

Postnatal development- and age-related changes in DNA-methylation patterns in the human genome

Paraskevi Salpea¹, Valya R. Russanova¹, Tazuko H. Hirai¹, Thomae G. Sourlingas², Kalliope E. Sekeri-Pataryas², Roberto Romero³, Jonathan Epstein⁴ and Bruce H. Howard^{1,*}

¹Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA, ²Institute of Biology, National Centre for Scientific Research ‘Demokritos’, Agia Paraskevi, 153 10, Attikis, Greece, ³Perinatology Research Branch and ⁴Molecular Genomics Laboratory, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received December 8, 2011; Revised March 22, 2012; Accepted March 25, 2012

ABSTRACT

Alterations in DNA methylation have been reported to occur during development and aging; however, much remains to be learned regarding post-natal and age-associated epigenome dynamics, and few if any investigations have compared human methylome patterns on a whole genome basis in cells from newborns and adults. The aim of this study was to reveal genomic regions with distinct structure and sequence characteristics that render them subject to dynamic post-natal developmental remodeling or age-related dysregulation of epigenome structure. DNA samples derived from peripheral blood monocytes and *in vitro* differentiated dendritic cells were analyzed by methylated DNA Immunoprecipitation (MeDIP) or, for selected loci, bisulfite modification, followed by next generation sequencing. Regions of interest that emerged from the analysis included tandem or interspersed-tandem gene sequence repeats (*PCDHG*, *FAM90A*, *HRNR*, *ECEL1P2*), and genes with strong homology to other family members elsewhere in the genome (*FZD1*, *FZD7* and *FGF17*). Our results raise the possibility that selected gene sequences with highly homologous copies may serve to facilitate, perhaps even provide a clock-like function for, developmental and age-related epigenome remodeling. If so, this would represent a fundamental feature of genome architecture in higher eukaryotic organisms.

INTRODUCTION

Aging is widely viewed as a complex process, distinct from development and characterized by accumulation of damaged molecules, cells and tissues, ultimately leading to malfunction and degeneration (1,2). Associated with old age is loss of the ability to maintain baseline tissue homeostasis and an increased risk for developing many diseases, including cancer, cardiovascular and neurodegenerative disorders (3,4). The aging phenotype varies among individuals, in part due to genetic factors, but also due to environmental influences and stochastic variation in epigenome structure (5,6).

Characterization of epigenome dysregulation with age is at a very early stage, and an even less explored area concerns the underlying role of epigenetic remodeling during post-natal development, despite numerous lines of evidence linking development and age-related variation, as well as age-related phenotypes (7–10). In mice, similarities have been reported between developmental and adult changes in DNA-methylation patterns (11). Furthermore, cumulative evidence links perinatal events to risk patterns for age-related diseases (12,13). Taken together with increasing focus on stem-cell integrity in the context of aging (14–17), such studies suggest that it will be worthwhile to investigate these links further, as well as to reexamine the widely accepted dichotomy between peri- and post-natal development versus aging.

DNA methylation, a major epigenetic mechanism in mammals, constitutes a post-replication modification at the 5'-position of cytosine, predominantly within CpG dinucleotides, that typically results in transcriptional repression of adjacent genes. DNA-methylation changes during early mammalian development, starting with

*To whom correspondence should be addressed. Tel: +1 301 496 9038; Fax: +1 301 480 9354; Email: howard@helix.nih.gov

demethylation during cleavage of the early embryo and followed by genome-wide *de novo* methylation after the embryonic implantation (18). A substantial fraction of CpG dinucleotides are clustered in domains termed 'CpG islands' (19). In normal cells, CpGs located in gene promoter-associated CpG islands tend to be unmethylated (20), whereas methylation predominates in most downstream transcribed gene regions (21–23) and in repetitive regions (satellite DNA, LINEs, SINEs, etc.) (24). Less dense CpG clusters than those in CpG islands, as well as sequences immediately adjacent to islands ('shores') are thought to exhibit differentiation-specific—but also highly variable—patterns of DNA methylation (25). Additionally, in differing immune cell types, differential methylation can occur in CpG islands that are not associated with transcription start sites (orphan CpG islands) (26).

In rodents, two early studies, one in mouse tissues and another in germ and liver rat cells, revealed consistent age-related increases in DNA-methylation of ribosomal genes (27,28). These studies suggest that ribosomal DNA repetitiveness constitutes a genomic structure susceptible to age-related hypermethylation. More recently, array-based epigenetic studies in mouse and rat tissues have reported hypomethylation or hypermethylation of specific genomic loci during aging (29,30).

A number of studies have examined selected genes or used array-based approaches to document DNA-methylation drift with age in humans. Hypermethylation occurs in the estrogen receptor gene CpG island in human colon samples with cancer, as well as with age (31). Likewise, genes hypermethylated in human prostate cancer are subject to progressive methylation in normal prostate tissues during aging (32). Given high variability of DNA methylation in many genome regions, including in monozygotic twins (33), several recent reports have presented population-based statistical analyses in support of age-related epigenetic disorder or drift (34–37).

In the present work, we compared methylome structures in DNAs from newborns and adults, with samples derived from cord blood and adult peripheral blood, respectively. Methylated DNA Immunoprecipitation followed by Next Generation Sequencing (MeDIP-Seq) was employed to search for age-related methylation differences on a whole genome scale. A notable advantage of this approach is that it allows a global survey of epigenetic marks without bias toward a specific genomic region, e.g. gene promoters or CpG islands.

Our aim was to reveal key genomic regions with distinct structure and sequence characteristics that are susceptible to developmental remodeling or to age-related dysregulation. As an experimental source we chose primary human peripheral blood monocytes. These cells are attractive for epigenome mapping, since they retain stem-cell-like characteristics (38–41) and can be easily differentiated into dendritic cells (DC). Reproducible changes detected across the span of post-natal development should be of strong inherent interest. Moreover, we reasoned that such changes, if detected, might shed light on more strictly age-related (young versus old adult) epigenome perturbations. To our knowledge, this is the first whole

human genome DNA-methylation study to provide gene locus-specific results from comparisons between different age groups.

MATERIALS AND METHODS

Ethics

All human materials used in this study were received under approval of the Institutional Review Board of the National Institute of Child Health and Human Development (NICHD), or of the Clinical Center, National Institutes of Health (NIH).

Human peripheral blood monocyte samples

Cord blood samples were obtained from healthy term newborns. Adult peripheral blood or monocyte-enriched apheresis samples were obtained from healthy volunteers from the NIH Department of Transfusion Medicine. In each of two independent experiments, cord blood and adult blood samples were pooled. For the first experiment, cord blood monocytes were derived from a set of five male and four female newborns; for the second experiment, the pooled set consisted of three male and seven female newborns. The corresponding adult sets included three male and six female donors in the first experiment, versus six male and four female in the second. The age range for adult samples in these experiments was 21 to 73 years old.

Cell purification and culture

Cord blood and adult (monocyte-enriched apheresis or whole blood) samples were diluted 2- to 4-fold in Dulbecco's PBS (DPBS) with 4% citrate dextrose solution (ACD). Samples were placed on a Ficoll-paque gradient (density 1.077) and centrifuged for 35 min at 900 *g* at room temperature. The interphase was transferred to a fresh tube, diluted to 40 ml using DPBS and centrifuged twice for 5 min at 200 *g* to reduce platelet contamination. Samples were resuspended in 20 ml of RPMI 1640, 10% fetal calf serum (FCS) and placed on a lower density Ficoll-paque gradient (20 min, 400 *g*, room temperature) (42). Interphase cells were collected and washed in DPBS followed by centrifugation for 8 min at 500 *g*. Pellets were resuspended in 20 ml PBE (PBS, 0.5% BSA, 2 mM EDTA) and centrifuged for 10 min at 500 *g*. Monocytes were further purified by subtractive magnetic cell sorting using a monocyte isolation kit (Miltenyi, Auburn, CA) following the manufacturer's protocol. Purified monocytes were washed in DPBS. To obtain DC, cells were plated in RPMI 1640, 10% FCS, 50 ng/ml GM-CSF, 20 ng/ml IL-4. Cultures were refed every other day and DC were harvested on Day 6. Purities of monocyte and DC preparations were assessed by flow cytometry (FACSCalibur; BD Biosciences, San Jose, CA) with the following antibodies: anti-CD14 FITC, anti-biotin PE (Miltenyi); anti-HLA-DR and anti-CD86 (BD Biosciences).

Methylated DNA immunoprecipitation

Methylated DNA immunoprecipitation (MeDIP) was performed as previously described in (43). We used 3–5 μ g of fragmented DNA for a standard MeDIP assay. We denatured the DNA for 10 min at 95°C and immunoprecipitated it for 2 h at 4°C with 10 μ l monoclonal antibody against 5-methylcytosine (Eurogentec) in a final volume of 500 μ l IP buffer [10 mM sodium phosphate (pH 7.0), 40 mM NaCl, 0.05% Triton X-100]. We incubated the mixture with 30 μ l of Dynalbeads with M-280 sheep antibody to mouse IgG (Dynal Biotech) for 2 h at 4°C and washed it three times with 700 μ l IP buffer. We then treated the beads with proteinase K for 3 h at 50°C and recovered the methylated DNA by phenol–chloroform extraction followed by ethanol precipitation.

Preparation and immunoprecipitation of formaldehydefixed chromatin

Purified monocytes or *in vitro* differentiated DCs were crosslinked with 1% formaldehyde solution. Nuclei were isolated and micrococcal nuclease (MN) digestion was carried out as described previously (44). The soluble chromatin fragments were diluted in ChIP buffer (50 mM Tris–HCl, pH 7.5, 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 5 mM Na–butyrate, protein inhibitors) to a final SDS concentration 0.1% and were used for Chromatin Immunoprecipitation (ChIP). ChIP was performed as described in previous studies (44,45) using polyclonal anti-trimethylated H3K27 antibody (Millipore #07-449).

MeDIP and ChIP sequencing library preparation

Sequencing libraries were prepared following Illumina ChIP Sequencing protocol. The DNA was end-repaired and ligated with Illumina sequencing paired end adaptors as described previously (46). After ligation, the DNA was enriched by 18 cycles of PCR with primers complementary to the paired end adaptor sequences followed by agarose gel size selection and purification.

Targeted bisulfite sequencing

DNA samples were bisulfite converted by using EpiTect Bisulfite Kit (Qiagen). Primer pairs were designed using a custom interface with primer3. Candidate initial and nested primers for upper and lower strands of the bisulfite-converted DNA template overlapped no more than one CpG, at a position ≥ 10 bases from the primer 3'-end. Where an internal primer base aligned to a potentially methylated CpG template cytosine, the base 'C' was designated in the primer, forcing either a C–T or C–C mismatch. Specificities and yields of the PCR reactions (initial followed by nested PCR) were analyzed by DHPLC (44) and real time PCR for 55 and 46 candidate primer sets targeted to CpGs within the ECEL1P2 and FZD1 regions, respectively. After final PCR products were combined based on primer validation and reaction yields, libraries were prepared as described above for

next-generation sequencing. Reads were initially aligned using Bowtie (47) to bisulfite-converted template databases with duplicate entries containing either CG or TG at all CpG positions. Alignment parameters (seed length = 18, maximum mismatches within seed = 3, maximum total quality scores at all mismatched read positions = 200) were based on pre-analysis confirming ≤ 6 CpGs/18 bp and ≤ 9 CpGs/36 bp read for all possible read positions within the regions of interest. In subsequent summation after string match alignment to score C/T read bases opposite 'Y' characters in a second template representation, positions were reported only for C+T calls ≥ 40 . Mismatched internal primer bases returned either a G/Y or CC/YG string match, and so could be discarded. Mean (geometric) C+T call depths for the four data sets reported here were ≥ 750 .

Illumina genome analyzer sequencing

Two lanes per sample of 36-bp single-ended sequencing were performed on the Illumina Genome Analyzer II according to the manufacturer's protocol. The sequencing was done at the Los Alamos National Laboratory (Los Alamos, NM 87545) and at the NIH Intramural Sequencing Center (Rockville, MD 20852).

Quantitative real time PCR

Selected samples were reverse transcribed in reactions containing 300 ng total RNA [TaqMan Reverse transcription Reagents, Applied Biosystems, Foster City, CA (ABI)]. RNA levels were assayed using a TaqMan Gene-expression Assays (ABI) (*FZD1*: Hs00268943_s1, *FZD7*: Hs00275833_s1, *TRIP6*: Hs01547009_m1, *FGF17*: Hs00182599_m1) according to the manufacturer's protocol and standardized against GAPDH levels (TaqMan Human GAPDH).

Genomics analysis

This study utilized the high-performance computational capabilities of the Helix Systems (<http://helix.nih.gov>) and the Biowulf Linux cluster (<http://biowulf.nih.gov>) at the NIH, Bethesda, MD. Bioinformatics applications blast, RepeatMasker (Smit, AFA, Hubley, R and Green, P. *RepeatMasker Open-3.0*.1996–2010 <<http://www.repeatmasker.org>>) and Bowtie (47) were applied as described in Results. Perl scripts and C programs for additional genomics analyses are provided in the Supplementary Materials or are available on request. Public Database Accession Number [Sequence Read Archive (SRA) division of GenBank]: SRP007666.1.

RESULTS

To investigate structural and dynamic features of the human methylome, we performed MeDIP-Seq. Two independent experiments each involved paired sets of pooled monocyte DNA samples from cord blood samples and from adult blood donors (21–73 years). Taken together, a total of 38 individuals contributed to generate four sample sets. We reasoned that pooling samples would minimize interference due to high inter-individual

variation in local DNA-methylation patterns. In support of this strategy, a similar approach has been reported to yield the comparable results to those obtained from individual samples (48).

Genome-wide methylation analysis

To gain a graphic overview of the results, enrichment patterns were first analyzed at very low (2 Mb) resolution. MeDIP output from reads aligned to single copy (RepeatMasker-excluded) DNA were visualized by a heat map representation of the human chromosomes (Figure 1A, panels a and b). This depiction reveals a striking feature of the organization of the human genome. In many chromosomes, multi-megabase telomere-proximal regions are enriched relative to the input control, whereas under-representation is more common for the internal regions of the chromosome arms. Variation roughly along these lines was noted earlier in fluorescence-based chromosome imaging studies (49). Interestingly, a heat map plot of CG dinucleotide densities, corrected for the

single copy (RepeatMasker-excluded) DNA content in 2-Mb windows, generates an almost identical pattern (Figure 1A, panel c). The most straightforward interpretation of these results is that fractional CpG methylation is nearly constant on a 2-Mb scale, so here MeDIP output largely reflects CpG content. Comparison of the two experimental heatmaps reveals only minimal large scale differences between the newborn and adult patterns (Supplementary Figure S1A).

Discrimination between non-methylated and methyl-CpG dinucleotides is shown in Figure 1B. This shows over/under-representation versus input for single copy DNA-aligned reads, with positions plotted relative to sense-oriented transcription start sites for all genes. After adjusting for the CpG content of these regions (Supplementary Figure S1B), the estimate for fractional DNA methylation around transcription start sites is clearly low, consistent with the known undermethylation of CpG islands located within promoters (20).

Interspersed repetitive sequences in the genome such as Alus, LINES, SINEs are generally methylated in normal

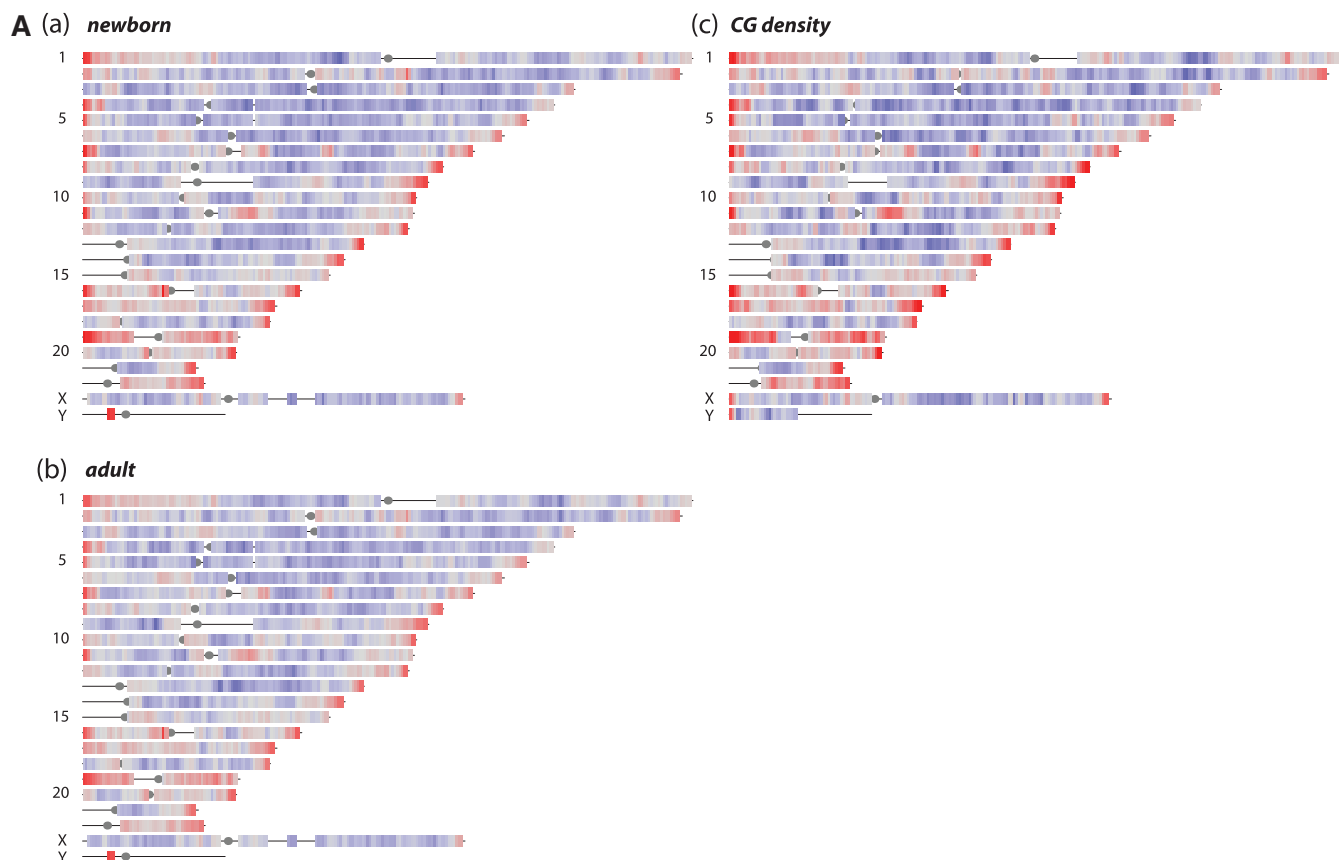


Figure 1. (A) (a and b) DNA-methylation levels on a whole genome scale. Sequence reads from precipitated MeDIP DNA were analyzed with a 2-Mb genomic window (1-Mb steps) and aligned to single copy (RepeatMasker excluded) regions of the human genome, build 36.1. Red color shows ≥ 4 -fold over-representation, and blue ≤ 4 -fold under-representation for MeDIP signals compared to input control DNA. (c) CG dinucleotide densities in single copy regions relative to genome average using same color scale. (B) Over- or under-representation of MeDIP sequence reads flanking transcription start sites ($N = 28\,939$) after normalization for CpG content [values in (a) in Supplementary Figure 1B] \times ave/CpG count ratio from (b) in Supplementary Figure 1B). Gray, newborn; black, adult; arrow shows gene orientation. (C) DNA-methylation levels of interspersed repetitive sequences (IRS). The graph represents estimated DNA-methylation levels from MeDIP-Seq for different families of IRS. Enrichment is compared to the control. Columns fully within gray area indicate depletion, i.e. apparent methylation levels lower than in control DNA (≤ 1 , x-axis). Numbers: see key on the right.

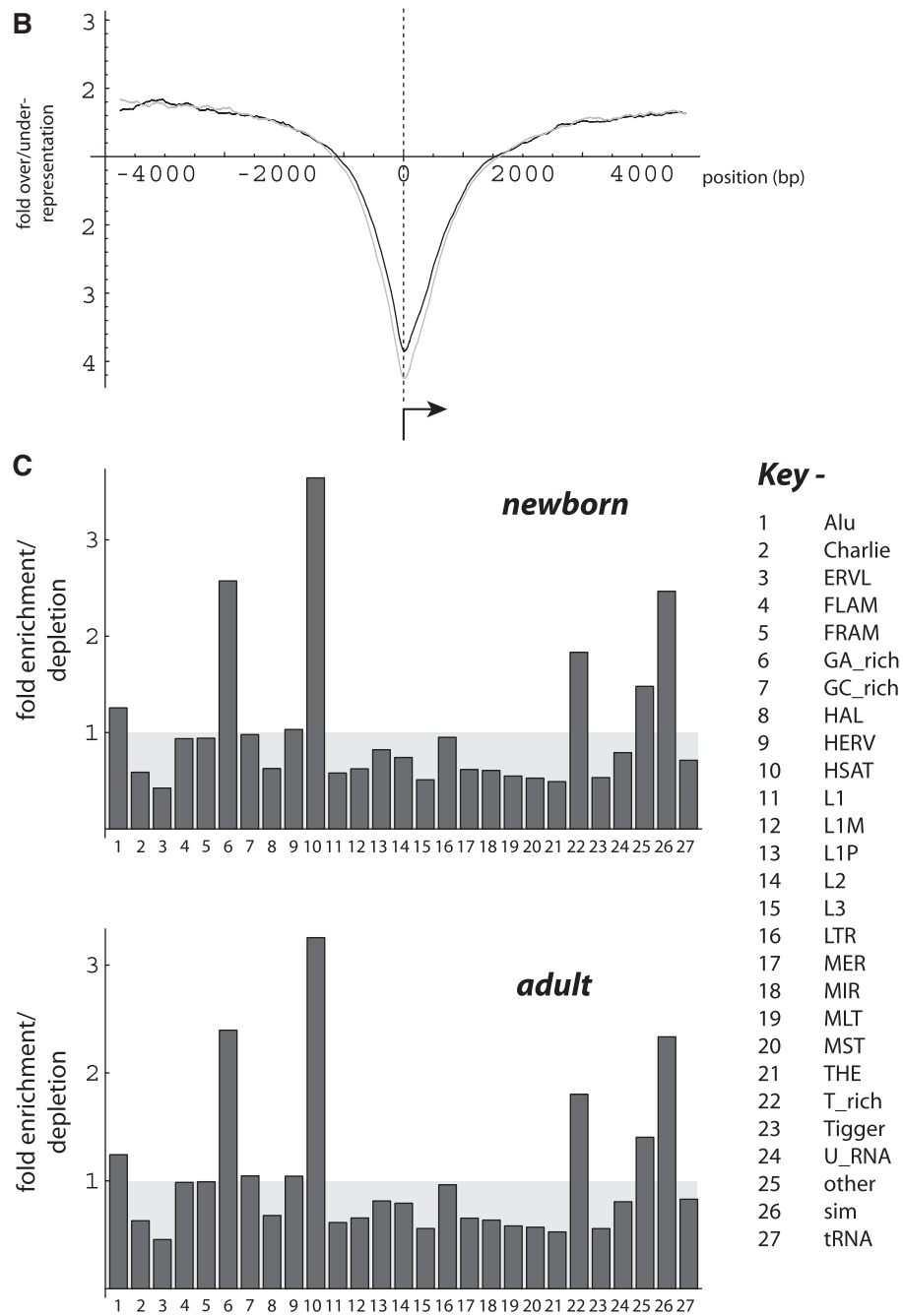


Figure 1. Continued

tissues (24), and can be hypomethylated in cancer cells (50,51). We examined the MeDIP enrichment or depletion relative to input for repetitive sequences grouped into 27 subsets (Figure 1C; shown are averages from the two independent experiments). Apparent DNA methylation levels vary for different subsets of repeated sequences, e.g. high in GA/CT-rich regions and in human satellite DNA, but no significant differences are evident between newborn and adults (it remains possible, of course, that locus-specific age-related differences may occur in such repeats).

Search parameters and variation in detection of methylation differences

Turning to the main focus of the analysis, we next compared MeDIP signals between newborn and adult samples using a sliding 2-kb genomic window. The number of regions sampled in this whole-genome scan is approximately 15 million (window position incremented by 200 bp), but the number of candidate loci for development- or age-related change was found to be strongly dependent on the search parameters selected. In Table 1

Table 1. Genes located within 50 kb of age-related change regions

Symbol	ChrNo	Pos	Fold enrichment	Fold change (nb/ad)	Distance from TS (bp)
ECEL1P3	chr2	232924316	9.58	0.14	35284
ALPP	chr2	232951704	9.58	0.14	7896
ECEL1P2	chr2	232959998	9.58	0.14	398
ALPPL2	chr2	232979797	9.58	0.14	20197
SLC1249	chr7	100288294	6.62	9.33	12906
TRIP6	chr7	100302981	6.62	9.33	1781
ARS2	chr7	100310671	6.62	9.33	9471

Threshold criteria: ≥ 5 -fold change and ≥ 6 -fold enrichment in each of three data sets, where fold enrichment is the higher of the newborn(nb)–adult(ad) paired values. ChrNo, chromosome number; Pos, gene position (human genome build 36.1); TS, transcription start.

are results obtained with very strict parameters: (i) consistent ≥ 5 -fold newborn to adult change in each experiment, and ii) read values elevated in either newborn or adult by ≥ 6 -fold relative to the 2-kb mean. Only two regions were returned from this search, which are listed in the table together with distances to nearby genes. Further details on these regions are described below.

Several strategies were adopted to maximize the number of loci detected while optimizing the signal-to-noise ratio. One option (see following) was to search for changes in MeDIP enrichment extending over larger regions. An alternative option was to include additional data sets. An independent library and sequencing results were available from *in vitro*-derived DC taken from the same pools of 10 newborns and 10 adults used in the second experiment. During *in vitro* differentiation, monocytes do not divide at a detectable level (unpublished results) (52,53), so it is unlikely that substantial changes in methylation patterns occur. We reasoned that inclusion of the DC data could be helpful specifically to minimize experimental error associated with library amplification and read sampling. This further provided three biological replicas, which were used to calculate the error bars shown in subsequent figures.

Figure 2 shows several analyses to assess agreement between experiments at 2-kb resolution. With respect to enrichment relative to the mean, the reproducibility with which specific subregions are identified clearly rises as the threshold is increased (Figure 2A). Plotting 173 K paired enrichment values in a binned heatmap format (Figure 2B) provides a measure of the degree of variation observed (correlation coefficient, $r = 0.61$). As anticipated, high scatter is encountered in calculated post-natal- or age-related change values; however, within a relatively narrow range of parameters, it is possible to detect small numbers of regions that exhibit a high likelihood for consistent change in methylation levels (Figure 2C and D; $P < 0.0001$ for both $r = 0.66$ and $r = 0.79$). A complementary approach to evaluate agreement between experiments is presented in Supplementary Figure S2.

Methylation differences extending over larger regions

Since epigenome domains are thought to extend in many instances over regions of 10s of kilobases or more, we examined the sequencing data with the scanning window

size varied incrementally ≤ 200 kb. What emerged from this approach, most strikingly at the 100-kb setting (Figure 3), was detection of a single region that encompasses members of the protocadherin G family (*PCDHG*). No *PCDHG* RNA signal could be detected in monocytes (unpublished results); however, *PCDHG* genes are expressed in the nervous system and encode for cell surface proteins, which play an important role in neuronal connectivity (54). This clustered family displays an unusual organization that is reminiscent of the immunoglobulin and T receptor loci, having multiple variable upstream exons spliced to a constant downstream exon (55,56).

We posited that the distinct structure of the *PCDHG* family cluster might reveal features that render certain genome regions susceptible to age-related epigenomic dysregulation. As seen in Figure 3, an increase in MeDIP enrichment with age spans a region containing *PCDHG* cluster variant coding exons. These coding exons contain numerous CpG islands and islets (overall content $\sim 38\%$), and remain undermethylated in both newborn and adult samples. A blast search revealed that there are extensive subregion homologies in the *PCDHG* cluster, primarily embedded within the variant coding exons (Figure 3). Further examination demonstrated that the regions interspersed between the exons have transition level CpG densities, and that is these latter regions, positioned between or flanking the highly homologous coding exons, which show evidence of methylation changes with age.

Clustered gene families and genes associated with tandem arrays

Motivated by the characteristics of the *PCDHG* gene cluster, we sought a general means to delineate regions that are normally classified as ‘single copy,’ but in fact have a high degree of homology with one or more regions elsewhere in the human genome. A computationally intensive genome-wide blast cross-homology search was deemed impractical, so virtual ‘sequence reads,’ generated in a uniform tiling pattern across the genome, were aligned using the Bowtie application (47). Such ‘reads’ were retained if: (i) they aligned to more than one genome position (matched at 35/36 positions) and (ii) were devoid of repetitive sequences as classified by RepeatMasker. The subset of sequences so defined (Multiply-Alignable RepeatMasker eXcluded, MAREX) could then be used to annotate the human genome for further analyses. Supplementary Figure S3 compares blast cross-homology output to regions demarcated by the virtual read/Bowtie alignment strategy, showing that similar patterns are produced for the *PCDHG* gene cluster. At the level of homology required here, MAREX annotated regions comprise 2.7% of the human genome.

We next analyzed the experimental data using a 2-kb sliding window, but scoring only sequence reads that partially or completely overlap a MAREX domain. Two examples of the results obtained are described. First, Gene family with sequence similarity 90 (*FAM90A*) (Figure 4) appeared to decrease modestly in DNA

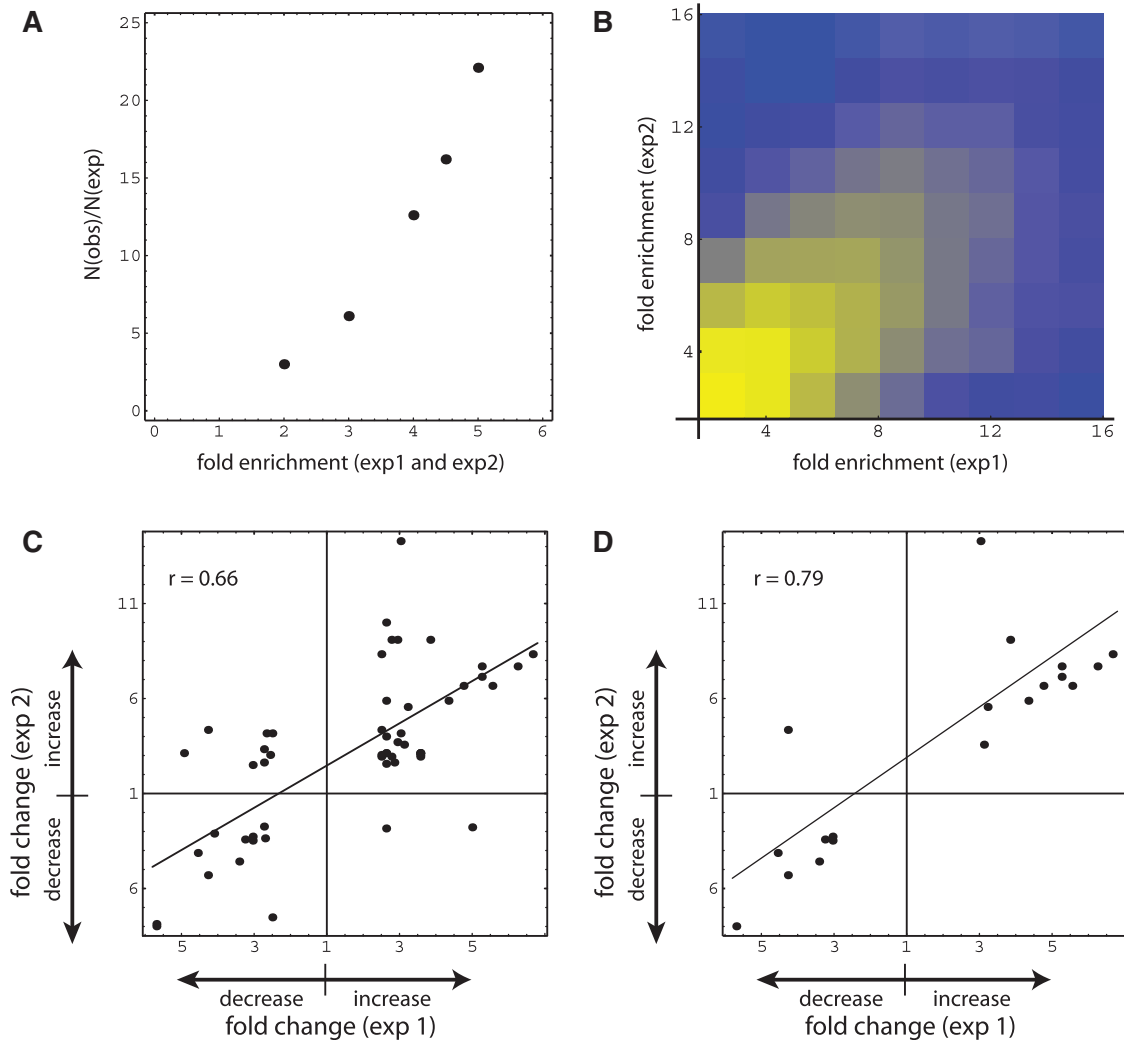


Figure 2. Experimental agreement between data sets. Enrichment and change values compare data from the first and second independent experiments, where values from the second experiment were derived by averaging the monocyte and dendritic cell data. (A) Abscissa values represent enrichment thresholds (region scored if at least one value \geq threshold in newborn-adult comparison). For axis values, $N(\text{exp})$ was calculated from the product of the fraction of regions scored in each experiment multiplied by regions sampled (1.5×10^7). $N(\text{obs})$ is the number of regions exceeding threshold in both experiments. (B) Enrichment relative to the mean as in (A), with values in range 2–16 binned and depicted in a heatmap representation. (C) Correlation plot with thresholds for enrichment parameter ≥ 4.5 and fold-change ≥ 2.5 ; regions plotted, $n = 56$. Probability (P) calculated from the Mathematica expression: $\text{Abs}[r] \text{Sqrt}[(n-2)/(1-r^2)] > \text{Quantile}[\text{StudentTDistribution}[n-2], 0.999]$. (D) Correlation plot with thresholds for enrichment parameter ≥ 4.5 and fold-change ≥ 3 ; regions plotted, $n = 20$.

methylation with age. This cluster of genes was created by multiple replication and rearrangement events (58), and contains a very high content of multiply-alignable domains within a 375-kb region on chromosome 8. While the magnitude of the decrease is small, the downward drift extends across two relatively large regions. A second interesting example is provided by the hornerin gene (*HRNR*) (Figure 5). *HRNR* encodes for a novel profilaggrin-like protein reported to be a differentiation-specific marker in mouse skin (59). In humans, it is expressed in psoriatic and regenerating skin after wounding (60), and it may be present at reduced levels in some cases of atopic dermatitis (61). Within the *HRNR*-coding region, a 4.4-kb region exhibited consistent increases in MeDIP signals with age in each of three data sets. This change region coincides with a set of tandem repeats that encode repeating *HRNR*

protein domain units. The region of putative methylation instability is rich in CpG micro-clusters (see Figure 3 legend for definition), but less strikingly so in CpG islands or islets. Of note is that neither of the above examples would have been found if the analysis had been limited to uniquely aligned reads. Whether conventional—but lower copy and more complex—repeats, are likewise subject to change, and whether these can be effectively probed with longer or paired end reads, are interesting issues for future study.

Examples of gene specific age-related methylation differences

Returning to the seven genes in two regions listed in Table 1, MeDIP enrichment patterns were examined in further detail. On chromosome 7, a region exhibiting a ≥ 4 -fold age-related decrease in each of three data sets is

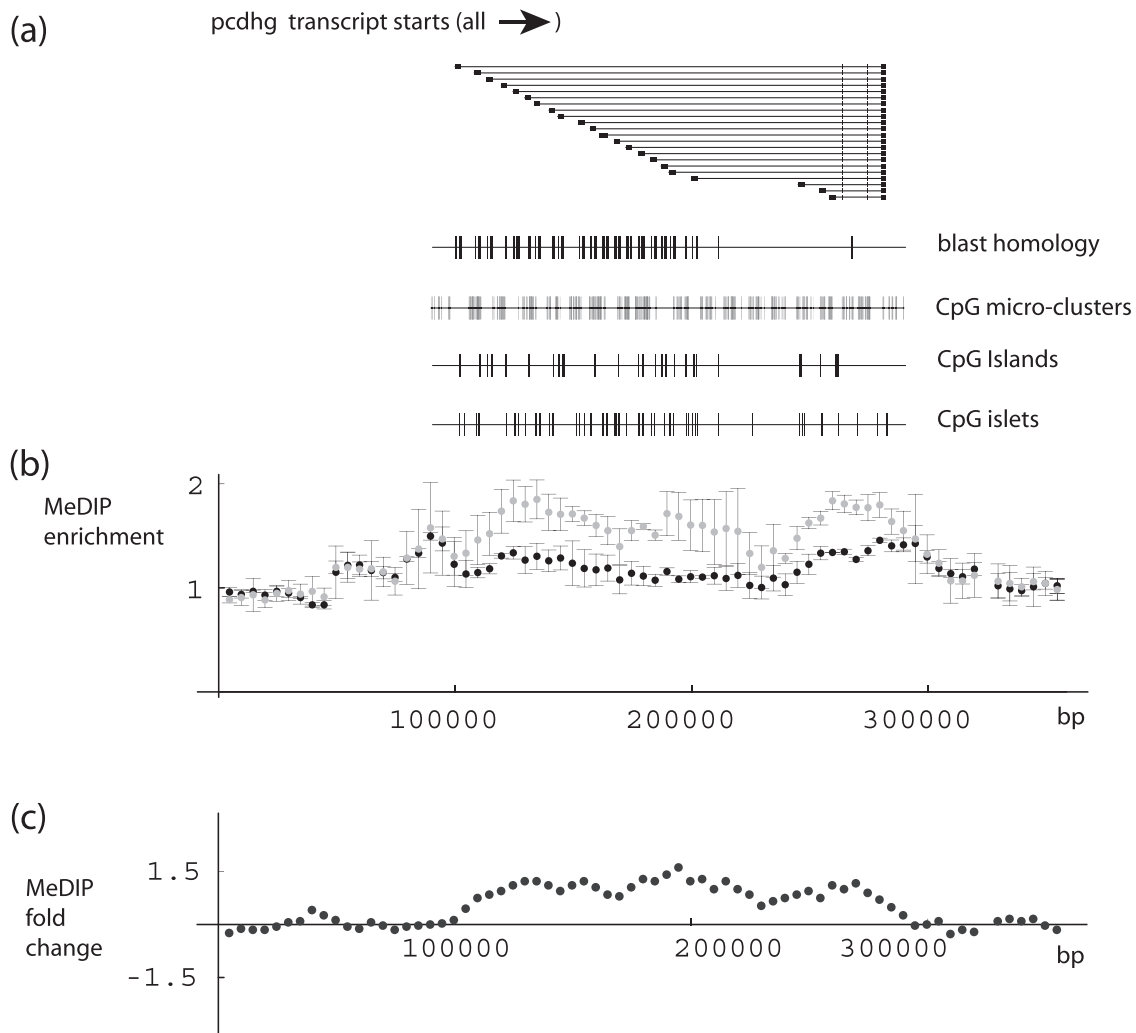


Figure 3. DNA-methylation levels and structure of the *PCDHG* gene cluster. (a) Schematic diagram of multiple protocadherin gamma transcripts (human genome build 37.2), together with diagrams showing positions of blast homology domains, CpG micro-clusters, CpG islands and CpG islets within 200-kb *PCDHG* gene cluster region. CpG islands were calculated according to the original algorithm (57), whereas islets were defined as smaller domains: 200–300 bp versus ≥ 300 bp for standard islands. CpG micro-clusters were defined by reducing the window size in the standard CpG island algorithm to 50 bp, and requiring the cluster length to fall in the range of 50–200 bp (note that this allows overlap with CpG islets and islands, if the latter contain inhomogeneities at the 50-bp level). Blast search queries for 200-kb region, and corresponding blast database, consisted of 200-bp subregions overlapped by 100-bp steps; search threshold 1×10^{-25} . (b) MeDIP enrichment within the *PCDHG* gene cluster. Each dot represents relative enrichment (read number for specific region (50-kb window; 5-kb step) relative to average reads for given window size). Error bars for each dot calculated for values from three data sets. Black: newborn; gray: adult. (c) DNA-methylation fold change between newborn and adult sample. Each dot represents average change value for the corresponding position in (b).

located within the *SLC12A9* gene, near (1800-bp upstream from) the *TRIP6* transcription start (Figure 6 and Supplementary Figure S4A). This region is devoid of CpG islands, but does contain a single islet. The entire 40-kb region plotted is rich (37%) in CpG micro-clusters (versus 9% for the entire human genome), although these do not coincide noticeably with the change region.

The second region of age-related change in Table 1, ≥ 5 -fold in each data set and 7-fold on average, encompasses a cluster of genes on chromosome 2 generated by multiple genome duplications (62,63). Endothelin converting enzyme-like 1 pseudogene 2 (*ECELIP2*) is situated between two alkaline phosphatase genes, the placental (*ALPP*) and the placental-like 2 (*ALPPL2*). The latter has also been termed germ-line alkaline phosphatase

(64). There is a 7-fold increase in methylation through development across a 3.8-kb region that covers the whole *ECELIP2* gene, including CpG island, islet and micro-clusters (Figure 7A and B). Plotted in this figure are MAREX domains, a subset of which coincides with the *ECELIP2* gene. Examination of Bowtie match patterns revealed, as expected, that the multiple alignments were due to *ECELI* homologs within the chromosome 2 cluster. Additionally plotted are monocyte histone H3–K27 ChIP-seq patterns, which show very substantial enrichment peaks aligned to the *ECELI* gene homologs, with a strong age-related decrease for *ECELIP2*. The H3–K27 ChIP-seq data were derived from different individuals (two newborns versus two adults) than those used for MeDIP-seq.

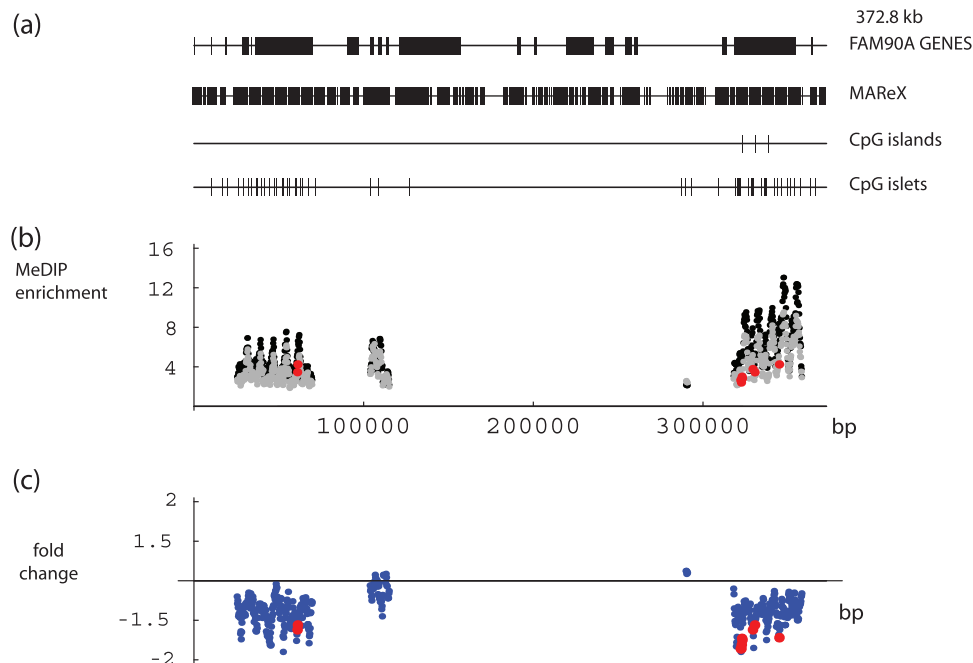


Figure 4. DNA-methylation levels in the *FAM90A* gene family region. (a) *FAM90A* gene family structure represented along a region spanning 372.8 kb, including multiply-alignable (MAREX) domains, CpG Islands and CpG islets. (b) MeDIP enrichment across the *FAM90A* gene family. Values calculated from reads that partially or completely overlap MAREX-annotated genome domains. Data pairs shown only for regions where one or both values ≥ 4 -fold enriched. Black: newborn; gray: adult; red: ≥ 1.5 -fold age-related difference in each of three data sets. (c) Apparent fold change in DNA methylation between newborn and adult, values from (b). Red: ≥ 1.5 -fold age-related difference in each of three data sets.

To extend the detection of age-associated methylation alterations associated with nearby genes, we scanned the genome with search parameters set to ≥ 6 -fold MeDIP enrichment and ≥ 2 -fold age-related change in apparent methylation levels in all three data sets. This yielded six additional newborn/adult change regions encompassing 10 genes that have transcription start sites located within a distance of 10 kb (not shown). From this group, fibroblast growth factor 17 (*FGF17*) ranked just after the *ECELIP2* region in terms of the average fold-change. Downstream of the 3'-end of *FGF17* at a distance of 930 bp, there is a 2.2-kb region in which DNA methylation is on average 5-fold higher in the adult samples compared to newborns (Figure 8 and Supplementary Figure S4B). This region contains a CpG islet and multiple CpG micro-clusters, and it also falls within a broad peak of H3-K27 enrichment centered on the *FGF17* gene.

Lastly, Frizzled homolog 1 (*FZD1*) (Figure 9 and Supplementary Figure S4C) and frizzled homolog 7 (*FZD7*) (Figure 10 and Supplementary Figure S4D) contained regions with 3- and 5-fold increases in apparent DNA methylation with age, respectively. Frizzled genes encode for transmembrane domain proteins that are receptors for Wnt-signaling proteins. Through the WNT [Wg (wingless) and Int] canonical pathway, FZD products are implicated in cell fate determination, and through the non-canonical pathway they are linked to the early mammalian development (65,66). These regions in both cases include CpG island, islet and micro-cluster features, and striking H3-K27 enrichment peaks are

again evident. MAREX domains, which occur within the FZD-coding regions, are due to stretches of high sequence homology shared by *FZD1*, *FZD2* and *FZD7*.

While not the primary focus of this study, RNA measurements were done for several loci. As shown in Supplementary Figure S5A, *FZD1* RNA levels increase with age, whereas *FZD7* (Supplementary Figure S5B) and *TRIP6* RNA (Supplementary Figure S5C) levels exhibit no substantial differences between the two age groups. *FGF17* was found to be expressed at such low levels in the monocytes that RNA estimates were considered to be unreliable. Determination of *ECELIP2* and *PCDHG* gene expression was not attempted, since high homologies between different members of these gene clusters render the development of transcript-specific assays difficult.

Validation of the differentially methylated regions with age by bisulfite sequencing

To validate representative MeDIP-Seq results, we employed targeted bisulfite sequencing, which provides single base resolution and is more quantitative than MeDIP. Two 2.2-kb domains were selected, overlapping change regions associated with the *ECELIP2* and *FZD1* genes. Pooled monocyte DNA samples from newborn and adult donors were from the same individuals as those used in the initial MeDIP-Seq experiment. To amplify the bisulfite-converted DNA regions of interest, we performed PCR reactions using primers targeted to bracket CpG dinucleotides, followed by a second amplification step with nested primer pairs. Yields of PCR reactions

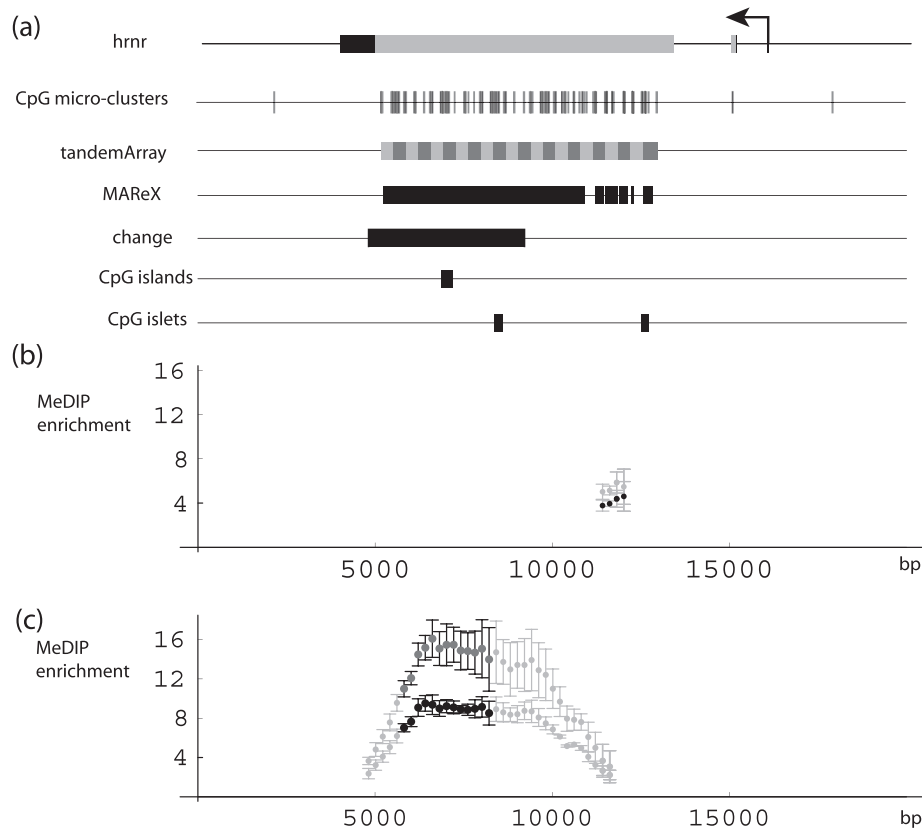


Figure 5. DNA-methylation levels in hornerin (*HRNR*) gene region. (a) *HRNR* gene representation, with coding and non-coding exon regions shown in gray and black, respectively, aligned with CpG micro-clusters, tandem array (repeat elements in alternating gray and black), MAREX domains, region with age-related change in DNA-methylation levels, CpG islands and CpG islets. (b) MeDIP enrichment in *HRNR* region, with values derived (as conventionally done) from uniquely alignable reads. Gray: adult; black: newborn. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched. Error bars in this and subsequent figures represent three MeDIP data sets. (c) MeDIP enrichment in *HRNR* region, with values calculated exclusively from reads that partially or completely overlap MAREX-annotated genome domains. Domains with consistent ≥ 2 -fold enrichment and ≥ 1.5 -fold age-related difference in each of three consecutive MeDIP data sets: black, newborn; dark gray, adult. Light gray, domains where one or both values ≥ 2 -fold enriched.

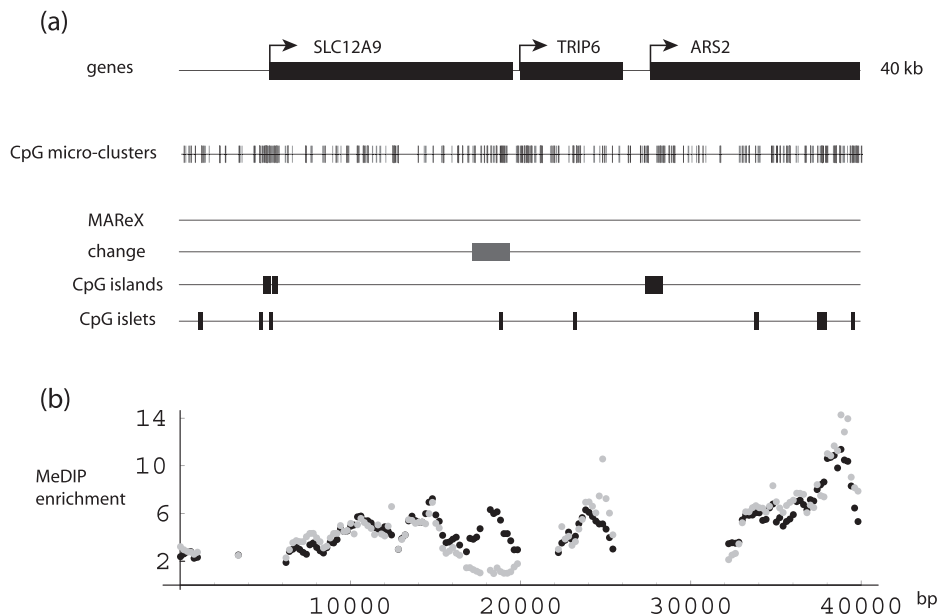


Figure 6. DNA-methylation levels in the *SLC12A9-TRIP6-ARS2* gene region. (a) *SLC12A9*, *TRIP6* and *ARS2* gene structures aligned with CpG micro-clusters, MAREX domains, DNA-methylation age-related change region, CpG islands and CpG islets. (b) *SLC12A9-TRIP6-ARS2* region MeDIP enrichment. Black: newborn; gray: adult. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched (note: Figures 7A and 8 through 10 show both MeDIP and H3K27me3 ChIP data; the latter are omitted in this figure because all values were below the minimal enrichment threshold).

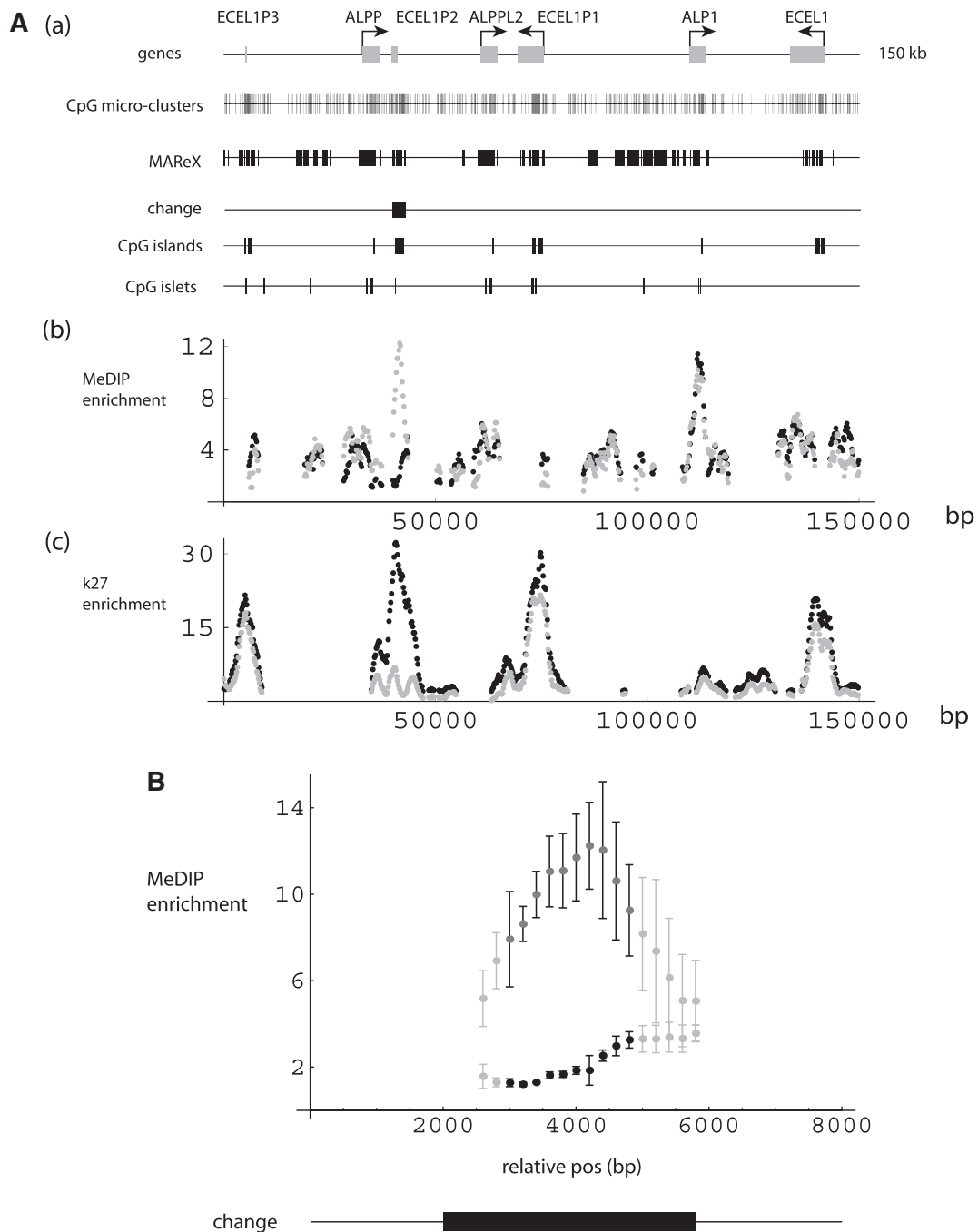


Figure 7. DNA-methylation levels of the *ECELI-ALP* reduplicated region. **(A)** (a) Schematic diagrams. *ECELI* and *ALP* gene family structures, with corresponding CpG micro-clusters, MAREX-annotated domains, age-related change region, CpG islands and CpG islets. **(b)** and **(c)** MeDIP and H3K27 trimethylation levels across the *ECELI-ALP* region. Black: newborn; gray: adult. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched. **(B)** *ECELI-ALP* subregion centered on area of change (black rectangle). Points plotted in black (newborn) or dark gray (adult) represent regions with ≥ 6 -fold enrichment relative to mean, and ≥ 2 -fold age-related change in the three data sets examined. Remaining data pairs plotted in light gray (newborn and adult) represent regions where one or both values enriched by ≥ 2 -fold.

varied widely, so products were normalized before being combined and subjected to next-generation sequencing. As shown in Figure 11, bisulfite analysis confirmed both the *ECELI1P2* and *FZD1* region MeDIP-Seq results. In the case of *ECELI1P2*, the percentage of methylated CpGs increased from 19% to 92% ($P < 0.0001$, *t*-test for

mean difference). Strikingly, the percentage of CpG positions measured to have $>90\%$ methylation rose from 5% to 77% ($P < 10^{-40}$, Fisher's exact test). For the *FZD1* gene, bisulfite analysis revealed a transition at about position 1000, so calculations were based on the downstream subregion, which presumably was the primary

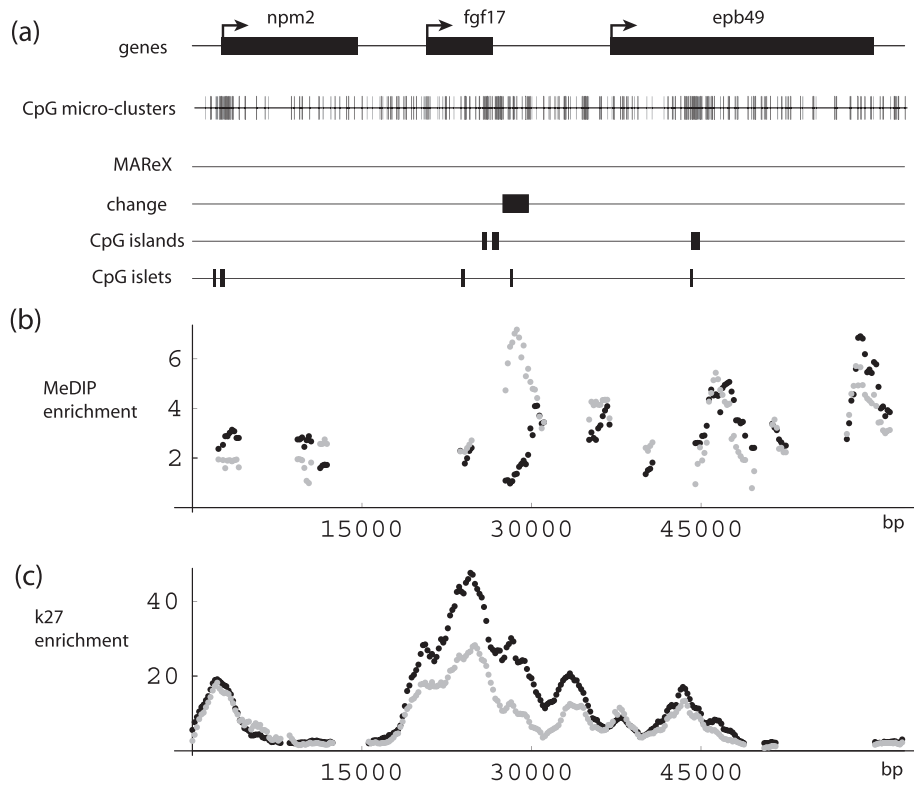


Figure 8. DNA-methylation levels in the *FGF17* gene region. (a) Schematic diagrams. *NPM2*, *FGF17* and *EPB49* gene structures, together with corresponding CpG micro-clusters, MAREX domains, age-related change region, CpG islands and CpG islets. (b and c) MeDIP and H3K27 trimethylation levels across the *FGF17* region. Black: newborn; gray: adult. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched.

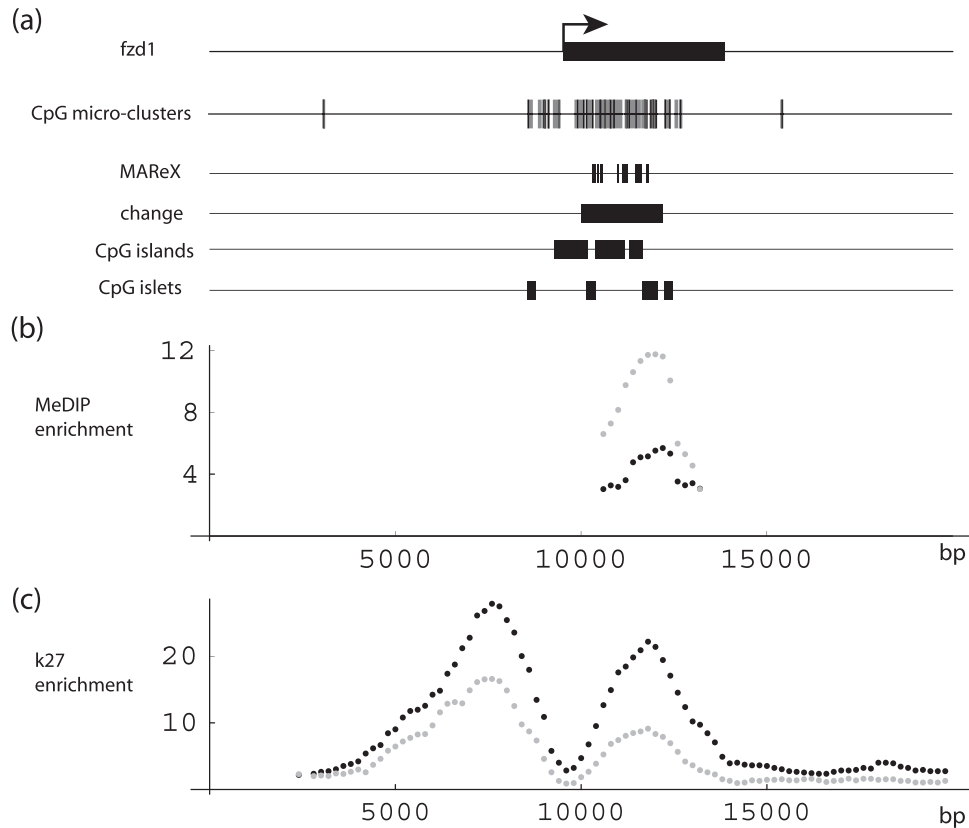


Figure 9. DNA-methylation levels in the *FZD1* gene region. (a) Schematic diagrams. *FZD1* gene structure and corresponding CpG micro-clusters, MAREX domains, age-related MeDIP change region, CpG islands and CpG islets. (b and c) MeDIP and H3K27 trimethylation levels across the *FZD1* region. Black: newborn; gray: adult. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched.

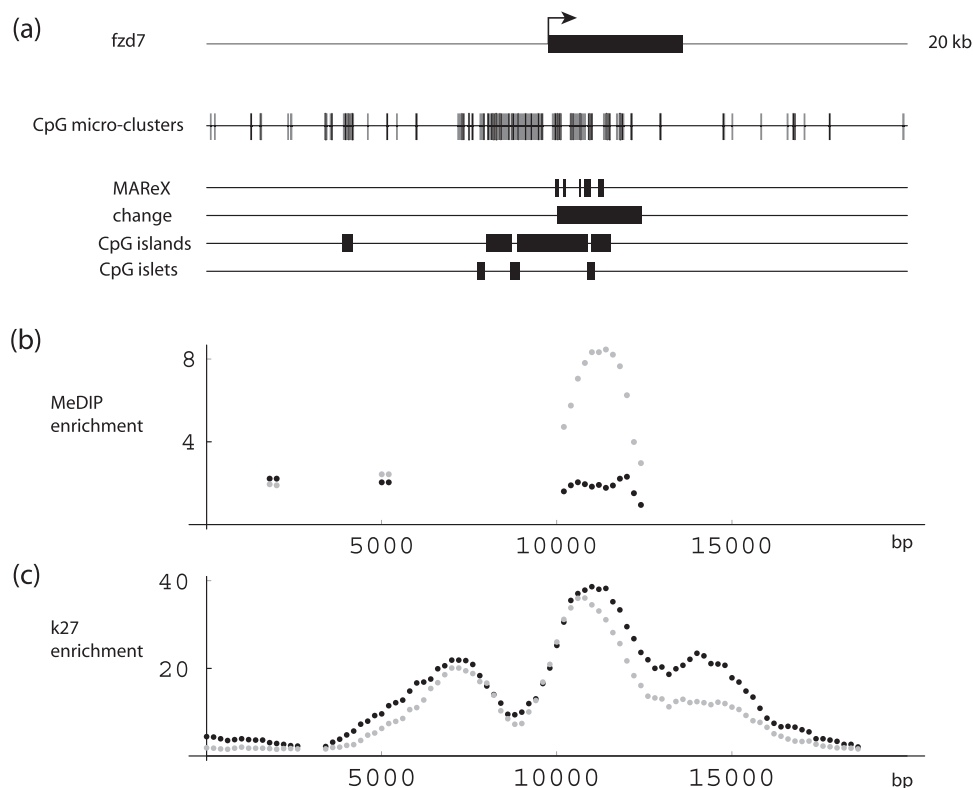


Figure 10. DNA-methylation levels in the *FZD7* gene region. (a) Schematic diagrams. *FZD7* gene structure and corresponding CpG micro-clusters, MAREX domains, age-related MeDIP change region, CpG islands and CpG islets. (b and c) MeDIP and H3K27 trimethylation levels across the *FZD7* region. Black: newborn; gray: adult. Data pairs shown only for regions where one or both values ≥ 2 -fold enriched.

determinant for MeDIP enrichment. Here, the percentage of methylated CpGs increased from 69% to 93% ($P < 0.0001$, t -test for mean difference), while the percentage of CpG positions with $>90\%$ methylation rose from 30% to 77% ($P < 2 \times 10^{-7}$, Fisher's exact test). These results compare to calculated MeDIP change values from the first experiment of 7-fold and 2.6-fold for the *ECEL1P2* and *FZD1* genes, respectively.

DISCUSSION

In the present study we examined, for the first time on a whole genome scale, age-related differences in human DNA-methylation patterns in cells from newborns and adults. MeDIP followed by next generation sequencing permits the study of DNA methylation without bias towards gene promoters, CpG islands, or other specific regions. This approach was coupled with an experimental design in which pairs of pooled chromatin samples, each derived from 9 or 10 individuals, were derived from cord blood or adult peripheral blood cells. Monocytes were the primary focus of the study, since these cells retain the capacity to differentiate into several different lineages, including DC. In analyzing our data, the primary goal was to identify DNA-methylation domains that undergo age-related epigenome remodeling at sufficient frequencies to emerge from a background of experimental noise and stochastic inter-individual fluctuation (67–72).

Despite the number of samples collected (the data were derived in total from 38 individuals), and that two independent experiments were done, we cannot exclude that factors other than age might have contributed to differences observed. Nevertheless, the findings provide a basis for proposing that a specific class of genome domains is predisposed to progressive post-natal DNA methylation. Once identified, such domains should enable more detailed studies to define how metastable epigenome states may contribute to developmental- and age-related disease.

Several analyses were done to assess MeDIP in this study. Heatmap views at 2-Mb resolution, aligned transcription start sites, and enrichment of interspersed repetitive sequences, all confirmed reproducibility for global and low resolution patterns. At smaller scales, our data support emergence of reproducible signals for developmental- or age-related methylome change primarily for regions that reach high local levels of CpG methylation. Using Bisulfite-Seq, we validated such levels directly for two representative 2-kb regions. These results revealed that transitional, heterogeneous methylation patterns can be present already in newborns. Further experiments will be required to evaluate detailed CpG methylation patterns in individuals, as opposed to pooled samples; however, prior to individual studies, a strategy should be determined, i.e. whether it is adequate to assay single (or small numbers of arbitrarily selected) CpG positions, or instead larger regions should be evaluated, as has been done here.

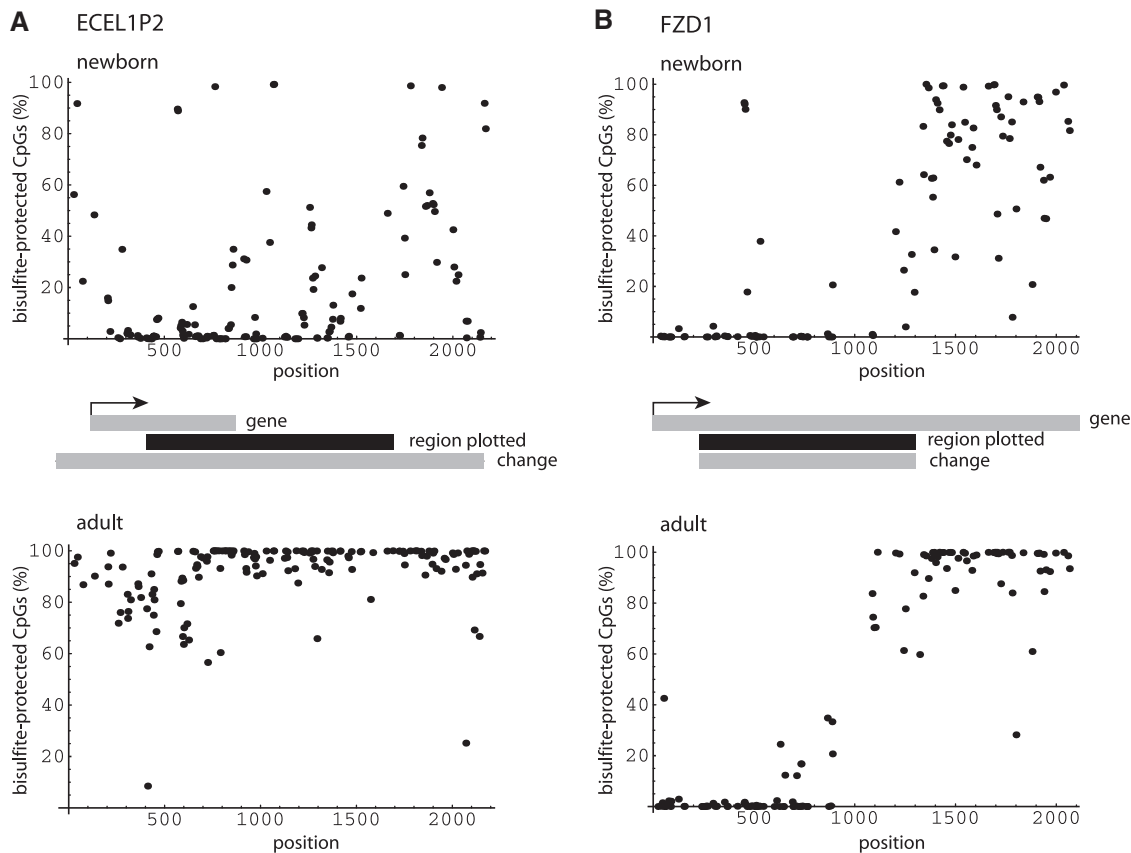


Figure 11. Methylation in ECEL1P2 and FZD1 by targeted Bisulfite Sequencing (A) Schematic diagram of the percentage of bisulfite protected CpGs (methyl-CpGs) relative to their positions in the region of interest at the ECEL1P2 pseudogene. Coverage of 182 CpGs within the region mapped was 79%. Change (gray rectangle) refers to corresponding region in Figure 7. Upper panel: Newborn. Lower panel: Adult. (B) Percentage of bisulfite protected CpGs relative to the FZD1 gene. Coverage of 79 CpGs within the region spanning positions 1000–2000 was 77%. Change (gray rectangle) refers to corresponding region in Figure 9. Upper panel: Newborn. Lower panel: Adult.

A striking finding to emerge from scanning the human methylome at a resolution of 100 kb is that the greater part of the protocadherin G (*PCDHG*) gene cluster is subject to age-related change. *PCDHG* genes are abundantly expressed in the central nervous system during embryonic development and adulthood (55,73,74), and play a pivotal role in the establishment of specific neuronal connectivity and synapse formation (54–56,74). The unusual genomic organization of this gene cluster is thought to give rise to significant molecular diversity in expression patterns (55,56). Interestingly, the *PCDHG* cluster is silenced by long-range epigenetic modifications in Wilm's tumor (75). This provides indirect support to our results, since there is evidence that genes silenced in cancer cells in some instances exhibit a tendency to increased methylation during aging (17,29,31,32,76–78). DNA methylation of another member of *PCDH* gene family, *PCDH10*, has been demonstrated to change with post-natal development in mice (29).

In further analysis of the *PCDHG* gene cluster, we used blast to quantify cross-homologies between variant exons, and to relate these in turn to DNA-methylation patterns. Here it was observed that lower homology regions interspersed between the CpG island-rich exons contribute preferentially to the extended region of age-related

change. Nevertheless, considerable evidence exists in the literature that tandemly repeated sequence arrays are prone to progressive DNA methylation and silencing (79–83). An obvious question was whether numerous genome regions might exist where interspersed-tandem arrays of complex 'single copy' sequences are predisposed to age-related methylation. This led us to implement a bioinformatic strategy for annotating 'non-repetitive' short sequence regions that have one or more near perfect copies elsewhere in the genome, i.e. are multiply-alignable and RepeatMasker excluded (MAReX). Restricting searches to such annotated domains yielded gene regions characterized by multiple duplication and rearrangement events (e.g. *FAM90A*), as well as genes with internal tandem repeating domains (e.g. *HRNR*) that would normally have eluded detection.

Strongly supportive of with the idea that selected tandemly arrayed gene clusters can be subject to developmental or age-related DNA methylation are results relating to the *ECEL1P2* pseudogene, where we observed a remarkable 7-fold age-related increase in MeDIP enrichment, validated by Bisulfite-Seq analysis. The 150 Kb region encompassing this pseudogene, as well as *ECEL1* and alkaline phosphatase gene family copies, has undergone duplication and partial triplication

events within the human genome (63). As expected, numerous MAREX annotated domains fall within and between these genes. We note further the strong peaks of H3–K27 enrichment corresponding to the set of *ECEL1*-related genes. Apparent DNA-methylation levels do not change in *ECEL1* and the other *ECEL1*-related genes, so it seems that a particular combination of CpG micro-cluster density, H3–K27 modification, and other variables not yet identified may be necessary for such age-related epigenome remodeling to occur.

Of particular interest for future study are several genes identified using optimized search parameters, namely *FGF17*, *FZD1* and *FZD7*, for which there are well established links to differentiation and developmental programs (65,66,84–89). The *FZD1* and *FZD7* transcripts are unspliced, so these might be considered processed genes that remain functional. In common with *ECEL1P2*, both *FZD* genes contain multiply-alignable (MAREX) domains, a consequence of highly homologous regions shared by *FZD1*, *FZD2* and *FZD7*. *FGF17* lacks MAREX domains by the criteria set here, but shares regions with *FGF8* and *FGF18*, which are nearly as homologous as those shared between the *FZD1*/*FZD2*/*FZD7* set. Change regions for *FZD1* and *FZD7* fall within transcription units, whereas the corresponding region for *FGF17* is located <1kb from the 3'-end of the gene. Finally, like *ECEL1P2*, and the *PCDHG* variable exon region (unpublished data), *FGF17*, *FZD1* and *FZD7* are all associated with strong peaks of H3–K27 enrichment.

QPCR analysis revealed an increase in *FZD1* expression in adult individuals relative to newborns (2.7-fold; $P < 10^{-5}$), whereas no significant differences in *FZD7* RNA levels were observed (Supplementary Figure S5A and S5B). For *FZD1*, the direction of expression change is inversely proportional to the H3K27me3 marks, consistent with the generally repressive function of such marks. Conversely, *FZD1* expression parallels the direction of DNA-methylation change. The latter result is unexplained, but we note that the CpG island located at the 5'-end of the gene remains unmethylated. Elevated DNA methylation in the gene bodies of expressed genes has been reported, and may play a role in transcription or alternative splicing linked to early stages of cell lineage formation (21–23,26). Bisulfite-Seq analysis confirmed a roughly 1-kb region with high methyl-CpG density within the *FZD1* gene body in adults, and it is conceivable that such densities can displace proteins associated with repressive H3K27me3 marks.

Finally, we would like to address intriguing commonalities between the regions reported here to exhibit increasing post-natal DNA methylation. In addition to the expected combinations of CpG islands, islets and micro-clusters, all encompass, or closely flank, domains for which highly homologous copies exist elsewhere in the human genome, and all are associated with H3–K27 enrichment. As already noted, tandemly arrayed transgenes are subject to H3–K27 and DNA methylation (79–83). Yet most tandemly arrayed genes may be silenced too rapidly to exhibit post-natal developmental change (90), whereas small, interspersed or more distant homology regions may interact with appropriate kinetics. Given this

background, we postulate that the latter feature will be observed frequently as more details emerge concerning age-related epigenome dynamics. Other investigators have postulated that gene function diversification can explain the high observed content of multi-copy gene families (including conserved pseudogenes), as well as complex reduplicated regions in mammalian genomes (91–95). Our results suggest that such highly homologous multi-copy regions may provide a mechanistic basis for developmental and age-related epigenome remodeling. If so, this fundamental feature of genome architecture in higher eukaryotic organisms merits much more extensive study.

ACCESSION NUMBERS

Public Database Accession Number [Sequence Read Archive (SRA) division of GenBank]: SRP007666.1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

The authors acknowledge the NIH Blood Bank and all of their staff, especially Cindy Matthews and Amy Melpolder (Department of Transfusion Medicine, National Institutes of Health) for their contribution in the collection of all the adult blood samples used in the present study. The authors are very grateful to Sandy Field (Perinatology Research lab) for her valuable help in the collection of cord blood samples, to Kirin Prasad for contributions to the QPCR analysis, and to the reviewers of the manuscript for helpful suggestions.

FUNDING

Funding for open access charge: Intramural Research Program of the National Institutes of Health (National Institute of Child Health and Human Development).

Conflict of interest statement. None declared.

REFERENCES

- de Boer,J., Andressoo,J.O., de Wit,J., Huijmans,J., Beems,R.B., van Steeg,H., Weeda,G., van der Horst,G.T., van Leeuwen,W., Themmen,A.P. *et al.* (2002) Premature aging in mice deficient in DNA repair and transcription. *Science*, **296**, 1276–1279.
- Maslov,A.Y. and Vijg,J. (2009) Genome instability, cancer and aging. *Biochim. Biophys. Acta*, **1790**, 963–969.
- Finkel,T., Serrano,M. and Blasco,M.A. (2007) The common biology of cancer and ageing. *Nature*, **448**, 767–774.
- Kirkwood,T.B. (2005) Understanding the odd science of aging. *Cell*, **120**, 437–447.
- Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**(Suppl), 245–254.
- Whitelaw,N.C. and Whitelaw,E. (2006) How lifetimes shape epigenotype within and across generations. *Hum. Mol. Genet.*, **15**(Spec No 2), R131–R137.

7. Howard, B.H. (1996) Replicative senescence: considerations relating to the stability of heterochromatin domains. *Exp. Gerontol.*, **31**, 281–293.
8. de Magalhaes, J.P. and Church, G.M. (2005) Genomes optimize reproduction: aging as a consequence of the developmental program. *Physiology*, **20**, 252–259.
9. Finch, C. (1990) *Longevity, Senescence, and the Genome*. University of Chicago Press, Chicago.
10. Martino, D.J., Tulic, M.K., Gordon, L., Hodder, M., Richman, T., Metcalfe, J., Prescott, S.L. and Saffery, R. (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics*, **6**, 1085–1094.
11. Takasugi, M. (2011) Progressive age-dependent DNA methylation changes start before adulthood in mouse tissues. *Mech. Ageing Dev.*, **132**, 65–71.
12. Godfrey, K.M., Lillycrop, K.A., Burdge, G.C., Gluckman, P.D. and Hanson, M.A. (2007) Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr. Res.*, **61**, 5R–10R.
13. Ozanne, S.E. and Constancia, M. (2007) Mechanisms of disease: the developmental origins of disease and the role of the epigenotype. *Nat. Clin. Pract. Endocrinol. Metab.*, **3**, 539–546.
14. Bocker, M.T., Hellwig, I., Breiling, A., Eckstein, V., Ho, A.D. and Lyko, F. (2011) Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging. *Blood*, **117**, e182–e189.
15. De Haan, G. and Gerrits, A. (2007) Epigenetic control of hematopoietic stem cell aging the case of *Ezh2*. *Ann. N Y Acad. Sci.*, **1106**, 233–239.
16. Pollina, E.A. and Brunet, A. (2011) Epigenetic regulation of aging stem cells. *Oncogene*, **30**, 3105–3126.
17. Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P. *et al.* (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.*, **20**, 440–446.
18. Jaenisch, R. (1997) DNA methylation and imprinting: why bother? *Trends Genet.*, **13**, 323–329.
19. Bird, A., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, **40**, 91–99.
20. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
21. Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q. and Church, G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.*, **27**, 361–368.
22. Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A.J., Socci, N.D. and Scandura, J.M. (2011) DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS One*, **6**, e14524.
23. Laurent, L., Wong, E., Li, G., Huynh, T., Tsigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J. *et al.* (2010) Dynamic changes in the human methylome during differentiation. *Genome Res.*, **20**, 320–331.
24. Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
25. Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S. *et al.* (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.*, **41**, 1350–1353.
26. Deaton, A.M., Webb, S., Kerr, A.R., Illingworth, R.S., Guy, J., Andrews, R. and Bird, A. (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.*, **21**, 1074–1086.
27. Oakes, C.C., Smiraglia, D.J., Plass, C., Trasler, J.M. and Robaire, B. (2003) Aging results in hypermethylation of ribosomal DNA in sperm and liver of male rats. *Proc. Natl Acad. Sci. USA*, **100**, 1775–1780.
28. Swisshelm, K., Disteche, C.M., Thorvaldsen, J., Nelson, A. and Salk, D. (1990) Age-related increase in methylation of ribosomal genes and inactivation of chromosome-specific rRNA gene clusters in mouse. *Mutat. Res.*, **237**, 131–146.
29. Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A. and Issa, J.P. (2010) Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res.*, **20**, 332–340.
30. Thompson, R.F., Atzmon, G., Gheorghe, C., Liang, H.Q., Lowes, C., Grealia, J.M. and Barzilay, N. (2010) Tissue-specific dysregulation of DNA methylation in aging. *Ageing Cell*, **9**, 506–518.
31. Issa, J.P., Ottaviano, Y.L., Celano, P., Hamilton, S.R., Davidson, N.E. and Baylin, S.B. (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat. Genet.*, **7**, 536–540.
32. Kwabi-Addo, B., Chung, W., Shen, L., Ittmann, M., Wheeler, T., Jelinek, J. and Issa, J.P. (2007) Age-related DNA methylation changes in normal human prostate tissues. *Clin. Cancer Res.*, **13**, 3796–3802.
33. Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA*, **102**, 10604–10609.
34. Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Asplund, T., Cui, H., Yu, W., Rongione, M.A., Ekstrom, T.J., Harris, T.B. *et al.* (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA*, **299**, 2877–2883.
35. Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S. and Ophoff, R.A. (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*, **4**, e6767.
36. Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R. *et al.* (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.*, **5**, e1000602.
37. Langevin, S.M., Houseman, E.A., Christensen, B.C., Wiencke, J.K., Nelson, H.H., Karagas, M.R., Marsit, C.J. and Kelsey, K.T. (2011) The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. *Epigenetics*, **6**, 908–919.
38. Kuwana, M., Okazaki, Y., Kodama, H., Izumi, K., Yasuoka, H., Ogawa, Y., Kawakami, Y. and Ikeda, Y. (2003) Human circulating CD14⁺ monocytes as a source of progenitors that exhibit mesenchymal cell differentiation. *J. Leukoc. Biol.*, **74**, 833–845.
39. Randolph, G.J., Inaba, K., Robbiani, D.F., Steinman, R.M. and Muller, W.A. (1999) Differentiation of phagocytic monocytes into lymph node dendritic cells in vivo. *Immunity*, **11**, 753–761.
40. Ruhnke, M., Ungefroren, H., Nussler, A., Martin, F., Brulport, M., Schormann, W., Hengstler, J.G., Klapper, W., Ulrichs, K., Hutchinson, J.A. *et al.* (2005) Differentiation of in vitro-modified human peripheral blood monocytes into hepatocyte-like and pancreatic islet-like cells. *Gastroenterology*, **128**, 1774–1786.
41. Zhou, L.J. and Tedder, T.F. (1996) CD14⁺ blood monocytes can differentiate into functionally mature CD83⁺ dendritic cells. *Proc. Natl Acad. Sci. USA*, **93**, 2588–2592.
42. Lehner, M. and Holter, W. (2002) Endotoxin-free purification of monocytes for dendritic cell generation via discontinuous density gradient centrifugation based on diluted Ficoll-Paque Plus. *Int. Arch. Allergy Immunol.*, **128**, 73–76.
43. Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. and Schubeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*, **37**, 853–862.
44. Russanova, V.R., Hirai, T.H., Tchernov, A.V. and Howard, B.H. (2004) Mapping development-related and age-related chromatin remodeling by a high throughput ChIP-HPLC approach. *J. Gerontol. A Biol. Sci. Med. Sci.*, **59**, 1234–1243.
45. Porras, A., Kozar, S., Russanova, V., Salpea, P., Hirai, T., Sammons, N., Mittal, P., Kim, J.Y., Ozato, K., Romero, R. *et al.* (2008) Developmental and epigenetic regulation of the human TLR3 gene. *Mol. Immunol.*, **46**, 27–36.

46. Down, T.A., Rakyán, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M. *et al.* (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.*, **26**, 779–785.
47. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
48. Docherty, S.J., Davis, O.S., Haworth, C.M., Plomin, R. and Mill, J. (2009) Bisulfite-based epotyping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenet. Chromatin*, **2**, 3.
49. Brock, G.J., Charlton, J. and Bird, A. (1999) Densely methylated sequences that are preferentially localized at telomere-proximal regions of human chromosomes. *Gene*, **240**, 269–277.
50. Issa, J.P. (2004) CpG island methylator phenotype in cancer. *Nat. Rev. Cancer*, **4**, 988–993.
51. Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
52. van Furth, R., Raeburn, J.A. and van Zwet, T.L. (1979) Characteristics of human mononuclear phagocytes. *Blood*, **54**, 485–500.
53. Pickl, W.F., Majdic, O., Kohl, P., Stockl, J., Riedl, E., Scheinecker, C., Bello-Fernandez, C. and Knapp, W. (1996) Molecular and functional characteristics of dendritic cells generated from highly purified CD14+ peripheral blood monocytes. *J. Immunol.*, **157**, 3850–3859.
54. Morishita, H. and Yagi, T. (2007) Protocadherin family: diversity, structure, and function. *Curr. Opin. Cell Biol.*, **19**, 584–592.
55. Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, **97**, 779–790.
56. Wu, Q., Zhang, T., Cheng, J.F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J.P., Zhang, M.Q., Myers, R.M. *et al.* (2001) Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.*, **11**, 389–404.
57. Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.
58. Bosch, N., Caceres, M., Cardone, M.F., Carreras, A., Ballana, E., Rocchi, M., Armengol, L. and Estivill, X. (2007) Characterization and evolution of the novel gene family FAM90A in primates originated by multiple duplication and rearrangement events. *Hum. Mol. Genet.*, **16**, 2572–2582.
59. Makino, T., Takaishi, M., Morohashi, M. and Huh, N.H. (2001) Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin. *J. Biol. Chem.*, **276**, 47445–47452.
60. Takaishi, M., Makino, T., Morohashi, M. and Huh, N.H. (2005) Identification of human hornerin and its expression in regenerating and psoriatic skin. *J. Biol. Chem.*, **280**, 4696–4703.
61. Henry, J., Hsu, C.Y., Haftek, M., Nachat, R., de Koning, H.D., Gardinal-Galera, I., Hitomi, K., Balica, S., Jean-Decoster, C., Schmitt, A.M. *et al.* (2011) Hornerin is a component of the epidermal cornified cell envelopes. *FASEB J.*, **25**, 1567–1576.
62. Knoll, B.J., Rothblum, K.N. and Longley, M. (1988) Nucleotide sequence of the human placental alkaline phosphatase gene. Evolution of the 5' flanking region by deletion/substitution. *J. Biol. Chem.*, **263**, 12020–12027.
63. Rump, A., Kasper, G., Hayes, C., Wen, G., Starke, H., Liehr, T., Lehmann, R., Lagemann, D. and Rosenthal, A. (2001) Complex arrangement of genes within a 220-kb region of double-duplicated DNA on human 2q37.1. *Genomics*, **73**, 50–55.
64. Beckman, G., Beckman, L., Sikstrom, C. and Millan, J.L. (1992) DNA polymorphism of alkaline phosphatase isozyme genes: linkage disequilibria between placental and germ-cell alkaline phosphatase alleles. *Am. J. Hum. Genet.*, **51**, 1066–1070.
65. Katoh, M. (2008) WNT signaling in stem cell biology and regenerative medicine. *Curr. Drug Targets*, **9**, 565–570.
66. Yu, H., Smallwood, P.M., Wang, Y., Vidaltamayo, R., Reed, R. and Nathans, J. (2010) Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. *Development*, **137**, 3707–3717.
67. Flanagan, J.M., Popendikyte, V., Pozdniakovaite, N., Sobolev, M., Assadzadeh, A., Schumacher, A., Zangeneh, M., Lau, L., Virtanen, C., Wang, S.C. *et al.* (2006) Intra- and interindividual epigenetic variation in human germ cells. *Am. J. Hum. Genet.*, **79**, 67–84.
68. Heijmans, B.T., Kremer, D., Tobi, E.W., Boomsma, D.I. and Slagboom, P.E. (2007) Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum. Mol. Genet.*, **16**, 547–554.
69. Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J. *et al.* (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.*, **6**, e22.
70. Sandovici, I., Kassovska-Bratinova, S., Loredó-Ostí, J.C., Leppert, M., Suarez, A., Stewart, R., Bautista, F.D., Schiraldi, M. and Sapienza, C. (2005) Interindividual variability and parent of origin DNA methylation differences at specific human Alu elements. *Hum. Mol. Genet.*, **14**, 2135–2143.
71. Schneider, E., Plushch, G., El Hajj, N., Galetzka, D., Puhl, A., Schorsch, M., Frauenknecht, K., Riepert, T., Tresch, A., Müller, A.M. *et al.* (2010) Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns. *Nucleic Acids Res.*, **38**, 3880–3890.
72. Turan, N., Katari, S., Coutifaris, C. and Sapienza, C. (2010) Explaining inter-individual variability in phenotype: is epigenetics up to the challenge? *Epigenetics*, **5**, 16–19.
73. Hirano, S., Ono, T., Yan, Q., Wang, X., Sonta, S. and Suzuki, S.T. (1999) Protocadherin 2C: a new member of the protocadherin 2 subfamily expressed in a redundant manner with OL-protocadherin in the developing brain. *Biochem. Biophys. Res. Commun.*, **260**, 641–645.
74. Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., Yasuda, M., Mishina, M. and Yagi, T. (1998) Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron*, **20**, 1137–1151.
75. Dallosso, A.R., Hancock, A.L., Szemes, M., Moorwood, K., Chilukamari, L., Tsai, H.H., Sarkar, A., Barasch, J., Vuononvirta, R., Jones, C. *et al.* (2009) Frequent long-range epigenetic silencing of protocadherin gene clusters on chromosome 5q31 in Wilms' tumor. *PLoS Genet.*, **5**, e1000745.
76. Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B. and Issa, J.P. (1998) Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res.*, **58**, 5489–5494.
77. Issa, J.P. (2000) CpG-island methylation in aging and cancer. *Curr. Top Microbiol. Immunol.*, **249**, 101–118.
78. Rakyán, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M. *et al.* (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res.*, **20**, 434–439.
79. Bender, J. (1998) Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. *Trends Biochem. Sci.*, **23**, 252–256.
80. Bestor, T. and Tycko, B. (1996) Creation of genomic methylation patterns. *Nat. Genet.*, **12**, 363–367.
81. Dorer, D.R. and Henikoff, S. (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics*, **147**, 1181–1190.
82. Garrick, D., Fiering, S., Martin, D.I. and Whitelaw, E. (1998) Repeat-induced gene silencing in mammals. *Nat. Genet.*, **18**, 56–59.
83. Mutskov, V. and Felsenfeld, G. (2004) Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.*, **23**, 138–149.
84. Hoshikawa, M., Ohbayashi, N., Yonamine, A., Konishi, M., Ozaki, K., Fukui, S. and Itoh, N. (1998) Structure and expression of a novel fibroblast growth factor, FGF-17, preferentially expressed in the embryonic brain. *Biochem. Biophys. Res. Commun.*, **244**, 187–191.

85. Huang,H.C. and Klein,P.S. (2004) The Frizzled family: receptors for multiple signal transduction pathways. *Genome Biol.*, **5**, 234.
86. Mariani,F.V., Ahn,C.P. and Martin,G.R. (2008) Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature*, **453**, 401–405.
87. Reya,T., Duncan,A.W., Ailles,L., Domen,J., Scherer,D.C., Willert,K., Hintz,L., Nusse,R. and Weissman,I.L. (2003) A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature*, **423**, 409–414.
88. Sun,X., Lewandoski,M., Meyers,E.N., Liu,Y.H., Maxson,R.E. Jr and Martin,G.R. (2000) Conditional inactivation of Fgf4 reveals complexity of signalling during limb bud development. *Nat. Genet.*, **25**, 83–86.
89. Winklbauer,R., Medina,A., Swain,R.K. and Steinbeisser,H. (2001) Frizzled-7 signalling controls tissue separation during Xenopus gastrulation. *Nature*, **413**, 856–860.
90. Robertson,G., Garrick,D., Wilson,M., Martin,D.I. and Whitelaw,E. (1996) Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res.*, **24**, 1465–1471.
91. Henikoff,S., Greene,E.A., Pietrokovski,S., Bork,P., Attwood,T.K. and Hood,L. (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science*, **278**, 609–614.
92. Itoh,N. and Ornitz,D.M. (2004) Evolution of the Fgf and Fgfr gene families. *Trends Genet.*, **20**, 563–569.
93. Marques,A.C., Vinckenbosch,N., Brawand,D. and Kaessmann,H. (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol.*, **9**, R54.
94. Ohno,S. (1970) *Evolution by Gene Duplication*. Springer Verlag, Berlin.
95. Young,J.M., Friedman,C., Williams,E.M., Ross,J.A., Tonnes-Priddy,L. and Trask,B.J. (2002) Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum. Mol. Genet.*, **11**, 535–546.