGBS have been deposited in the National Bioscience Database Center (NBDC) [20] (accession number JGA00000000026), and also can be browsed in the IHEC database [21].

Chromatin immunoprecipitation–sequencing

About 1×10^7 purified normal hepatocytes were crosslinked for 5 min using 1% formaldehyde at room temperature and then neutralized using 0.2 M glycine solution as described previously [22]. Normal hepatocytes were resuspended in SDS lysis buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 1% SDS, 1 mM EDTA). Cell suspensions, including the DNA–protein complexes, were subjected to fragmentation (ten cycles for 30 s, output level 40%) using a SLPe 40 sonifier (Branson, CT, USA) on ice. The lysate was centrifuged at 13,000 r.p.m. for 10 min, and the supernatant was recovered. The supernatant containing 30 μ g of sheared DNA was incubated with 4 μ g of mouse monoclonal antibody against H3K4me3, H3K9me3, H3K27me3, H3K27AC, H3K4me1 and H3K36me3 [23 and references therein] overnight at 4°C with rotation, and then the immunocomplexes were collected with 35 μ l of the magnetic beads (Dynabeads M-280 Sheep anti-mouse IgG, Life Technologies). These antibodies have been employed by many research teams participating in the IHEC, and equivalency of the ChIP protocols has been confirmed by concordance of the results and QC metrics based on analysis of the same samples by participating teams under the supervision of the Assay Standard Working Group of the IHEC. After precipitation, quantitative real-time PCR using primer sets encompassing the enhancer region of the *GAPDH*, *FOXA1* and *ALB* genes was performed for validation of H3K27AC and/or H3K4me1 precipitation. Similarly, that using primer sets encompassing the promoter region of the *GAPDH*, *EEF1A1*, *CNR1* and *KCNA1* genes was performed for validation of H3K4me3 or H3K27me3 precipitation, and that using primer sets encompassing the body of the *GAPDH* gene and Satellite II was performed for validation of H3K36me3 and H3K9me3, respectively. It was confirmed that an appropriate amount of DNA was co-precipitated almost equally from all samples (data not shown). ChIP-Seq libraries were generated from DNA extracted from the immunoprecipitates. Single-end 100-nt reads were generated using the HiSeq 2000 or HiSeq 2500 platforms (Illumina) in accordance with the manufacturer's protocol.

ChIP-seq data for each histone modification were mapped to the human reference genome (UCSC Genome Browser hg19 assembly) using ELAND (Illumina). Using model-based analysis of the ChIP-Seq (MACS) 2 pipeline [24,25] with default parameters, the peaks of each modification were identified. The QC metrics of ChIP-seq for each batch of six samples are summarized in Supplementary Table 3 (page 10). Using ChromHMM [26,27] based on a multivariate hidden Markov model, chromatin states were characterized from ChIP-Seq data. All the results of ChIP-seq have been deposited in the NBDC [20] (accession number JGAS00000000027), and also can be browsed in the IHEC database [21].

Whole transcriptome sequencing (RNA-seq)

cDNA libraries (fragments of 150–200 bp) were prepared from 2 μ g of total RNA using the TruSeq RNA Sample Preparation Kit (Illumina). The RNA-seq libraries were subjected to paired-end 100-bp sequencing using the HiSeq2000 (Illumina) in accordance with the manufacturer's instructions. Sequencing data were mapped to known RNA sequences in the NCBI Reference Sequence database [28], Ensembl [29] and Long noncoding RNA (LincRNA) databases [30] using the Bowtie program [31] as described previously [32]. The QC metrics of RNA-seq for each batch of six samples are summarized in Supplementary Table 4 (page 11). After selecting the best hits with the proper spacing and orientation, the number of reads per kilobase pair per million reads (RPKM) was calculated. All the results of RNA-seq have been deposited in the NBDC [20] (accession number JGAS00000000028), and also can be browsed in the IHEC database [21].

Whole-genome sequencing

DNA was fragmented using an ultrasonic solubilizer (Covaris, MA, USA) using a combination of quick bursts and sonication. DNA fragments of 200–300 bp were gel purified to exclude any inappropriate DNA fusions during library construction. The short fragmented DNA libraries were generated using a pair-end DNA sample prep kit (Illumina) and subjected to paired-end sequencing on the HiSeq2000 (Illumina), in accordance with the manufacturer's instructions. Paired-end reads were aligned to the human reference genome (UCSC Genome Browser hg19 assembly) using Burrows–Wheeler Aligner (version 0.7.12) [33]. Because duplicated reads were generated during the PCR amplification process, pair-end reads that aligned to the same genomic positions were removed using SAMtools [34] (version 1.2) [35] and a program developed in-house. The QC metrics of WGS for each batch of six samples are summarized in Supplementary Table 5 (page 12). Single nucleotide variations (SNVs) and insertions and deletions (indels) were called using SAMtools with stringent confidence filtering conditions we had developed [36]. Copy number information was detected using CNVnator algorithm [37,38].

Pathway & statistical analysis

MetaCore software [39] was used for pathway analysis and statistical analysis was performed using programming language R, both as described in Supplementary Methods (page 1).

Transfection with siRNA

The human liver cancer cell lines KIM-1 and KYN-2 were transfected with either negative control siRNA, *AKT3* siRNA or *EDNRA* siRNA. MTS cell viability assay, apoptosis assay and cell migration assay were then performed, as described in Supplementary Methods (page 1).

Results

CpG methylation status in human normal purified hepatocytes

In order to confirm the reliability of WGBS using PBAT [16], WGBS using the MethylC-Seq method [40] and Infinium Assay [41] using a HumanMethylation450 BeadChip (Illumina) were performed on the same sample, HPC6. Significant correlations of the results between WGBS using PBAT and WGBS using MethylC-Seq, and between WGBS using PBAT and the Infinium assay, were confirmed (Supplementary Figure 3A & B [page 305]).

CpG sites with a read depth of 10 or more were subjected to further analysis. All chromosomes were divided into 30,956,951 100-bp-length windows. Occurrence of *de novo* SNV on CpG sites ($1.944 \pm 0.064\%$) was significantly higher than that on all chromosomes (0.251 ± 0.008) on average for all six samples (Supplementary Table 6 [page 13]). For accurate analysis, CpG sites showing *de novo* SNV in at least one sample were omitted from calculation of the average DNA methylation levels of the 100-bp windows for all six samples.

Each window was annotated as (a) from 200 bp upstream of the transcription start site (TSS) to 1500 bp upstream of it (TSS1500), (b) from TSS to 200 bp upstream of TSS (TSS200), (c) the 5' untranslated region (UTR), (d) the first coding exon (excluding the 5' UTR from the first exon), (e) the first intron, (f) the gene body (second exon and downstream), (g) the 3' UTR and (h) the intergenic region based on the Reference Sequence database [28]. If a window was included in (a) to (g) of multiple genes, it was annotated for the nearest three genes. The number of windows annotated as (a) to (h) are summarized in Supplementary Table 7A (page 14). Each window was also categorized into (i) CpG island, (j) N-Shelf (2000 bp region 5' adjacent to N-Shore), (k) N-Shore (2000 bp region 5' adjacent to CpG island), (l) S-Shore (2000 bp region 3' adjacent to CpG island), (m) S-Shelf (2000 bp region 3' adjacent to S-Shore) and (n) open sea based on the UCSC Genome Browser [42]. The number of windows annotated as (i) to (n) are summarized in Supplementary Table 7B (page 14).

The mean values of the average levels of CpG methylation in 100-bp windows belonging to (a) to (h) for all six samples are shown in Figure 1A. CpG methylation levels of TSS200s and TSS1500s were significantly lower than other regions. Within first exons, CpG methylation levels of coding regions (first coding exons) were significantly lower than those of 5' UTRs. The mean values of the average levels of CpG methylation in 100-bp windows belonging to (i) to (n) for all six samples are shown in Figure 1A. CpG methylation levels of CpG islands were significantly lower than other regions. Variations within individual samples are shown by box and whisker plots in Supplementary Figure 4 (page 307). Average CpG methylation levels of 100-bp windows for genes showing five or more RPKM and those showing less than five are shown in Figures 1B & C, respectively. CpG methylation levels of TSS200s, first coding exons and CpG islands were lower in actively transcribed genes (Figure 1B) than in nonactive genes (Figure 1C). With respect to 100-bp windows belonging to lincRNAs, CpG methylation levels of

Figure 1. Average CpG methylation levels of 100-bp windows for the mean of all six samples (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35). (A) CpG methylation levels of TSS200s (*) were significantly lower than those of TSS1500 ($p = 8.923 \times 10^{-18}$), 5' UTR ($p = 1.062 \times 10^{-15}$), first coding exon ($p = 2.193 \times 10^{-13}$), first intron ($p = 1.942 \times 10^{-14}$), gene body ($p = 5.025 \times 10^{-15}$), 3' UTR ($p = 2.122 \times 10^{-18}$) and intergenic region ($p = 2.122 \times 10^{-18}$). CpG methylation levels of TSS1500s (†) were significantly lower than those of 5' UTR ($p = 1.299 \times 10^{-12}$), first intron ($p = 8.598 \times 10^{-12}$), gene body ($p = 6.184 \times 10^{-13}$), 3' UTR ($p = 7.758 \times 10^{-16}$) and intergenic region ($p = 1.011 \times 10^{-8}$). Within first exons, CpG methylation levels of coding regions (first coding exons, ¶) were significantly lower than those of 5' UTRs ($p = 6.260 \times 10^{-16}$). CpG methylation levels of CpG islands (§) were significantly lower than those of N-Shelf ($p = 7.002 \times 10^{-18}$), N-Shore ($p = 4.147 \times 10^{-16}$), S-Shore ($p = 4.799 \times 10^{-16}$), S-Shelf ($p = 6.696 \times 10^{-18}$) and open sea ($p = 2.820 \times 10^{-14}$). Error bars, standard deviation among windows or samples. CpG methylation levels of TSS1500s (** $p = 6.372 \times 10^{-7}$), TSS 200s (†† $p = 1.132 \times 10^{-8}$), first coding exons (¶¶ $p = 5.201 \times 10^{-9}$) and CpG islands (§§ $p = 2.830 \times 10^{-4}$) in actively transcribed genes (B) were significantly lower than in suppressed genes (C). RPKM: Read per kilobase pair per million read; TSS: Transcription start site; UTR: Untranslated region.

TSS1500s, TSS200s, first exon, CpG islands, N-Shelf, N-Shore and S-Shelf showing RPKM values of 0.5 or more were significantly lower than those showing RPKM values of less than 0.5 (Supplementary Figure 5 [page 313]).

Methylome data for Roadmap liver tissue [43] and Encode purified hepatocytes [44,45] are available from the public database. Although low CpG methylation in TSS200, the first coding exon and CpG island was confirmed even in the Encode data, the CpG methylation levels of Encode purified hepatocytes were generally lower than in our six purified hepatocyte samples and Roadmap liver tissue (Supplementary Figure 6 [page 314]). Such differences between our purified hepatocytes and Roadmap liver tissue on the one hand and Encode purified hepatocytes on the other are probably due to differences in analytical procedure: the Encode protocol involves reduced representation bisulfite sequencing.

Non-CpG methylation status of human normal purified hepatocytes

The number of non-CpG sites (CHG and CHH sites) showing 20% or more cytosine methylation in each of the six samples, and the averaged values for all six samples, are summarized in Supplementary Table 8 (page 15). The mean number of CHG and CHH sites showing such methylation in all six samples (mean $\pm$ standard deviation) were $468,807 \pm 50,737$ and $1,059,551 \pm 75,647$, respectively. The number of CHG sites showing 20% or more cytosine methylation on autosomes and the ratios of CHG sites showing 20% or more cytosine methylation for all CHG sites on autosomes are shown in Figure 2A. The sample-by-sample details are shown in Supplementary Figure 7 (page 315). The number of CHG sites showing 20% or more cytosine methylation on the sense and antisense strands on autosomes are shown in Figure 2A. Average methylation levels (less than 60 vs 60% or more) and the average expression levels of the genes (RPKM value by RNA-seq) are shown in Figure 2B. CHH sites showing 20% or more cytosine methylation are shown in Figure 2C & D in the same way.

Twenty to 40% CHG (Figure 2A) and CHH (Figure 2C) methylation was observed more frequently than 40% or more methylation. CHG (Figure 2A) and CHH (Figure 2C) methylation was observed less frequently in TSS200, the first coding exon and TSS1500. Twenty percent or more CHH methylation on the antisense strand was observed more frequently than on the sense strand (coding strand) in the 5' UTR, first intron, gene body and 3' UTR (Figure 2C), whereas CHG methylation did not show such significant differences between the sense and antisense strands (Figure 2A). CHG and CHH methylation was significantly correlated with higher methylation levels at neighboring CpG sites (Supplementary Table 9 [page 17]). CHG and CHH methylation showed significant inverse association with gene expression levels (Figure 2B & D). The association for lincRNAs is shown in Supplementary Figure 5C & D (page 313).

pDMRs among human normal purified hepatocytes

We focused on personal variation of CpG methylation profiles. If the average DNA methylation level of a 100-bp window in a sample was 0.3 or more higher than the maximum DNA methylation level of the other five samples and DNA methylation levels of the other five samples were within 0.3, the window was defined as 'pDMR showing DNA hypermethylation' for the sample (Supplementary Figure 8A [page 316]). 'pDMR showing DNA hypomethylation' for the sample was defined in the same way (Supplementary Figure 8B [page 316]). Representative pDMRs showing DNA hypermethylation and DNA hypomethylation are shown in Figure 3A. Since PBAT shows good coverage for GC-rich regions, more accurate evaluation can be performed for GC-rich pDMRs in comparison to MethylC-seq (Supplementary Figure 3C [page 306]). The number of pDMRs for each sample and the mean of all six samples are summarized in Supplementary Table 10 (page 18).

Pie charts in Figure 3B show the ratios of pDMRs belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region for all pDMRs of each sample. Differences in the ratio among samples were not obvious among the pie charts. The bar graph in Figure 3B shows the average ratios of the number of pDMRs per total number of 100-bp windows on autosomes containing CpG sites and belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region of all six samples. Personal variations in CpG methylation levels were observed less frequently in TSS200s, first coding exons and TSS1500s (bar graph in Figure 3B). Pie charts in Figure 3C show the ratio of pDMRs belonging to CpG island, N-Shelf, N-Shore, S-Shore and S-Shelf for all pDMRs in each sample. Differences in the ratio among samples were not obvious among the pie charts. Personal variations in CpG methylation levels were observed less frequently in CpG islands (bar graph in Figure 3C).

Among all the examined 100-bp windows on autosomes, 21,682 and 21 windows showed significant correlations between the average CpG methylation levels and age and sex of the six individuals, respectively. Among the 21,682

Figure 2. Non-CpG (CHG and CHH) methylation status based on the mean values for human normal purified hepatocytes (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35). (A) Methylation status on CHG sites. The number of methylated CHG sites in TSS200s (*) were significantly lower than those of TSS1500 ($p = 7.544 \times 10^{-6}$), 5' UTR ($p = 3.649 \times 10^{-6}$), first intron ($p = 4.154 \times 10^{-6}$), gene body ($p = 2.786 \times 10^{-6}$), 3' UTR ($p = 3.318 \times 10^{-6}$) and intergenic region ($p = 3.182 \times 10^{-6}$). The number of methylated CHG sites in first coding exons (†) were significantly lower than those of TSS1500 ($p = 7.213 \times 10^{-6}$), TSS200 ($p = 9.236 \times 10^{-5}$), 5' UTR ($p = 3.453 \times 10^{-6}$), first intron ($p = 3.785 \times 10^{-6}$), gene body ($p = 2.673 \times 10^{-6}$), 3' UTR ($p = 3.060 \times 10^{-6}$) and intergenic region ($p = 3.147 \times 10^{-6}$). The number of methylated CHG sites in TSS1500 (‡) were significantly lower than those of 5' UTR ($p = 4.864 \times 10^{-6}$), first intron ($p = 5.772 \times 10^{-6}$), gene body ($p = 3.512 \times 10^{-6}$), 3' UTR ($p = 3.930 \times 10^{-6}$) and intergenic region ($p = 3.461 \times 10^{-6}$). The ratios of CHG sites showing 20% or more cytosine methylation among all CHG sites belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region are shown in the middle panel. The number of CHG sites (mean $\pm$ standard error for all six samples) showing 20% or more cytosine methylation on the sense and antisense strands are shown in the right panel. **(B)** Association between CHG methylation and gene expression. CHG methylation in TSS1500, TSS200 and first coding exon ($p = 4.022 \times 10^{-4}$) and in 5' UTR, first intron, gene body and 3' UTR ($p = 2.369 \times 10^{-3}$) was inversely associated with expression levels. **(C)** Methylation status on CHH sites. The number of methylated CHH sites in TSS200s (*) were significantly lower than those of TSS1500 ($p = 1.041 \times 10^{-6}$), 5' UTR ($p = 8.590 \times 10^{-7}$), first intron ($p = 5.310 \times 10^{-7}$), gene body ($p = 3.369 \times 10^{-6}$), 3' UTR ($p = 1.649 \times 10^{-6}$) and intergenic region ($p = 1.939 \times 10^{-7}$). The number of methylated CHH sites in first coding exons (†) were significantly lower than those of TSS1500 ($p = 9.096 \times 10^{-7}$), TSS200 ($p = 1.714 \times 10^{-6}$), 5' UTR ($p = 8.214 \times 10^{-7}$), first intron ($p = 4.849 \times 10^{-7}$), gene body ($p = 3.260 \times 10^{-6}$), 3' UTR ($p = 1.514 \times 10^{-6}$) and intergenic region ($p = 1.924 \times 10^{-7}$). The number of methylated CHH sites in TSS1500 (‡) were significantly lower than those of 5' UTR ($p = 1.161 \times 10^{-6}$), first intron ($p = 4.606 \times 10^{-7}$), gene body ($p = 4.555 \times 10^{-6}$), 3' UTR ($p = 2.801 \times 10^{-6}$) and intergenic region ($p = 2.075 \times 10^{-7}$). The ratio of CHH sites showing 20% or more cytosine methylation among all CHH sites belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region is shown in the middle panel. Twenty percent or more CHH methylation was observed more frequently on the antisense strand than on the sense strand (coding strand) in the 5' UTR ($p = 1.312 \times 10^{-3}$), first intron ($p = 1.214 \times 10^{-4}$), gene body ($p = 2.598 \times 10^{-3}$) and 3' UTR ($p = 1.885 \times 10^{-2}$) (right panel of C), whereas CHG methylation did not show such significant differences between the sense and antisense strands (right panel of A). **(D)** Association between CHH methylation and gene expression. CHH methylation in TSS1500, TSS200 and first coding exon ($p = 1.170 \times 10^{-13}$) and in 5' UTR, first intron, gene body and 3' UTR ($p = 3.337 \times 10^{-4}$) was inversely associated with expression levels. Error bars, standard error. TSS: Transcription start site; UTR: Untranslated region.

Figure 2. Non-CpG (CHG and CHH) methylation status based on the mean values for human normal purified hepatocytes (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35) (cont.). (A) Methylation status on CHG sites. The number of methylated CHG sites in TSS200s (*) were significantly lower than those of TSS1500 ($p = 7.544 \times 10^{-6}$), 5' UTR ($p = 3.649 \times 10^{-6}$), first intron ($p = 4.154 \times 10^{-6}$), gene body ($p = 2.786 \times 10^{-6}$), 3' UTR ($p = 3.318 \times 10^{-6}$) and intergenic region ($p = 3.182 \times 10^{-6}$). The number of methylated CHG sites in first coding exons (†) were significantly lower than those of TSS1500 ($p = 7.213 \times 10^{-6}$), TSS200 ($p = 9.236 \times 10^{-5}$), 5' UTR ($p = 3.453 \times 10^{-6}$), first intron ($p = 3.785 \times 10^{-6}$), gene body ($p = 2.673 \times 10^{-6}$), 3' UTR ($p = 3.060 \times 10^{-6}$) and intergenic region ($p = 3.147 \times 10^{-6}$). The number of methylated CHG sites in TSS1500 (¶) were significantly lower than those of 5' UTR ($p = 4.864 \times 10^{-6}$), first intron ($p = 5.772 \times 10^{-6}$), gene body ($p = 3.512 \times 10^{-6}$), 3' UTR ($p = 3.930 \times 10^{-6}$) and intergenic region ($p = 3.461 \times 10^{-6}$). The ratios of CHG sites showing 20% or more cytosine methylation among all CHG sites belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region are shown in the middle panel. The number of CHG sites (mean $\pm$ standard error for all six samples) showing 20% or more cytosine methylation on the sense and antisense strands are shown in the right panel. (B) Association between CHG methylation and gene expression. CHG methylation in TSS1500, TSS200 and first coding exon ($p = 4.022 \times 10^{-4}$) and in 5' UTR, first intron, gene body and 3' UTR ($p = 2.369 \times 10^{-3}$) was inversely associated with expression levels. (C) Methylation status on CHH sites. The number of methylated CHH sites in TSS200s (*) were significantly lower than those of TSS1500 ($p = 1.041 \times 10^{-6}$), 5' UTR ($p = 8.590 \times 10^{-7}$), first intron ($p = 5.310 \times 10^{-7}$), gene body ($p = 3.369 \times 10^{-6}$), 3' UTR ($p = 1.649 \times 10^{-6}$) and intergenic region ($p = 1.939 \times 10^{-7}$). The number of methylated CHH sites in first coding exons (†) were significantly lower than those of TSS1500 ($p = 9.096 \times 10^{-7}$), TSS200 ($p = 1.714 \times 10^{-6}$), 5' UTR ($p = 8.214 \times 10^{-7}$), first intron ($p = 4.849 \times 10^{-7}$), gene body ($p = 3.260 \times 10^{-6}$), 3' UTR ($p = 1.514 \times 10^{-6}$) and intergenic region ($p = 1.924 \times 10^{-7}$). The number of methylated CHH sites in TSS1500 (¶) were significantly lower than those of 5' UTR ($p = 1.161 \times 10^{-6}$), first intron ($p = 4.606 \times 10^{-7}$), gene body ($p = 4.555 \times 10^{-6}$), 3' UTR ($p = 2.801 \times 10^{-6}$) and intergenic region ($p = 2.075 \times 10^{-7}$). The ratio of CHH sites showing 20% or more cytosine methylation among all CHH sites belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region is shown in the middle panel. Twenty percent or more CHH methylation was observed more frequently on the antisense strand than on the sense strand (coding strand) in the 5' UTR ($p = 1.312 \times 10^{-3}$), first intron ($p = 1.214 \times 10^{-4}$), gene body ($p = 2.598 \times 10^{-3}$) and 3' UTR ($p = 1.885 \times 10^{-2}$) (right panel of C), whereas CHG methylation did not show such significant differences between the sense and antisense strands (right panel of A). (D) Association between CHH methylation and gene expression. CHH methylation in TSS1500, TSS200 and first coding exon ($p = 1.170 \times 10^{-13}$) and in 5' UTR, first intron, gene body and 3' UTR ($p = 3.337 \times 10^{-4}$) was inversely associated with expression levels. Error bars, standard error. TSS: Transcription start site; UTR: Untranslated region.

age-related windows, 53 were included in pDMRs: the average DNA methylation levels of less than 1% of the pDMRs showed significant correlations with age (Supplementary Table 11 [page 19]). Among the 21 sex-related windows (summarized in Supplementary Table 12 [page 20]), none was included in the pDMRs.

Histone modification profiles in human normal purified hepatocytes

Representative wiggle tracks for each mark of the *ALB* and *PLK* genes are shown in Supplementary Figure 9 (page 317). Differing from the ChIP-seq data (H3K27AC and H3K27me3) for three hepatocyte samples without detailed metadata deposited in public databases [46–48], our data consist of six marks (H3K4me3, H3K9me3, H3K27me3, H3K27AC, H3K4me1 and H3K36me3) and were obtained from healthy hepatocytes with accurate metadata, as shown in Supplementary Table 1 (page 8), for which DNA methylation and expression profiles are available.

When we focused on representative genes encoding transcription factors that are highly expressed in hepatocytes and reportedly important for hepatocyte function, such as *HNF1* and *HNF4*, lower DNA methylation levels and

Figure 3. Personal differentially methylated regions in human normal purified hepatocytes. (A) Representative pDMRs (blue boxes) showing DNA hypermethylation (a: HPC6, HPC20 and HPC28) and DNA hypomethylation (b: HPC25) based on the definition shown in Supplementary Figure 8 (page 316). (B) The ratio of pDMRs belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region for all pDMRs in each sample (pie charts) and the average ratios of the number of pDMRs per total number of 100-bp windows on autosomes that contain CpG sites and belong to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region in all six samples (bar graph). pDMRs in TSS200s (*) were less frequent than those in TSS1500 ($p = 2.550 \times 10^{-2}$), 5' UTR ($p = 1.755 \times 10^{-2}$) and intergenic region ($p = 6.466 \times 10^{-3}$). pDMRs in first coding exons (†) were less frequent than those in TSS1500 ($p = 7.647 \times 10^{-3}$), 5' UTR ($p = 3.385 \times 10^{-3}$), first intron ($p = 5.665 \times 10^{-3}$), gene body ($p = 1.030 \times 10^{-2}$), 3' UTR ($p = 6.319 \times 10^{-3}$) and intergenic region ($p = 6.294 \times 10^{-3}$). pDMRs in TSS1500 (¶) were less frequent than those in 5' UTR ($p = 5.451 \times 10^{-4}$), first intron ($p = 4.724 \times 10^{-3}$), gene body ($p = 2.846 \times 10^{-2}$) and intergenic region ($p = 6.743 \times 10^{-3}$). (C) pDMRs in CpG island (§) were less frequent than those in N-Shelf ($p = 2.010 \times 10^{-2}$), N-Shore ($p = 5.850 \times 10^{-3}$), S-Shore ($p = 6.111 \times 10^{-3}$), S-Shelf ($p = 8.473 \times 10^{-3}$) and open sea ($p = 5.695 \times 10^{-3}$). Chr: Chromosome; pDMR: Personal differentially methylated region; UTR: Untranslated region.

Figure 3. Personal differentially methylated regions in human normal purified hepatocytes (cont.). (A) Representative pDMRs (blue boxes) showing DNA hypermethylation (a: HPC6, HPC20 and HPC28) and DNA hypomethylation (b: HPC25) based on the definition shown in Supplementary Figure 8 (page 316). (B) The ratio of pDMRs belonging to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region for all pDMRs in each sample (pie charts) and the average ratios of the number of pDMRs per total number of 100-bp windows on autosomes that contain CpG sites and belong to TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region in all six samples (bar graph). pDMRs in TSS200s (*) were less frequent than those in TSS1500 ($p = 2.550 \times 10^{-2}$), 5' UTR ($p = 1.755 \times 10^{-2}$) and intergenic region ($p = 6.466 \times 10^{-3}$). pDMRs in first coding exons (†) were less frequent than those in TSS1500 ($p = 7.647 \times 10^{-3}$), 5' UTR ($p = 3.385 \times 10^{-3}$), first intron ($p = 5.665 \times 10^{-3}$), gene body ($p = 1.030 \times 10^{-2}$), 3' UTR ($p = 6.319 \times 10^{-3}$) and intergenic region ($p = 6.294 \times 10^{-3}$). pDMRs in TSS1500 (¶) were less frequent than those in 5' UTR ($p = 5.451 \times 10^{-4}$), first intron ($p = 4.724 \times 10^{-3}$), gene body ($p = 2.846 \times 10^{-2}$) and intergenic region ($p = 6.743 \times 10^{-3}$). (C) pDMRs in CpG island (§) were less frequent than those in N-Shelf ($p = 2.010 \times 10^{-2}$), N-Shore ($p = 5.850 \times 10^{-3}$), S-Shore ($p = 6.111 \times 10^{-3}$), S-Shelf ($p = 8.473 \times 10^{-3}$) and open sea ($p = 5.695 \times 10^{-3}$).

Chr: Chromosome; pDMR: Personal differentially methylated region; UTR: Untranslated region.

H3K4me3 and H3K27AC marks around the TSS were observed in all six of our purified hepatocyte samples without exception (representative data are shown in Figure 4A), whereas such marks were not observed in genes encoding organ-specific transcription factors that are not expressed in hepatocytes, such as *PAX8* and *CDX2* (all panels for all six samples of the *HNF1*, *HNF4*, *PAX8* and *CDX2* genes are shown in Supplementary Figure 10 [page 319]). In addition, the number of peaks revealed by ChIP-seq that were detected in all three samples from males but not in all three samples from females, or vice versa, are summarized in Supplementary Table 13 (page 21). The H3K4me3 and H3K27AC peaks showing such sex differences frequently resulted in significant overexpression in samples from males in comparison with those from females, or vice versa, respectively (Supplementary Table 14 [page 22]).

All 100-bp windows on all autosomes of all six samples were grouped into ten subclusters based on ChIP-seq data by the ChromHMM pipeline. The heatmap obtained by ChromHMM is shown in Figure 4B and the emission parameters for each cluster are summarized in Supplementary Table 15 (page 23). We annotated each cluster as follows: clusters 3, 4 and 5: H3K4me1 and H3K27AC predominant pattern, clusters 1 and 2: H3K4me3 and H3K27AC predominant pattern, clusters 9 and 10: H3K36me3 predominant pattern, cluster 6: no distinctive mark pattern, cluster 8: H3K27me3 predominant pattern and cluster 7: H3K9me3 predominant pattern. Analogous with the annotation suggested by Ernst *et al.* [49], the 3K4me1 and H3K27AC predominant pattern, the H3K4me3 and H3K27AC predominant pattern, the H3K36me3 predominant pattern, the no distinctive mark pattern, the H3K27me3 predominant pattern and the H3K9me3 predominant pattern have also been interpreted as the strong enhancer cluster, active promoter cluster, transcribed gene cluster, low signal cluster, polycomb-repressed gene cluster

Figure 4. Histone modification status based on the mean values for human normal purified hepatocytes (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35). (A) Representative wiggle tracks of genes encoding transcription factors that are highly expressed in hepatocytes and reportedly important for hepatocyte function, such as *HNF1* and *HNF4*, in HPC6 and HPC27 samples. Panels of the *HNF1* and *HNF4* genes in four other samples and for genes encoding organ-specific transcription factors that are not expressed in hepatocytes, such as *PAX8* and *CDX2*, in the six purified hepatocyte samples are shown in Supplementary Figure 10 (page 319). (B) Heatmap of ChIP-seq data in HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35 generated by the ChromHMM pipeline [27]. The ratios (mean of all six samples) for all autosomal windows (C), windows annotated to genes showing five or more RPKMs (D), windows annotated to genes showing RPKM values of less than 5 (E), and pDMRs (F) are shown.

ChIP-seq: Chromatin immunoprecipitation–sequencing; pDMR: Personal differentially methylated region; RPKM: Read per kilobase pair per million read.

Figure 4. Histone modification status based on the mean values for human normal purified hepatocytes (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35) (cont.). (A) Representative wiggle tracks of genes encoding transcription factors that are highly expressed in hepatocytes and reportedly important for hepatocyte function, such as *HNF1* and *HNF4*, in HPC6 and HPC27 samples. Panels of the *HNF1* and *HNF4* genes in four other samples and for genes encoding organ-specific transcription factors that are not expressed in hepatocytes, such as *PAX8* and *CDX2*, in the six purified hepatocyte samples are shown in Supplementary Figure 10 (page 319). (B) Heatmap of ChIP-seq data in HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35 generated by the ChromeHMM pipeline [27]. The ratios (mean of all six samples) for all autosomal windows (C), windows annotated to genes showing five or more RPKMs (D), windows annotated to genes showing RPKM values of less than 5 (E), and pDMRs (F) are shown. ChIP-seq: Chromatin immunoprecipitation–sequencing; pDMR: Personal differentially methylated region; RPKM: Read per kilobase pair per million read.

and heterochromatin region cluster, respectively. The ratio of autosomal windows showing histone modifications of the H3K4me1 and H3K27AC predominant pattern (strong enhancer: clusters 3, 4 and 5), H3K4me3 and H3K27AC predominant pattern (active promoter: clusters 1 and 2), H3K36me3 predominant pattern (transcribed gene: clusters 9 and 10), no distinctive mark pattern (low signal: cluster 6), H3K27me3 predominant pattern (polycomb-repressed gene: cluster 8) and H3K9me3 predominant pattern (heterochromatin region: cluster 7) are shown in Figure 4C. The sample-by-sample details are shown in Supplementary Figure 11 (page 323). Differing from the DNA methylation profiles shown in Figure 3B & C, considerable variations in the histone modification profiles among samples shown in Supplementary Figure 11 (page 323) were observed. On the other hand, the finding that H3K4 methylation and H3K27 acetylation were frequently observed in TSS200, the first coding exon and CpG island, especially in genes showing RPKM values of 5 or more (Figure 4D), whereas H3K27me3 suppression was frequently observed in genes showing RPKM values of less than 5 (Figure 4E), was not unexpected.

The histone modification patterns of pDMRs are shown in Figure 4E. The sample-by-sample details are shown in Supplementary Figure 11G (page 325). Histone modification patterns that were observed frequently or less frequently (binomial test; $p < 0.05$) in pDMRs than in all autosomes are summarized in Supplementary Table 16 (page 24). When compared with all autosomes, histone modification profiles of pDMRs did not show a common tendency in TSS1500, TSS200, 5' UTR, first coding exon, first intron, gene body, 3' UTR and intergenic region, or in CpG island, N-Shelf, N-Shore S-Shore, S-Shelf and open sea: histone modifications frequently observed in pDMRs differed from each other among each batch of six samples (Supplementary Table 16 [page 24]).

Expression profiles in human normal purified hepatocytes

We focused on genes showing personal variations in expression levels. If a gene showed an RPKM value that was at least twice the maximum for the other five samples, it was defined as personal overexpression in that sample. If a gene showed an RPKM value that was a half or less of the minimum for other five samples, it was defined as personal underexpression in that sample. Genes showing personal over- or underexpression in each sample are summarized in Supplementary Table 17 (page 25): 6864, 2648, 2700, 2121, 2828 and 3628 genes for HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35, respectively.

Figure 4. Histone modification status based on the mean values for human normal purified hepatocytes (HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35) (cont.). (A) Representative wiggle tracks of genes encoding transcription factors that are highly expressed in hepatocytes and reportedly important for hepatocyte function, such as *HNF1* and *HNF4*, in HPC6 and HPC27 samples. Panels of the *HNF1* and *HNF4* genes in four other samples and for genes encoding organ-specific transcription factors that are not expressed in hepatocytes, such as *PAX8* and *CDX2*, in the six purified hepatocyte samples are shown in Supplementary Figure 10 (page 319). (B) Heatmap of ChIP-seq data in HPC6, HPC20, HPC25, HPC27, HPC28 and HPC35 generated by the ChromHMM pipeline [27]. The ratios (mean of all six samples) for all autosomal windows (C), windows annotated to genes showing five or more RPKMs (D), windows annotated to genes showing RPKM values of less than 5 (E), and pDMRs (F) are shown.

ChIP-seq: Chromatin immunoprecipitation–sequencing; pDMR: Personal differentially methylated region; RPKM: Read per kilobase pair per million read.

If a pDMR showing DNA hypermethylation (or hypomethylation) was located within 1500 bp upstream or downstream of TSS of a gene showing personal underexpression (or overexpression) in the same sample, respectively, we considered that the pDMR was potentially associated with personal expression variation in normal purified hepatocytes from the same individual: genes showing such an association are summarized in Supplementary Table 18 (page 281).

In order to validate the association between variations in DNA methylation levels and variations in expression levels of the 30 genes listed in Supplementary Table 18 (page 281), DNA methylation and expression data deposited in a public database (Integrative Disease Omics Database [50]) were examined. Significant inverse correlations between DNA methylation levels and expression levels for the *AURKB*, *BOK*, *C11orf21*, *CLEC2B*, *CSDC2*, *CYP4F12*, *GRB7*, *KLHDC7B*, *MYOC*, *NPR2*, *PLLP*, *RASGRF2*, *SLC28A1*, *SLC44A3*, *STEAP2*, *TFF1* and *TMEM105* genes were confirmed in the database (Supplementary Table 19 [page 282]).

Genome–epigenome correlation

The number of SNVs that were observed in only one sample but not in the other five, in each sample batch are summarized in Supplementary Table 20A (page 283). The number of pDMRs existing in the vicinity, in other words, within 1000 bp upstream or downstream, of SNVs in each sample are summarized in Supplementary Table

Figure 5. Genome–epigenome correlation in human normal purified hepatocytes. (A) Histogram of pDMRs according to the distance from SNVs in six samples of normal hepatocytes. pDMRs were observed more frequently in the vicinity, in other words, within 1000 bp upstream or downstream, of SNVs than on all autosomes in all six samples (also see Table 1). (B) Histogram of pDMRs according to the distance from indels in six samples of normal hepatocytes. pDMRs were observed more frequently in the vicinity, in other words, within 1000 bp upstream or downstream, of indels than on all autosomes in five of the six samples (also see Table 1). (C) Histogram of regions showing CNVs according to the distance from pDMRs in six samples of normal hepatocytes. CNVs were observed more frequently in the vicinity, in other words, within 1000 bp upstream or downstream, of pDMRs (especially pDMRs showing DNA hypomethylation), than on all autosomes in five of the six samples (also see Table 1).
bp: Base pair; CNV: Copy number variation; indel: Insertion/deletion; pDMR: Personal differentially methylated region; SNV: Single nucleotide variation.

20A (page 283): pDMRs were observed more frequently in the vicinity of SNVs than on all autosomes in all six samples of hepatocytes (binomial test, Table 1). Such a correlation between SNVs and pDMRs is also shown in Figure 5A. The number of pDMRs existing in the vicinity, in other words, within 1000 bp upstream or downstream, of indels in each sample are summarized in Supplementary Table 20B (page 283): pDMRs were observed more frequently in the vicinity of indels than on all autosomes in five of the six hepatocyte samples examined (binomial test, Table 1). Such a correlation between SNVs and pDMRs is also shown in Figure 5B.

The number of regions showing copy number variations (CNVs) in each sample are summarized in Supplementary Table 20C (page 284). The number of CNVs existing in the vicinity, in other words, within 1000 bp upstream or downstream, of pDMRs in each sample are summarized in Supplementary Table 20C (page 284): CNVs were observed more frequently in the vicinity of pDMRs, especially pDMRs showing DNA hypomethylation, than on all autosomes in five of the six hepatocyte samples examined (binomial test, Table 1). Such correlations between pDMRs and CNVs are also shown in Figure 5C.

Table 1. Genome–epigenome correlation: personal differentially methylated regions, genetic polymorphisms[†] and copy number variations.

The number of pDMRs in the vicinity of SNVs				
Sample ID	pDMRs	Number of pDMRs per one base		p-value (binomial test)
		Within 1000 bp upstream and downstream from SNVs	On all autosomes	
HPC6	DNA hypermethylation	8.778E-07	5.894E-07	1.539E-19*
	DNA hypomethylation	1.087E-06	7.518E-07	9.060E-21*
	Total	1.965E-06	1.341E-06	1.702E-38*
HPC20	DNA hypermethylation	4.455E-07	2.898E-07	4.982E-13*
	DNA hypomethylation	1.270E-06	1.052E-06	2.522E-08*
	Total	1.716E-06	1.342E-06	6.079E-17*
HPC25	DNA hypermethylation	1.372E-06	1.060E-06	6.271E-14*
	DNA hypomethylation	1.147E-06	8.240E-07	4.501E-18*
	Total	2.519E-06	1.884E-06	6.826E-30*
HPC27	DNA hypermethylation	5.53E-07	3.356E-07	5.974E-19*
	DNA hypomethylation	2.39E-06	1.900E-06	1.502E-18*
	Total	2.94E-06	2.236E-06	2.964E-31*
HPC28	DNA hypermethylation	1.020E-06	7.046E-07	1.539E-19*
	DNA hypomethylation	1.095E-06	6.991E-07	2.546E-29*
	Total	2.115E-06	1.404E-06	1.281E-46*
HPC35	DNA hypermethylation	4.896E-07	2.496E-07	2.715E-28*
	DNA hypomethylation	9.732E-07	6.081E-07	4.726E-29*
	Total	1.463E-06	8.577E-07	1.661E-53*
The number of pDMRs in the vicinity of indels				
Sample ID	pDMRs	Number of pDMRs per one base		p-value (binomial test)
		Within 1000 bp upstream and downstream from indels	On all autosomes	
HPC6	DNA hypermethylation	1.177E-06	5.894E-07	7.553E-06*
	DNA hypomethylation	1.041E-06	7.518E-07	3.650E-02*
	Total	2.218E-06	1.341E-06	3.526E-06*
HPC20	DNA hypermethylation	3.631E-07	2.898E-07	3.031E-01
	DNA hypomethylation	8.217E-07	1.052E-06	1.203E-01
	Total	1.185E-06	1.342E-06	3.702E-01
HPC25	DNA hypermethylation	1.207E-06	1.060E-06	3.257E-01
	DNA hypomethylation	1.436E-06	8.240E-07	2.153E-05*
	Total	2.643E-06	1.884E-06	2.765E-04*
HPC27	DNA hypermethylation	6.19E-07	3.356E-07	1.785E-03*
	DNA hypomethylation	2.89E-06	1.900E-06	3.265E-06*
	Total	3.51E-06	2.236E-06	3.939E-08*
HPC28	DNA hypermethylation	1.203E-06	7.046E-07	2.173E-04*
	DNA hypomethylation	1.095E-06	6.991E-07	2.705E-03*
	Total	2.298E-06	1.404E-06	2.249E-06*
HPC35	DNA hypermethylation	6.908E-07	2.496E-07	8.506E-07*
	DNA hypomethylation	1.136E-06	6.081E-07	4.710E-05*
	Total	1.827E-06	8.577E-07	8.473E-10*
The number of regions showing CNVs in the vicinity of pDMRs				
Sample ID	pDMRs	Number of CNVs per one base		p-value (binomial test)
		Within 1000 bp upstream and downstream from pDMRs	On all autosomes	
HPC6	DNA hypermethylation	9.151E-07	4.731E-07	2.040E-01
	DNA hypomethylation	3.159E-06		5.194E-08*
	Total	2.205E-06		3.282E-07*

[†]Single nucleotide variations and insertions/deletions.
 *p-values of <0.05.
 CNV: Copy number variation; indel: Insertion/deletion; pDMR: Personal differentially methylated region; SNV: Single nucleotide variation.

Table 1. Genome–epigenome correlation: personal differentially methylated regions, genetic polymorphisms[†] and copy number variations (cont.).

The number of regions showing CNVs in the vicinity of pDMRs				
Sample ID	pDMRs	Number of CNVs per one base		p-value (binomial test)
		Within 1000 bp upstream and downstream from pDMRs	On all autosomes	
HPC20	DNA hypermethylation	1.292E-06	4.044E-07	1.306E-01
	DNA hypomethylation	1.673E-06		2.132E-04*
	Total	1.594E-06		8.155E-05*
HPC25	DNA hypermethylation	3.207E-07	4.616E-07	1
	DNA hypomethylation	6.961E-07		4.572E-01
	Total	4.741E-07		8.201E-01
HPC27	DNA hypermethylation	1.544E-06	2.211E-07	9.606E-03*
	DNA hypomethylation	7.138E-07		4.034E-03*
	Total	8.364E-07		2.252E-04*
HPC28	DNA hypermethylation	2.577E-06	6.786E-07	4.144E-04*
	DNA hypomethylation	1.729E-06		2.235E-02*
	Total	2.144E-06		4.870E-05*
HPC35	DNA hypermethylation	7.289E-07	1.604E-07	1.975E-01
	DNA hypomethylation	8.285E-07		2.125E-02*
	Total	8.011E-07		9.105E-03*

[†]Single nucleotide variations and insertions/deletions.
 *p-values of <0.05.
 CNV: Copy number variation; indel: Insertion/deletion; pDMR: Personal differentially methylated region; SNV: Single nucleotide variation.

Pathway analysis

Supplementary Table 21 (page 285) summarizes 179 genes showing both genetic variations (SNVs and/or indels) and epigenetic variations (pDMRs) in two or more human normal purified hepatocyte samples. MetaCore pathway analysis by GeneGo using these 179 genes revealed that genes showing both genetic and epigenetic variations were significantly accumulated in 168 pathways ($p < 0.05$, Supplementary Table 22 [page 289]). After exclusion of pathways specific to organs other than the liver, 46 pathways, which are potentially correlated with liver function, epithelial cell function or immune response, and showed a p-value of 0.01 or less, are summarized in Table 2. Two representative signaling pathways, ‘Development_Endothelin-1/EDNRA transactivation of EGFR’ ($p = 9.282 \times 10^{-6}$) [51,52] and ‘G-protein signaling_G-Protein α -q signaling cascades’ ($p = 5.671 \times 10^{-5}$) [53–55] have been illustrated schematically in Supplementary Figure 12 (page 326).

Transfection with siRNA

Knockdown of the *AKT3* gene, a representative gene frequently showing both SNV and pDMR and listed in Table 2, in hepatocellular carcinoma KIM-1 and KYN-2 cells, which show higher levels of *AKT3* expression, resulted in increased cell viability ($p = 1.627 \times 10^{-2}$ for KYN-2), reduced apoptosis ($p = 1.765 \times 10^{-5}$ for KIM-1) and enhanced cell migration ($p = 7.721 \times 10^{-3}$ for KYN-2) (Figure 6A). Knockdown of another representative gene, *EDNRA*, in KIM-1 and KYN-2 cells, which show higher levels of *EDNRA* expression, resulted in reduced apoptosis ($p = 4.011 \times 10^{-6}$ for KIM-1) and enhanced cell migration ($p = 7.046 \times 10^{-3}$ for KIM-1 and $p = 1.241 \times 10^{-2}$ for KYN-2) (Figure 6B).

Discussion

Epigenome profiles satisfying the IHEC recommendations have been obtained using highly reliable methods, such as PBAT [16], in multiple samples of human normal hepatocytes highly purified immediately after surgery. Although six patients had liver metastases from cancers arising in organs other than the liver, and their hepatocytes were free from hepatitis virus infection and liver-specific precancerous conditions, with no remarkable histological features, they were of relatively advanced age and may have been affected by carcinogenetic factors capable of inducing cancers of organs other than the liver. Therefore, the possibility that our hepatocytes may have represented ‘totally normal materials’ is debatable. In addition, since simultaneous WGBS, ChIP-seq, RNA-seq and WGS for the same purified sample involved laborious work, such a study of a larger cohort would have been impractical. Therefore,

Table 2. Signaling pathways in which genes showing both genetic variations[†] and epigenetic variations[‡] were significantly accumulated, as revealed by MetaCore pathway analysis[§].

Pathways	p-value	Included genes showing genetic and epigenetic variations
Development_Endothelin-1/EDNRA transactivation of EGFR	9.282E-06	<i>EDNRA, ITPR1, AKT3, PTK2B, PLCB4</i>
Immune response_IL-16 signaling pathway	2.061E-05	<i>ITPR1, CD4, SKP2, AKT3, SKP2</i>
G-protein signaling_G-Protein α -q signaling cascades	5.671E-05	<i>ITPR1, AKT3, PTK2B, PLCB4</i>
Development_ACM2 and ACM4 activation of ERK	1.447E-04	<i>CALDI1, ITPR1, PTK2B, PLCB4</i>
Cell adhesion_Histamine H1 receptor signaling in the interruption of cell barrier integrity	1.731E-04	<i>ITPR1, PTK2B, FAK1, PLCB4</i>
Signal transduction_Calcium signaling	1.731E-04	<i>SLC8A1, ITPR1, IP3R1, PLCB4</i>
Development_A3 receptor signaling	2.415E-04	<i>ITPR1, AKT3, PTK2B, PLCB4</i>
Development_A2B receptor: action via G-protein α s	2.613E-04	<i>ITPR1, AKT3, PTK2B, PLCB4</i>
Development_Endothelin-1/EDNRA signaling	3.276E-04	<i>EDNRA, ITPR1, AKT3, PTK2B</i>
Immune response_Inhibitory PD-1 signaling in T cells	3.276E-04	<i>CD4, SKP2, AKT3, SKP2</i>
Cytoskeleton remodeling_FAK signaling	4.338E-04	<i>ITPR1, AKT3, FAK1, PLCB4</i>
Development_Activation of ERK by Kappa-type opioid receptor	9.359E-04	<i>ITPR1, PTK2B, PLCB4</i>
Development_Angiotensin signaling via STATs	1.028E-03	<i>ITPR1, PTK2B, PLCB4</i>
Development_Angiotensin activation of ERK	1.126E-03	<i>ITPR1, PTK2B, PLCB4</i>
Immune response_CXCR4 signaling via second messenger	1.229E-03	<i>ITPR1, AKT3, PTK2B</i>
Immune response_Role of the Membrane attack complex in cell survival	1.339E-03	<i>ITPR1, AKT3, PLCB4</i>
G-protein signaling_S1P2 receptor signaling	1.339E-03	<i>ITPR1, AKT3, FAK1</i>
HIV-1 signaling via CCR5 in macrophages and T lymphocytes	1.836E-03	<i>ITPR1, CD4, PTK2B</i>
Development_VEGF-family signaling	2.123E-03	<i>NRP1, ITPR1, AKT3</i>
Immune response_PIP3 signaling in B lymphocytes	2.276E-03	<i>ITPR1, AKT3, PLCB4</i>
Development_Growth hormone signaling via PI3K/AKT and MAPK cascades	2.276E-03	<i>AKT3, PTK2B, FAK1</i>
Development_Angiotensin signaling via PYK2	2.436E-03	<i>ITPR1, PTK2B, PLCB4</i>
Immune response_HTR2A-induced activation of cPLA2	2.436E-03	<i>ITPR1, KCNMA1, PLCB4</i>
Immune response_Substance P-stimulated expression of proinflammatory cytokines via MAPKs	2.436E-03	<i>ITPR1, AKT3, PLCB4</i>
Development_S1P3 receptor signaling pathway	2.436E-03	<i>ITPR1, AKT3, PLCB4</i>
Chemotaxis_CCL16-, CCL20-, CXCL16- and CCL25-mediated cell migration	2.436E-03	<i>ITPR1, AKT3, PLCB4</i>
Immune response_CRT2 signaling in Th2 cells	2.603E-03	<i>ITPR1, AKT3, PLCB4</i>
Development_S1P1 signaling pathway	2.603E-03	<i>ITPR1, AKT3, PLCB4</i>
G-protein signaling_Regulation of cAMP levels by ACM	2.776E-03	<i>ITPR1, PDE1C, PLCB4</i>
Development_G-Proteins mediated regulation MAPK-ERK signaling	2.957E-03	<i>ITPR1, PTK2B, PLCB4</i>
Immune response_PGE2 common pathways	3.144E-03	<i>ITPR1, AKT3, PLCB4</i>
Immune response_C3a signaling	3.339E-03	<i>ITPR1, AKT3, PLCB4</i>
Signal transduction_IP3 signaling	3.541E-03	<i>ITPR1, AKT3, PLCB4</i>
Tissue Factor signaling in cancer via PAR1 and PAR2	3.541E-03	<i>ITPR1, AKT3, PTK2B</i>
Development_EDNRB signaling	3.750E-03	<i>ITPR1, AKT3, PLCB4</i>
G-protein signaling_Proinsulin C-peptide signaling	4.191E-03	<i>ITPR1, AKT3, PLCB4</i>
Cell cycle_Influence of Ras and Rho proteins on G1/S Transition	4.423E-03	<i>SKP2, AKT3, FAK1</i>
Immune response_IFN gamma signaling pathway	4.663E-03	<i>ITPR1, AKT3, PTK2B</i>
Immune response_Platelet activating factor/PTAFR pathway signaling	4.910E-03	<i>ITPR1, AKT3, PLCB4</i>
Cell adhesion_Integrin inside-out signaling	5.165E-03	<i>ITPR1, FAK1, PLCB4</i>
Immune response_CCR5 signaling in macrophages and T lymphocytes	5.698E-03	<i>ITPR1, AKT3, PTK2B</i>
Immune response_LTBR1 signaling	6.264E-03	<i>ITPR1, AKT3, PLCB4</i>
Role of Tissue factor-induced Thrombin signaling in cancerogenesis	8.157E-03	<i>ITPR1, AKT3, FAK1</i>
Cytoskeleton remodeling_ESR1 action on cytoskeleton remodeling and cell migration	8.280E-03	<i>AKT3, FAK1</i>
Development_S1P4 receptor signaling pathway	8.280E-03	<i>ITPR1, PLCB4</i>
Proteolysis_Putative ubiquitin pathway	9.032E-03	<i>RARK2, SKP2</i>

[†]Single nucleotide variations and insertions and deletions.[‡]Personal differentially methylated regions.[§]www.genego.com.

After exclusion of pathways specific to organs other than the liver, 46 pathways that were potentially correlated with liver function, epithelial cell function or immune response, and showed a p-value of 0.01 or less, were collected for this table. All 168 significant (p < 0.05) pathways are summarized in Supplementary Table 22.

for a comparison of six samples, the statistical power would not have been conclusive, and therefore this represents a clear limitation of this study. Since it is not easy to obtain totally normal human tissue samples, it seems that in order to obtain a comprehensive picture of totally normal reference hepatocyte epigenomes, a comparison of data for various hepatocytes that have been, or will be, deposited in the IHEC data portal and other databases will be necessary.

DNA methylation profiles such as low CpG methylation levels in TSS1500, TSS200 and CpG island were obtained especially in actively transcribed genes (Figure 1). Even within the first exon, differing from the 5' UTR, the first coding exon in particular was shown to have a low CpG methylation level (Figure 1). Such low CpG methylation in TSS200, the first coding exon and CpG island was also confirmed in an adult liver tissue sample deposited in the Roadmap database [56] (accession number: GSM916049) (Supplementary Figure 6 [page 314]). Additional analysis using the Roadmap database revealed that low CpG methylation in such regions may not be a liver-specific phenomenon, but common in organs other than the liver (Supplementary Figure 6 [page 314]).

Unlike CHG sites, CHH sites are not located symmetrically on the sense and antisense strands along genomic DNA. In addition, CHH methylation was observed more frequently on the antisense strands than on the sense

Figure 6. Significance of *AKT3* and *EDNRA* for hepatocyte function revealed by transfection with siRNA. (A) Knockdown of the *AKT3* gene by siRNA transfection was performed in HCC cell lines KIM-1 and KYN-2, which show higher levels of *AKT3* expression. Knockdown of the *AKT3* gene resulted in increased cell viability ($p = 1.63 \times 10^{-2}$ for KYN-2), reduced apoptosis ($p = 1.77 \times 10^{-5}$ for KIM-1) and enhanced cell migration ($p = 0.77 \times 10^{-2}$ for KYN-2). **(B)** Knockdown of the *EDNRA* gene by siRNA transfection was performed in HCC cell lines KIM-1 and KYN-2, which show higher levels of *EDNRA* expression. Knockdown of the *EDNRA* gene resulted in reduced apoptosis ($p = 4.011 \times 10^{-6}$ for KIM-1) and enhanced cell migration ($p = 0.70 \times 10^{-2}$ for KIM-1 and $p = 0.70 \times 10^{-2}$ for KYN-2). These data indicated that *AKT3* and *EDNRA*, representative genes frequently showing both genetic and epigenetic variations (single nucleotide variation and personal differentially methylated region) and included in the top pathway of Table 2, play a significant role in hepatocyte functions such as proliferation, apoptosis and migration.

HCC: Hepatocellular carcinoma.

Figure 6. Significance of *AKT3* and *EDNRA* for hepatocyte function revealed by transfection with siRNA (cont.). (A) Knockdown of the *AKT3* gene by siRNA transfection was performed in HCC cell lines KIM-1 and KYN-2, which show higher levels of *AKT3* expression. Knockdown of the *AKT3* gene resulted in increased cell viability ($p = 1.63 \times 10^{-2}$ for KYN-2), reduced apoptosis ($p = 1.77 \times 10^{-5}$ for KIM-1) and enhanced cell migration ($p = 0.77 \times 10^{-2}$ for KYN-2). (B) Knockdown of the *EDNRA* gene by siRNA transfection was performed in HCC cell lines KIM-1 and KYN-2, which show higher levels of *EDNRA* expression. Knockdown of the *EDNRA* gene resulted in reduced apoptosis ($p = 4.011 \times 10^{-6}$ for KIM-1) and enhanced cell migration ($p = 0.70 \times 10^{-2}$ for KIM-1 and $p = 0.70 \times 10^{-2}$ for KYN-2). These data indicated that *AKT3* and *EDNRA*, representative genes frequently showing both genetic and epigenetic variations (single nucleotide variation and personal differentially methylated region) and included in the top pathway of Table 2, play a significant role in hepatocyte functions such as proliferation, apoptosis and migration. HCC: Hepatocellular carcinoma.

strands in the 5' UTR, first intron, gene body and 3' UTR (Figure 2C). Such observations are consistent with the more frequent non-CpG methylation on the antisense strand reported previously [57]. It has been reported that non-CpG methylation was significantly correlated with gene expression levels in human H1 embryonic stem cells [57], whereas non-CpG methylation was inversely correlated with expression in adult tissues [58]. Our data for purified hepatocytes are consistent with the latter inverse association. Since CHG and CHH methylation was significantly correlated with higher CpG methylation levels in neighboring regions (Supplementary Table 9 [page 17]), this inverse association with expression levels was not unexpected.

Since epigenome profiles of multiple human normal purified hepatocytes were obtained, we focused on personal variation of CpG methylation profiles. pDMRs were observed less frequently in TSS200, the first coding exon, TSS1500 and CpG island where CpG methylation levels were low, indicating that the regions important for expression regulation may be protected from personal variations of CpG methylation. Unlike the histone modification profiles for which personal variations among samples are shown in Supplementary Figure 11 (page 323), variations of DNA methylation profiles were not obvious among the samples (pie charts in Figure 3B & C).

Based on ChIP-seq data, presumptive histone modification profiles [59], such as the histone H3K4 methylation and H3K27 acetylation predominant active marks in TSS200 and CpG island and the H3K27me3 predominant

suppressive mark in the intergenic region and open sea, were obtained. On the other hand, 114 genes in which TSS 200 regions showing a H3K4me3 and H3K27AC-dominant histone modification pattern (active promoter cluster) and low DNA methylation levels (<0.5) showed low level expression (less than 0.01 RPKM) in two or more samples (Supplementary Table 23 [page 292]). Such genes that appeared to be less reliant on known epigenetic regulation revealed additional complexities of gene regulation mechanisms.

Differing from DNA methylation profiles, considerable personal variations in the histone modification profiles were observed among samples (Supplementary Figure 11 [page 323]), suggesting that histone modification profiles might be easily altered according to physiological conditions. Since histone modification can induce DNA methylation (e.g., histone methyltransferases G9a and SUV39H1 can recruit DNA methyltransferases) [60,61], some of the histone modification alterations based on physiological conditions (personal variations in histone modification patterns) may result in alterations of DNA methylation (personal variations in DNA methylation profiles such as pDMRs). However, the histone modification patterns of pDMRs did not show a common tendency among normal hepatocyte samples (Supplementary Table 16 [page 24]), and the specific histone modifications that precede and/or induce DNA methylation variations among individuals remain unclear.

Associations between CpG methylation variations (pDMRs) and expression variations (personal over- or under-expression) were observed in the genes listed in Supplementary Table 18 (page 281). Among these genes, expression regulation due to CpG methylation of the *LRP1B* [62], *RASGRF* [63] and *TFF1* [64] genes has already been reported in normal and diseased states. Moreover, data deposited in a reliable public database [50,65] validated the significant inverse correlations between DNA methylation levels and expression levels of the *AURKB*, *BOK*, *C11orf21*, *CLEC2B*, *CSDC2*, *CYP4F12*, *GRB7*, *KLHDC7B*, *MYOC*, *NPR2*, *PLL1*, *RASGRF2*, *SLC28A1*, *SLC44A3*, *STEAP2*, *TFF1* and *TMEM105* genes (Supplementary Table 19 [page 282]). Taken together, although further analysis is needed, the possibility that personal variations of CpG methylation (pDMRs) may result in personal variations in the expression levels of specific genes in normal cells can be considered.

In addition to the standard recommendation by the IHEC, we performed WGS using human normal purified hepatocytes and focused on genome–epigenome correlation. It was clearly shown that pDMRs are located more frequently in the vicinity of loci showing genetic variation (SNVs and/or indels) than on all autosomes (Table 1 & Figure 5): 15.7–22.9% of pDMRs and 0.7–2.1% of pDMRs were located in the vicinity of SNVs and indels, respectively (Supplementary Table 20A & B [page 283]). Although further study is needed to clarify the molecular background of such genome–epigenome correlation, we speculate that SNVs and indels may affect the binding of noncoding RNAs and/or protein complexes that induce histone modification or CpG methylation alterations around the SNV and indel loci. If this possibility were to be extended to a larger number of individuals, the association of DNA methylation status with rare SNVs and indels may have resulted in rare phenotypic individuality in this cohort. SNVs and indels showing a certain frequency may have stratified the phenotypes in the cohort through different DNA methylation profiles. In addition, ethnic variation in the frequencies of SNVs and indels may generate interethnic differences in DNA methylation profiles.

On the other hand, it is known that hotspots of DNA methylation alterations are frequently concordant with hotspots of copy number alterations in human cancers. DNA hypomethylation can induce chromosomal instability via alterations of chromatin configuration: mice carrying a hypomorphic allele of the *DNMT1* gene developed aggressive tumors that displayed a high frequency of chromosomal abnormalities [66]. By analogy with such findings, we examined the associations between variations of CpG methylation (pDMRs) and CNVs, and found that CNVs were more frequent in the vicinity of pDMRs, especially pDMRs showing DNA hypomethylation (Table 1), although we cannot conclude with certainty whether this represented a causal relationship or simple concordance.

Genes showing both genetic and epigenetic variations in multiple samples were significantly accumulated in signaling pathways participating in various functions in normal and diseased epithelial cells including hepatocytes (Table 2). Even when we examined genes showing SNVs and/or indels in 5% or more of the Japanese public cohort deposited in the Integrative Japanese Genome Variation Database [67] and pDMRs in our six purified hepatocyte samples (Supplementary Table 24 [page 293]), such genes were accumulated in 43 of the 46 pathways listed in Table 2 (Supplementary Table 25 [page 301]), indicating that the enriched pathways did not solely reflect our samples alone, but were of universal significance in a larger Japanese cohort. Knockdown for *AKT3* and *EDNRA*, representative genes, frequently showing both SNV/indel and pDMR and included in the top pathway in Table 2, resulted in increased cell proliferation, reduced apoptosis and enhanced migration of hepatocellular carcinoma cell lines (Figure 6), indicating that such genes have significant importance for hepatocyte function. Genetic variations

can induce epigenetic variations in a *cis*-acting manner and may generate individual differences in the phenotypes of hepatocytes.

Table 2 also includes the pathways participating in the functions of cells other than hepatocytes, such as inflammatory response pathways. Shi *et al.* reported that methylation quantitative trait loci (meQTL) in the lung were replicated in both breast and kidney tissues [68]. Among the replicated meQTL documented by Shi *et al.*, the *STARTD3*, *BRF1*, *AGA*, *SLC35A3* and *PABPC4* genes were also included in our pDMRs. In addition, Volkov *et al.* [69] and Grundberg *et al.* [70] identified 39,386 *cis*-mQTL in adipose tissue that were also found in pancreatic islets and 31,735 meQTL in adipose tissue that were also replicated in whole blood samples, respectively. These previous reports suggested that genome–epigenome correlations observed in a specific cell lineage, such as our sample of hepatocytes, may be shared by other cell lineages in the same individuals. If the same genome–epigenome correlation shown in Table 2 is established even in other cell lineages, such as those with roles in the immune system, then the interaction may affect a wider variety of cell functions or disease susceptibility.

Conclusion

pDMRs were frequently observed around the TSSs of genes whose expression is reportedly regulated by CpG methylation. In addition, they were frequently observed in the vicinity of SNVs and indels, suggesting possible *cis*-acting genome–epigenome interaction. Genes showing both genetic (SNVs and indels) and epigenetic (pDMRs) variations in multiple samples were accumulated in signaling pathways participating in hepatocyte function and disease susceptibility. These data indicated that genetic variations may induce epigenetic variations, generating individual differences in the phenotypes of normal hepatocytes through variations in expression.

Future perspective

In parallel, Supplementary Table 20 (page 283) revealed that about 80% of pDMRs are located independently from SNVs and/or indels. Moreover, less than 1% of pDMRs (Supplementary Table 11 [page 19]) and no pDMR showed significant correlations with age and sex, respectively. At least a proportion of the age-related pDMRs may reflect the cumulative effects of environmental exposures throughout life. Moreover, the majority of CpG methylation variations (about 80%) may be induced by environmental and/or epidemiological factors including exposure to carcinogens, rather than by genetic variations, aging-related insults and sex. In particular, even though six patients did not have hepatitis virus infection or NASH, their hepatocytes may have been exposed to some degree of inflammation (Table 2) reflecting the lifestyle habits and metabolic state of each patient. Such inflammation-related epigenome profiles may represent a very early phase of inflammation-related diseases. After accumulation of numerous reference epigenome profile data in the IHEC database, such correlations between epigenome variations and environmental factors will be examined further.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at:

<https://www.futuremedicine.com/doi/suppl/10.2217/epi-2017-0111>

Financial & competing interest disclosure

This work was supported by AMED-CREST, Japan Agency for Medical Research and Development (16gm0510003h0006). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

All of the patients included in this study provided written informed consent. This study (2012-095) was approved by the Ethics Committees of the National Cancer Center, Tokyo, and Keio University School of Medicine, and was performed in accordance with the Declaration of Helsinki.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Summary points

- In the course of our research as a member group of the International Human Epigenome Consortium, we subjected hepatocytes purified from six Japanese patients without hepatitis virus infection or hepatocellular carcinoma who had undergone partial hepatectomy to whole-genome bisulfite sequencing using postbisulfite adaptor tagging, chromatin immunoprecipitation–sequencing, RNA sequencing and whole-genome sequencing to reveal the epigenome landscape of normal human hepatocytes.
- CpG methylation levels were low in TSS200, TSS1500 and the first coding exon.
- CHH methylation on the antisense strand was more frequent than on the sense strand. CHG and CHH methylation was correlated with higher methylation at neighboring CpG sites and was inversely associated with gene expression levels.
- Personal differentially methylated regions (pDMRs) were less frequent in TSS200, TSS1500, the first coding exon and CpG island where CpG methylation levels were low, indicating that the regions important for expression regulation may be protected from personal variations of CpG methylation.
- Histone modification profiles of pDMRs differed considerably among samples.
- pDMRs were observed around the transcription start sites of genes whose expression is reportedly regulated by CpG methylation.
- pDMRs were frequently observed in the vicinity of single nucleotide variations (SNVs) and insertions and deletions (indels), suggesting possible *cis*-acting genome–epigenome correlation.
- Genes showing both genetic (SNVs and indels) and epigenetic (pDMRs) variations in multiple samples were accumulated in signaling pathways participating in hepatocyte function and disease susceptibility.
- These data suggested that genetic variations may induce epigenetic variations, generating individual differences in normal hepatocyte phenotypes through variations in expression.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Claycombe KJ, Brissette CA, Ghribi O. Epigenetics of inflammation, maternal infection, and nutrition. *J. Nutr.* 145(5), 1109S–1115S (2015).
2. Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. *Nat. Rev. Cancer* 11(10), 726–734 (2011).
3. Keating ST, El-Osta A. Epigenetics and metabolism. *Circ. Res.* 116(4), 715–736 (2015).
4. Tian Y, Wong VW, Wong GL *et al.* Histone deacetylase HDAC8 promotes insulin resistance and β -catenin activation in NAFLD-associated hepatocellular carcinoma. *Cancer Res.* 75(22), 4803–4816 (2015).
5. Nagashio R, Arai E, Ojima H, Kosuge T, Kondo Y, Kanai Y. Carcinogenetic risk estimation based on quantification of DNA methylation levels in liver tissue at the precancerous stage. *Int. J. Cancer* 129(5), 1170–1179 (2011).
6. Arai E, Ushijima S, Gotoh M *et al.* Genome-wide DNA methylation profiles in liver tissue at the precancerous stage and in hepatocellular carcinoma. *Int. J. Cancer* 125(12), 2854–2862 (2009).
7. Kanai Y, Arai E. Multilayer-omics analyses of human cancers: exploration of biomarkers and drug targets based on the activities of the International Human Epigenome Consortium. *Front. Genet.* 5, 24 (2014).
8. International Human Epigenome Consortium. <http://ihec-epigenomes.org/>
9. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* 167(7), 1145–1149 (2016).
- **Achievements of International Human Epigenome Consortium teams to gather and interpret comprehensive epigenome datasets of major human cell types were introduced.**
10. International Human Epigenome Consortium Team Japan. <http://crest-ihec.jp/english/index.html>
11. Fasano M, Theise ND, Nalesnik M *et al.* Immunohistochemical evaluation of hepatoblastomas with use of the hepatocyte-specific marker, hepatocyte paraffin 1, and the polyclonal anti-carcinoembryonic antigen. *Mod. Pathol.* 11(10), 934–938 (1998).
12. Maslak E, Gregorius A, Chlopicki S. Liver sinusoidal endothelial cells (LSECs) function and NAFLD; NO-based therapy targeted to the liver. *Pharmacol. Rep.* 67(4), 689–694 (2015).
13. Marcos R, Correia-Gomes C. Long live the liver: immunohistochemical and stereological study of hepatocytes, liver sinusoidal endothelial cells, Kupffer cells and hepatic stellate cells of male and female rats throughout ageing. *Cell Tissue Res.* 366(3), 639–649 (2016).
14. Kudryavtsev BN, Kudryavtseva MV, Sakuta GA, Stein GI. Human hepatocyte polyploidization kinetics in the course of life cycle. *Virchows Arch. B Cell Pathol.* 64(6), 387–393 (1993).

15. Nishiyama N, Arai E, Nagashio R *et al.* Copy number alterations in urothelial carcinomas: their clinicopathological significance and correlation with DNA methylation alterations. *Carcinogenesis* 32(4), 462–469 (2011).
16. Miura F, Enomoto Y, Dairiki R, Ito T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* 40(17), e136 (2012).
- **Postbisulfite adaptor tagging, which can generate a substantial number of unamplified reads from as little as subnanogram quantities of DNA, was first conceived for whole-genome bisulfite sequencing.**
17. Post-bisulfite adaptor tagging protocol revision 12. <http://crest-ihc.jp/english/epigenome/index.html>
18. BMap. <http://itolab.med.kyushu-u.ac.jp/BMap/>
19. Reference epigenome criteria of International Human Epigenome Consortium. <http://ihc-epigenomes.org/research/reference-epigenome-standards/>
20. National Bioscience Database Center. <http://humandbs.biosciencedbc.jp/en/>
21. International Human Epigenome Consortium Data Portal. <http://epigenomesportal.ca/ihc/>
22. Takeshima H, Yamashita S, Shimazu T, Niwa T, Ushijima T. The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res.* 19(11), 1974–1982 (2009).
23. Kimura H. Histone modifications for human epigenome analysis. *J. Hum. Genet.* 58(7), 439–445 (2013).
24. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9(9), R137 (2008).
25. Model-based analysis for ChIP-Seq. <http://liulab.dfci.harvard.edu/MACS/>
26. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9(3), 215–216 (2012).
27. ChromHMM: chromatin state discovery and characterization. <http://compbio.mit.edu/ChromHMM/>
28. NCBI Reference Sequence Database. www.ncbi.nlm.nih.gov/refseq/
29. Ensembl Asia. http://asia.ensembl.org/Homo_sapiens/Info/Index?redirect=no
30. Long noncoding RNA catalog. www.broadinstitute.org/genome.bio/human.lincrnas/?q=lincRNA_catalog
31. Bowtie. <http://bowtie-bio.sourceforge.net/index.shtml>
32. Totoki Y, Yoshida A, Hosoda F *et al.* Unique mutation portraits and frequent COL2A1 gene alteration in chondrosarcoma. *Genome Res.* 24(9), 1411–1420 (2014).
33. Burrows–Wheeler Aligner. <http://bio-bwa.sourceforge.net/>
34. Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16), 2078–2079 (2009).
35. SAMtools. <http://samtools.sourceforge.net/>
36. Totoki Y, Tatsuno K, Yamamoto S *et al.* High-resolution characterization of a hepatocellular carcinoma genome. *Nat. Genet.* 43(5), 464–469 (2011).
37. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 21(6), 974–984 (2011).
38. CNVnator. <http://sv.gersteinlab.org/cnvator/>
39. Clarivate Analytics. www.genego.com
40. Lister R, O'Malley RC, Tonti-Filippini J *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133(3), 523–536 (2008).
41. Bibikova M, Le J, Barnes B *et al.* Genome-wide DNA methylation profiling using Infinium[®] assay. *Epigenomics* 1(1), 177–200 (2009).
42. UCSC Genome Browser. <https://genome.ucsc.edu/>
43. Methylome data for Roadmap liver tissue. www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM916049
44. Methylome data for Encode purified hepatocyte. www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM683764
45. Methylome data for Encode purified hepatocyte. www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM683872
46. ChIP-seq data for hepatocyte sample. <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1112nnn/GSM1112808/suppl/GSM1112808%5FBI.Adult%5FLiver.H3K27ac.3.bed.gz>
47. ChIP-seq data for hepatocyte sample. <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1112nnn/GSM1112809/suppl/GSM1112809%5FBI.Adult%5FLiver.H3K27ac.4.bed.gz>
48. LChIP-seq data for hepatocyte sample. <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1112nnn/GSM1112814/suppl/GSM1112814%5FBI.Adult%5FLiver.H3K27me3.4.bed.gz>
49. Ernst J, Kheradpour P, Mikkelsen TS *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345), 43–49 (2011).
50. Integrative Disease Omics Database. <http://gemdbj.ncc.go.jp/omics/>

51. Bahde R, Kapoor S, Viswanathan P, Spiegel HU, Gupta S. Endothelin-1 receptor A blocker darusentan decreases hepatic changes and improves liver repopulation after cell transplantation in rats. *Hepatology* 59(3), 1107–1117 (2014).
52. Lu JW, Liao CY, Yang WY *et al.* Overexpression of endothelin 1 triggers hepatocarcinogenesis in zebrafish and promotes cell proliferation and migration through the AKT pathway. *PLoS ONE* 9(1), e85318 (2014).
53. Dou L, Meng X, Sui X *et al.* MiR-19a regulates PTEN expression to mediate glycogen synthesis in hepatocytes. *Sci. Rep.* 5, 11602 (2015).
54. Sun CK, Man K, Ng KT *et al.* Proline-rich tyrosine kinase 2 (Pyk2) promotes proliferation and invasiveness of hepatocellular carcinoma cells through c-Src/ERK activation. *Carcinogenesis* 29(11), 2096–2105 (2008).
55. Bouchard MJ, Wang L, Schneider RJ. Activation of focal adhesion kinase by hepatitis B virus HBx protein: multiple functions in viral replication. *J. Virol.* 80(9), 4406–4414 (2006).
56. Roadmap database. www.ncbi.nlm.nih.gov/epigenomics
57. Lister R, Pelizzola M, Dowen RH *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271), 315–322 (2009).
- **Genome-wide, single-base resolution maps of methylated cytosine in a mammalian genome was first reported. This paper showed that non-CpG methylation was significantly correlated with gene expression levels in human H1 ES cells.**
58. Schultz MD, He Y, Whitaker JW *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523(7559), 212–216 (2015).
- **Single-base resolution human epigenome maps revealed widespread tissue-specific differentially methylated regions. This paper showed that non-CpG methylation was inversely correlated with expression in adult tissues.**
59. Fischle W, Mootz HD, Schwarzer D. Synthetic histone code. *Curr. Opin. Chem. Biol.* 28, 131–140 (2015).
60. Yang Y, Liu R, Qiu R *et al.* CRL4B promotes tumorigenesis by coordinating with SUV39H1/HP1/DNMT3A in DNA methylation-based epigenetic silencing. *Oncogene* 34(1), 104–118 (2015).
61. Bittencourt D, Lee BH, Gao L, Gerke DS, Stallcup MR. Role of distinct surfaces of the G9a ankyrin repeat domain in histone and DNA methylation during embryonic stem cell self-renewal and differentiation. *Epigenetics Chromatin* 7, 27 (2014).
62. Sonoda I, Imoto I, Inoue J *et al.* Frequent silencing of low density lipoprotein receptor-related protein 1B (LRP1B) expression by genetic and epigenetic mechanisms in esophageal squamous cell carcinoma. *Cancer Res.* 64(11), 3741–3747 (2004).
63. Chen H, Suzuki M, Nakamura Y *et al.* Aberrant methylation of RASGRF2 and RASSF1A in human non-small cell lung cancer. *Oncol. Rep.* 15(5), 1281–1285 (2006).
64. Carvalho R, Kayademir T, Soares P *et al.* Loss of heterozygosity and promoter methylation, but not mutation, may underlie loss of TFF1 in gastric carcinoma. *Lab. Invest.* 82(10), 1319–1326 (2002).
65. Arai E, Gotoh M, Tian Y *et al.* Alterations of the spindle checkpoint pathway in clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Int. J. Cancer* 137(11), 2589–2606 (2015).
66. Gaudet F, Hodgson JG, Eden A *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* 300(5618), 489–492 (2003).
67. Integrative Japanese Genome Variation Database. <https://ijgvd.megabank.tohoku.ac.jp>
68. Shi J, Marconett CN *et al.* Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat. Commun.* 5, 3365 (2014).
69. Volkov P, Olsson AH *et al.* A genome-wide mQTL analysis in human adipose tissue identifies genetic variants associated with DNA methylation, gene expression and metabolic traits. *PLoS ONE* 11(6), e0157776 (2016).
70. Grundberg E, Meduri E *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* 93(5), 876–890 (2013).