

Intra- and Interindividual Epigenetic Variation in Human Germ Cells

James M. Flanagan, Violeta Pependikyte, Natalija Pozdniakovaite, Martha Sobolev, Abbas Assadzadeh, Axel Schumacher, Masood Zangeneh, Lynette Lau, Carl Virtanen, Sun-Chong Wang, and Arturas Petronis

Epigenetics represents a secondary inheritance system that has been poorly investigated in human biology. The objective of this study was to perform a comprehensive analysis of DNA methylation variation between and within the germlines of normal males. First, methylated cytosines were mapped using bisulphite modification–based sequencing in the promoter regions of the following disease genes: presenilins (*PSEN1* and *PSEN2*), breast cancer (*BRCA1* and *BRCA2*), myotonic dystrophy (*DM1*), and Huntington disease (*HD*). Major epigenetic variation was detected within samples, since the majority of sperm cells of the same individual exhibited unique DNA methylation profiles. In the interindividual analysis, 41 of 61 pairwise comparisons revealed distinct DNA methylation profiles ($P = .036$ to 6.8×10^{-14}). Second, a microarray-based epigenetic profiling of the same sperm samples was performed using a 12,198-feature CpG island microarray. The microarray analysis has identified numerous DNA methylation–variable positions in the germ cell genome. The largest degree of variation was detected within the promoter CpG islands and pericentromeric satellites among the single-copy DNA fragments and repetitive elements, respectively. A number of genes, such as *EED*, *CTNNA2*, *CALM1*, *CDH13*, and *STMN2*, exhibited age-related DNA methylation changes. Finally, allele-specific methylation patterns in *CDH13* were detected. This study provides evidence for significant epigenetic variability in human germ cells, which warrants further research to determine whether such epigenetic patterns can be efficiently transmitted across generations and what impact inherited epigenetic individuality may have on phenotypic outcomes in health and disease.

Phenotypic differences among individuals have traditionally been attributed to genetic (DNA sequence) variation and environmental differences. Over the past several decades, documentation of DNA sequence variants has been one of the top priorities in biomedical research. Numerous major international projects—from the sequencing of the Human Genome^{1,2} to the creation of SNP databases (dbSNP, now called “Entrez SNP”) and the Haplotype Map³ (HapMap)—have contributed significantly to the understanding of the position, degree, and structure of DNA polymorphisms. However, SNPs and other DNA sequence differences are relatively rare, and DNA sequences of two unrelated individuals are 99.5% identical. Furthermore, only a small fraction of these polymorphisms are functional—that is, polymorphisms that change amino acid sequence in the protein or have an impact on gene expression. Sequencing of the chimpanzee (*Pan troglodytes*) genome revealed 98.67% DNA sequence identity to the human genome, and, again, only a fraction of polymorphisms appear to result in structural or functional gene differences.⁴ Such findings raise the question: is this low DNA sequence variation across unrelated individuals and our closest related species sufficient to account for all major differences in physiological and psychological phenotypic outcomes?

One potential, although poorly investigated, source of

phenotypic differences is epigenetic variation. By definition, “epigenetics” refers to the regulation of various genomic functions that are controlled by partially stable modifications of DNA and chromatin proteins.⁵ Epigenetic signals are critical to the proper functioning of the genome, as seen in *Dnmt1*-knockout mice that die in early embryogenesis,⁶ in several rare pediatric syndromes, and in cancer.⁷ One important feature of epigenetic regulation is partial epigenetic stability, or metastability. Epigenetic profiles in different cells of the same organism can be quite different, and developmental programs, environmental factors, or stochastic events in the nucleus of a cell can induce this variation. The first systematic effort to document DNA methylation differences and similarities across different genome regions, the Human Epigenome Project, was recently launched. The pilot study of the *MHC* locus on chromosome 6 investigated seven cell types (adipose, brain, breast, lung, liver, prostate, and muscle) across 32 individuals.⁸ In this study, which was not controlled for sex and age, around half (118/253) of the tested loci showed some interindividual variability in at least one tissue. The next phase of the Human Epigenome Project, which will be controlled for the above parameters, will provide coverage of >5,000 loci (~3,000 genes) from chromosomes 6, 13, 20, and 22 across >20 tissue types.⁸ The Human Epigenome Project and other smaller-scale studies

From The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health (J.M.F.; M.S.; A.A.; A.S.; M.Z.; S.-C.W.; A.P.), and Microarray Centre, The Ontario Cancer Institute (L.L.; C.V.), Toronto; and Institute of Biotechnology, Vilnius, Lithuania (V.P.; N.P.)

Received December 19, 2005; accepted for publication March 29, 2006; electronically published May 25, 2006.

Address for correspondence and reprints: Dr. Arturas Petronis, The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, 250 College Street, Toronto, ON, M5T 1R8, Canada. E-mail: arturas_petronis@camh.net

Am. J. Hum. Genet. 2006;79:67–84. © 2006 by The American Society of Human Genetics. All rights reserved. 0002-9297/2006/7901-0009\$15.00

have investigated epigenetic variation primarily in somatic cells. However, there has been very little effort to document epigenetic variation in the germline, apart from imprinted genes^{9,10} and isolated cases of germ cell epimutations.^{11,12}

There are several reasons to believe that the germline may contain substantial epigenetic variation. Epigenetic reprogramming during gametogenesis, fertilization, and embryogenesis involves dramatic chromatin remodeling.¹³ Methylation reprogramming during gametogenesis involves the erasure and reestablishment of methylation of imprinted genes and other nonimprinted genes and, then, a second wave of reprogramming during fertilization (paternal) and embryogenesis (maternal).¹³ This process is thought to (1) ensure that both gametes acquire the appropriate sex-specific epigenetic states and establish the epigenetic states required for early embryonic development and toti- or pluripotency and (2) allow the erasure of epimutations that adult germ cells may have inherited or developed during their lifetime.^{14,15} In parallel with DNA methylation, chromatin changes during spermatogenesis involve the compaction of the haploid genome by the replacement of the core histones via transition proteins to the much smaller basic protamines 1 and 2.¹⁶ However, a number of testis-specific histones and histone variants—such as TSH2B, histones H2A, H3, and H4, variants of H2B, and CENP-A—are present, to some extent, in the mature spermatozoa.^{17–19} How these remaining histones are arranged and to what extent interindividual variability in histone placement and modification can affect development and phenotype are subjects yet to be investigated. Despite dramatic changes, not all epigenetic signals are erased in the germline, and recent studies in mice have suggested that this phenomenon could underlie epigenetic inheritance.^{20,21} Therefore, there is ample opportunity during these phases of reprogramming to either maintain or generate substantial epigenetic variability in the germ cells.

Although there is evidence that some individual loci exhibit partial epigenetic stability during meiosis in mice and in other organisms, to further understand the degree, mechanisms, and importance of epigenetic inheritance across generations in humans, three main questions need to be addressed in the following order. (1) Is there any evidence for epigenetic variation in the germ cells? (2) To what extent is the epigenetic variation meiotically stable? (3) What is the impact of epigenetic variation in germ cells on phenotypic differences? In this study, we attempted to answer the first question by estimating the intra- and interindividual epigenetic variation detectable in the mature sperm of healthy individuals. For this goal, we used two different laboratory strategies. The first approach focused on promoter regions of several disease-related genes—such as *PSEN1* (MIM 104311), *PSEN2* (MIM 600759), *BRCA1* (MIM 113705), *BRCA2* (MIM 600185), *DM1* (MIM 160900), and *HD* (MIM 143100)—in healthy individuals, with the use of bisulphite modification–based

mapping of methylated cytosines and measured epigenetic “distances” between individuals. The second strategy was to perform a microarray-based epigenetic profiling of sperm DNA with the use of a CpG island microarray, which provides genomewide information on methylation variability across different unique and repetitive DNA sequences. Several loci of interest identified in the microarray experiments were further investigated using methylation-sensitive single-nucleotide primer extension (MS-SNuPE) reaction.

Material and Methods

Samples

Two sperm sample sets were used in this study. The first sample set was received from the Fairfax Cryobank, Genetics & IVF Institute, in Fairfax, VA, and consisted of 25 sperm samples collected from healthy white sperm donors at an average age of 27 years (range 22–35 years). The second set of sperm samples was collected at the Centre for Addiction and Mental Health in Toronto from 21 healthy white individuals at an average age of 39 years (range 24–56 years). This study was approved by an institutional ethics board, and informed consent was obtained from all participants. Some aspects of sperm DNA data analysis required a nonsperm tissue of reference; for this purpose, postmortem brain tissues were used (donated by The Stanley Medical Research Institute’s brain collection, courtesy of Drs. M. B. Knable, E. F. Torrey, M. J. Webster, S. Weis, and R. H. Yolken). These brain samples were from 22 white males who had an average age at death of 46 years (range 31–66 years). Extraction of DNA was performed using standard salt and phenol/chloroform extraction.

Bisulphite Modification–Based Mapping of Methylated Cytosines

Bisulphite modification–based mapping of methylated cytosines was performed as described elsewhere.²² In brief, genomic DNA (700 ng) was digested with *Bgl*II (Fermentas) for 1 h at 37°C, was denatured at 100°C for 5 min, was chilled on ice, and was then incubated at 50°C for 15 min in 0.3 M NaOH. The DNA was then mixed with 2% low-melting point agarose (SeaPlaque Agarose, FMC) and was dropped into ice-cold mineral oil to form seven beads of ~10 μ l, and, finally, the beads were placed into a freshly prepared solution containing 2.5 M sodium bisulphite (pH 5.0) plus 1 mM hydroquinone (both from Sigma). The beads were then incubated on ice for 30 min, followed by incubation at 50°C for 3.5 h. The beads were washed in four changes of Tris-EDTA (TE) (pH 8.0) for 1 h and then were desulphonated in 0.2 M NaOH for 30 min. After desulphonation, the beads were washed a second time in three changes of TE for 30 min. Before amplification, the beads were washed in H₂O for 30 min. PCR amplification of the target sequences consisted of 5 μ l of agarose beads containing the bisulphite-treated DNA, 2 mM MgCl₂, 0.2 mM deoxynucleotide triphosphates (dNTPs), 0.4 μ M each of forward and reverse primer, 250 ng/ml BSA, and 2.5 U *Taq* polymerase (New England Biolabs) in 1 \times PCR buffer, to a total volume of 50 μ l. PCR was performed using either a seminested or a fully nested approach, with the first PCR consisting of one cycle at 97°C for 4 min, 53°C for 2 min, and 72°C for 2 min, followed by 24 cycles at 94°C for 45 s, 53°C for 1 min, and 72°C for 1 min. The second PCR used 5 μ l of the first PCR as a template and consisted of one cycle at

97°C for 2 min, 53°C for 2 min, and 72°C for 1 min, followed by 24 cycles at 94°C for 45 s, 55°C for 45 s, and 72°C for 1 min. CpG islands in the 5' promoter sequences were analyzed in six genes: *PSEN1* (Entrez GeneID 5663; chromosome 14: 72672525–72673163), *PSEN2* (GeneID 5664; chromosome 1: 223365273–223365990), *BRCA1* (GeneID 672; chromosome 17: 38530561–38531181), *BRCA2* (GeneID 675; chromosome 13: 31787367–31788153), *HD* (GeneID 3064; chromosome 4: 3113281–3113816), and *DM1* (GeneID 1760; chromosome 19: 50964670–50965254). An intronic CpG island within the *CDH13* gene (GeneID 1012; chromosome 16: 81218597–81218988) was also analyzed by bisulphite genomic sequencing. Nucleotide positions given are from the May 2004 Genome (hg17) version (UCSC Genome Browser). The primers used for amplification of bisulphite-modified DNA fragments are available in table 1.

PCR products were electrophoresed on an agarose gel. DNA fragments were excised, were cleaned using Qiagen Gel Extraction Kit (Qiagen), and were cloned into the pGEM-T vector (Promega). Thirty clones from each PCR product (locus/individual) were sequenced. To evaluate the degree of intraindividual variation, we sequenced an additional 30 clones from separate bisulphite reactions in five cases: two in *BRCA1*, one in *BRCA2*, and two in *PSEN2*. A total of 1,020 clones were analyzed, which required >1,500 sequencing reactions, since some longer fragments had to be sequenced from both ends.

Analyses of DNA Methylation Variation in Bisulphite Modification-Based Experiments

The degree of epigenetic diversity within and across individuals was evaluated using the concept of epigenetic "distance."²³ Each of the 30 sequenced clones was binary coded, with "0" for an unmethylated cytosine and "1" for a methylated cytosine. Each clone was, therefore, represented by a row vector of n 0 and 1, where n is the number of cytosines in the tested region.

Estimation of intraindividual variation.—Unique methylation profiles were identified for each set of 30 clones. For example, a set of clones 0101, 0101, 0111, and 1100 exhibits three types of methylation profiles (1/2, 3, and 4), and, therefore, the proportion of unique methylation profiles is 3/4. This calculation was performed for every set of 30 clones, and then the mean and SD of the proportion of unique clones across individuals were calculated for each locus. In the second round of analysis, because of possible imperfect C→T conversion with bisulphite treatment, two clones different by a single position were treated as identical. With use of the above example, profiles 0101 and 0111 are now treated as identical, and the degree of uniqueness is 2/4. In the final analysis, the tolerance was increased to two differences—that is, the clones that exhibited two or fewer differences were treated as identical.

Comparison of DNA methylation distances across individuals.—The average methylation-intensity vector for each locus/individual was calculated by dividing the sum of the methylated cytosines by 30 for each different cytosine position. The degree of epigenetic dissimilarity was measured by Euclidean distance, by use of the following equation:

$$d_{12} = \sqrt{\sum_{i=1}^n (m_{1i} - m_{2i})^2},$$

where m_1 is the average methylation vector of individual 1, m_2

Table 1. Primer Sequences

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

is the average methylation vector of individual 2, and d_{12} is the Euclidean DNA methylation distance between individuals 1 and 2. The larger the distance, the more dissimilar the two individuals' methylation profiles are to each other. With this metric, we calculated the distances between all possible pairs of individuals for each promoter locus of *BRCA1*, *BRCA2*, *HD*, *DM1*, *PSEN1*, and *PSEN2*. To test statistical significance of methylation differences, we performed the following analysis. For each locus, all clones from all individuals were pooled together, and two sets of 30 randomly selected clones from the pool formed the methylation profiles of two pseudo-individuals. The epigenetic distance between the two pseudo-individuals was then calculated with the same procedure as above, and this procedure was repeated 100,000 times to generate 100,000 distances, the density distribution of which was plotted, and the mean (± 2 SD) was calculated. The (one-tailed) P value of a distance was then obtained by finding the area under the distribution curve, from the left up to the calculated distance. An epigenetic distance in two real individuals with $P < .05$ (i.e., >2 SD) indicates that difference in the DNA methylation of two individuals is statistically significant.

Microarray-Based DNA Methylation Analysis

Microarrays.—Genomewide epigenetic profiling was performed using the 12,192 CpG island microarrays²⁴ purchased from the University Health Network Microarray Facility in Toronto.

Enrichment of unmethylated DNA.—We used our developed technology for enrichment of the unmethylated DNA fraction and for epigenetic profiling described in detail elsewhere.²⁵ The general principle of the DNA methylation profiling consists of interrogation of the unmethylated fraction of genomic DNA on the microarray. Intensity of hybridization inversely correlates with the DNA methylation status at the genomic locus homologous to a specific DNA fragment on the array. In brief, methylation-sensitive restriction enzymes were used to digest 1 μ g of genomic DNA, and two enzyme scenarios were used in this project. First, sperm DNA samples from 25 individuals were analyzed using methylation-sensitive enzymes *HpaII*, *Hin6I*, and *AciI* (designated "sperm DNA-HHA array" set). This enzyme "cocktail" strategy, however, is not ideal for GC-rich regions, such as CpG islands, since these three enzymes would generate DNA fragments too small for efficient amplification and hybridization. Therefore, a single-digestion approach with *HpaII* alone was used on a second set of sperm DNA samples from 21 individuals (designated "sperm DNA-HpaII" array set). DNA adaptors (annealing products of two primers, U-CG1a [5'-CGTGGAGACTGACTACCAGAT-3'] and U-CG1b [5'-AGTTACATCTGGTAGTCACTCCA-3']) were ligated to the restricted DNA fragments, followed by treatment with *McrBC* (New England Biolabs), which cleaves the fragments containing two or more methylated cytosines, thereby further enriching the unmethylated fraction. Adaptor-PCR amplification of the ligated products, with the use of primers complementary to the adaptor sequence, consisted of 250 ng of ligated DNA, 2.5 mM MgCl₂, 0.2 mM aminoallyl-dNTPs (15 mM aminoallyl-2'-deoxyuridine 5'-triphosphate, 10 mM 2'-deoxythymidine 5'-triphosphate, and 25 mM each of 2'-deoxycytidine 5'-triphosphate, 2'-deoxyguanosine 5'-triphosphate, and 2'-deoxyadenosine 5'-tri-

phosphate), 200 pmol primer U-CG1b, and 5 U *Taq* polymerase (New England Biolabs) in 1 × PCR reaction buffer (Sigma), to a final volume of 100 μ l. PCR conditions are adjusted in such a way that only fragments <1.5 kb (i.e., short, digested, and, therefore, unmethylated) will amplify preferentially. Cycling consisted of an initial cycle at 72°C for 5 min and 95°C for 1 min, 25 cycles at 95°C for 40 s and 68°C for 2 min 30 s, and a final extension at 72°C for 5 min. Equal amounts of amplicons from each sample were mixed to form the pooled control, which was labeled with Cy3 and was cohybridized against each individual amplicon labeled with Cy5. Hybridization was performed at 42°C with the use of standard procedure.²⁵

For comparison with the sperm DNA methylation profiles, DNA samples from postmortem brains of 22 individuals who did not have any known brain disease were subjected to the same microarray-based DNA methylation profiling that used a single-digestion approach with *HpaII* (designated “brain DNA-*HpaII*” array set).

Microarray data processing and analysis.—Methylation differences between the individuals and the pooled control were analyzed by the ratio of hybridization intensities of Cy5 (individual samples) over Cy3 (pooled control). As we have learned from our previous analyses of arrays used for DNA methylation analysis, such ratios show normal distribution; therefore, the data can be treated similarly to those in classical microarray experiments. The array data were normalized in two steps—first, in a global intensity normalization, to adjust the Cy5: Cy3 ratio to 1:1 across the entire array, followed by a block-by-block LOWESS normalization. The data were trimmed to remove spots with ambiguous genome locations, including spots with no sequence or annotation (647 spots), spots with >30% repetitive elements (2,706 spots), and translocation hotspots (633 spots). The spots for which the microarray clones represented identical sequences were averaged, which resulted in ~4,970 unique loci. Coefficient of variation (CV) was calculated for each remaining spot by dividing the SD in Cy5/Cy3 by the mean of Cy5/Cy3 across all individuals. The sperm DNA-*HHA* experiments were performed in duplicate, and the data were averaged ratios. The sperm DNA-*HpaII* and the brain DNA-*HpaII* data sets consisted of one array per individual, because we opted for increased biological replicates rather than for increased technical replicates for the number of microarrays available.

Age-covariate analysis.—For the CpG island microarray experiment, the age-covariate analysis was performed using a correlation coefficient between two series of quantities, to measure the linear relationship between the series. Pearson correlation coefficient was calculated between the mean fold change (log Cy5/Cy3) across individuals and the ages across individuals, for each spot on the microarray. A large absolute value ($|r| > 0.5$) of the coefficient indicates that the methylation intensity at the locus covariates with age in a positive or a negative way. To test their statistical significance, the ages across individuals were permuted, and, again, the coefficients were computed using the permuted age series. For each spot, the permutation was repeated 5,000 times to get 5,000 coefficients. The one-tailed *P* value of the coefficient was then obtained by finding the fraction of times that the coefficients were larger (or smaller) than the original coefficient. Although a *P*-value cutoff at .05 may lead to many false positives, the adjustment of *P* value for multiple testing, by controlling the probability of making at least one false positive, dramatically lowers the power of the experiment and is also considered too conservative for microarray studies.²⁶ “False discovery

rate” (FDR) is defined as the expected proportion of false-positive predictions among the positive predictions. For example, in the 100 positives declared by FDR at 0.1, 90 are expected to be true positives on average. The FDR criterion has increasingly been adopted over *P* value in microarray analysis. We have, therefore, applied FDR for the findings described in this article.

The autocorrelation clustering analysis for the CpG island microarray experiment was performed using the autocorrelation function $ACF(x)$, which measures how strongly two methylation intensities x loci apart influence each other.

Measurement of Densities of Methylated Cytosines in the Selected Loci

Further analysis of a selected set of DNA fragments identified as the most variable was performed using the MS-SNuPE reaction on the ABI SNaShot platform accommodated for measuring the C/T ratios in the bisulphite-treated genomic DNA.²⁷ In brief, genomic DNA was digested with *NdeI* (Fermentas), followed by treatment with sodium bisulphite, as described above. The loci of interest were amplified using nested PCR (primers available in table 1). Typical PCR amplification consisted of one cycle at 95°C for 1 min, then 40 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 40 s, followed by a final extension at 72°C for 5 min. Quantitative interrogation of the bisulphite-induced C→T transition at CpG dinucleotides in such amplicons was performed with primers targeted to the CpG dinucleotides within the restriction sites for *HpaII*, *Hin6I*, or *AclI*.

Results

Intra- and Interindividual DNA Methylation Differences in the Promoters of BRCA1, BRCA2, HD, DM1, PSEN1, and PSEN2

The bisulphite modification-based mapping of methylated cytosines for all of these genes demonstrated that numerous individual clones (representing individual sperm cells) demonstrated quite different DNA methylation profiles within individuals (fig. 1A). This finding was confirmed by the analysis of the degree of uniqueness of DNA methylation profiles (fig. 1B). In the case of *HD*, ~80% of all clones exhibited unique patterns of methylated cytosine distribution. This estimate did not change dramatically when potential bisulphite modification-induced artifacts were taken into account; on average, 72% of clones were different from one another when one methylated cytosine difference was tolerated, and up to 53% were different when two differences were allowed. The latter situation is a very conservative estimate of the degree of uniqueness, since such a high artifactual C→T non-conversion rate is unrealistic. In our experiments, the artifactual C→T bisulphite conversion was always <1%. The lowest degree of intraindividual DNA methylation uniqueness was detected for *PSEN2*: 36%, 20%, and 13% for 0, 1, and 2 levels of tolerance, respectively. This analysis of uniqueness is, however, related to the clone length and correlates specifically with the density of the CpGs analyzed (Pearson $R = 0.64, 0.93, \text{ and } 0.98$ for 0, 1, and 2 levels of tolerance, respectively), since more methylatable CpG sites allow more opportunity for variation.

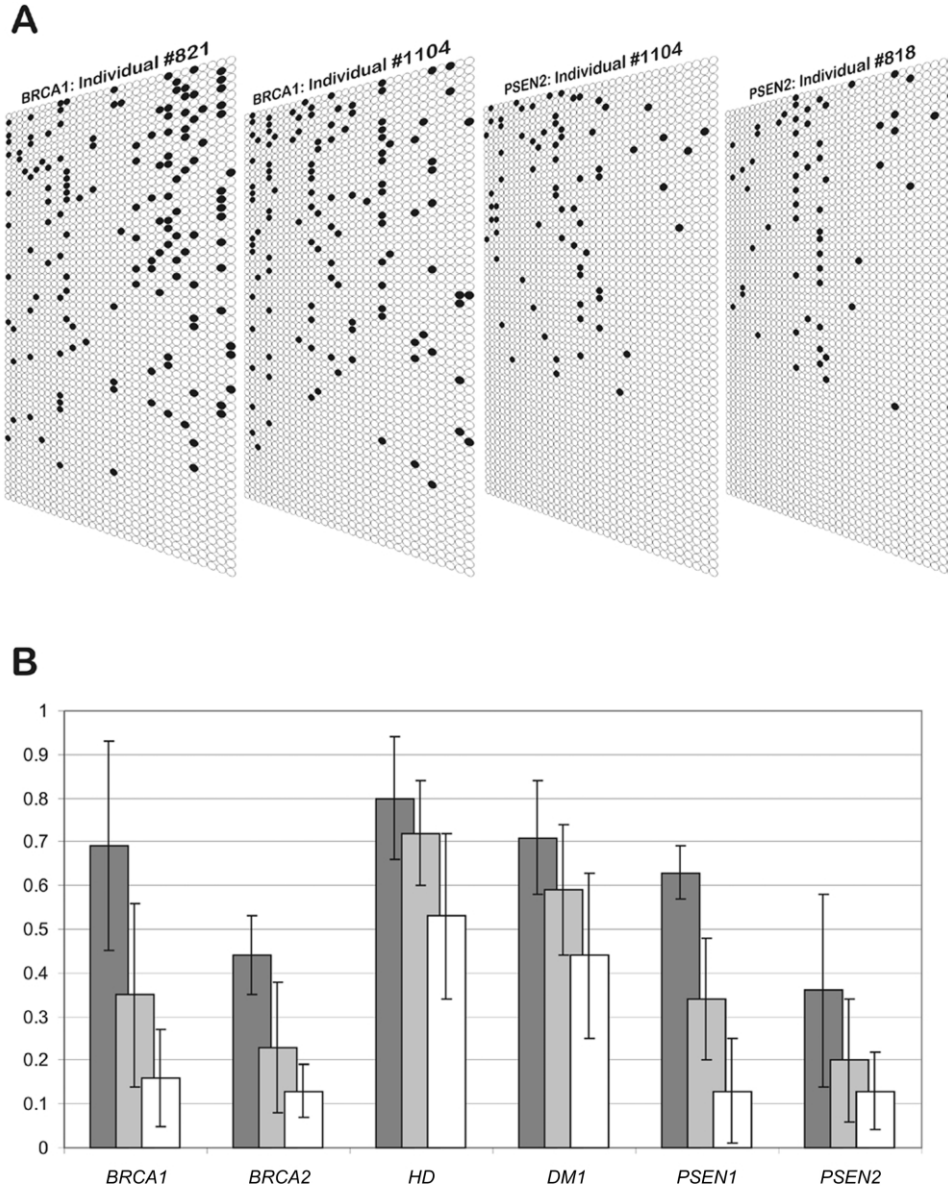


Figure 1. Intraindividual variability of DNA methylation. *A*, DNA methylation profiles of the promoter CpG islands of *BRCA1* and *PSEN2*, determined on the basis of sequencing 60 clones of bisulphite-modified sperm DNA. The *BRCA1* locus covered 32 CpGs, and the *PSEN2* region included 45 CpGs. Nine monomorphic (unmethylated) CpGs (*BRCA1* or *PSEN2*) were excluded from the figure. Each individual is represented, with individual CpG dinucleotides from left to right (*black* = methylated cytosines; *white* = unmethylated cytosines) and individual clones from top to bottom. Like the presented *BRCA1* and *PSEN2* cases, a substantial proportion of clones in other loci (*HD*, *DM1*, *BRCA2*, and *PSEN1*) revealed unique DNA methylation profiles. *B*, Estimates of the proportion of unique methylation profiles in the promoter regions of the six analyzed genes. The Y-axis shows the proportion of clones carrying unique methylation profiles over the total number of sequenced clones; the X-axis shows the proportion of unique profiles that contain at least 1, 2, and 3 differences (*left*, *middle*, and *right* bars, respectively), compared with the other profiles at the same locus in the same individual.

Whereas the intraindividual analysis can show variability within an individual, significantly variable methylation patterns between individuals were also revealed (fig. 2). The gene-specific results were as follows: in *BRCA1*, $n = 4$, 32 CpGs were analyzed, and 5/6 pairwise comparisons exhibited statistically significant differences (average $P = 2.53 \times 10^{-5}$); in *BRCA2*, $n = 4$, 36 CpGs were analyzed, and 3/6 pairwise comparisons were significant (average $P = 8.56 \times 10^{-7}$); in *PSEN1*, $n = 3$, 43 CpGs were analyzed, and 2/3 pairwise comparisons were significant (average $P = 1.89 \times 10^{-4}$); in *PSEN2*, $n = 5$, 45 CpGs were analyzed, and 6/10 pairwise comparisons were significant

(average $P = 5.11 \times 10^{-3}$); in *DM1*, $n = 7$, 99 CpGs were analyzed, and 13/21 pairwise comparisons were significant (average $P = 5.60 \times 10^{-4}$); and, in *HD*, $n = 6$, 108 CpGs were analyzed, and 12/15 pairwise comparisons were significant (average $P = 1.66 \times 10^{-3}$). Overall, 67% (41/61) of the pairwise comparisons were significantly different, which suggests a high overall level of interindividual variability in the methylation patterns of the tested genes. For the five cases in which 60 sequenced clones were available for *BRCA1*, *BRCA2*, and *PSEN2*, the comparisons were performed using two randomly selected groups of 30 clones (fig. 2). As a validation of this statisti-

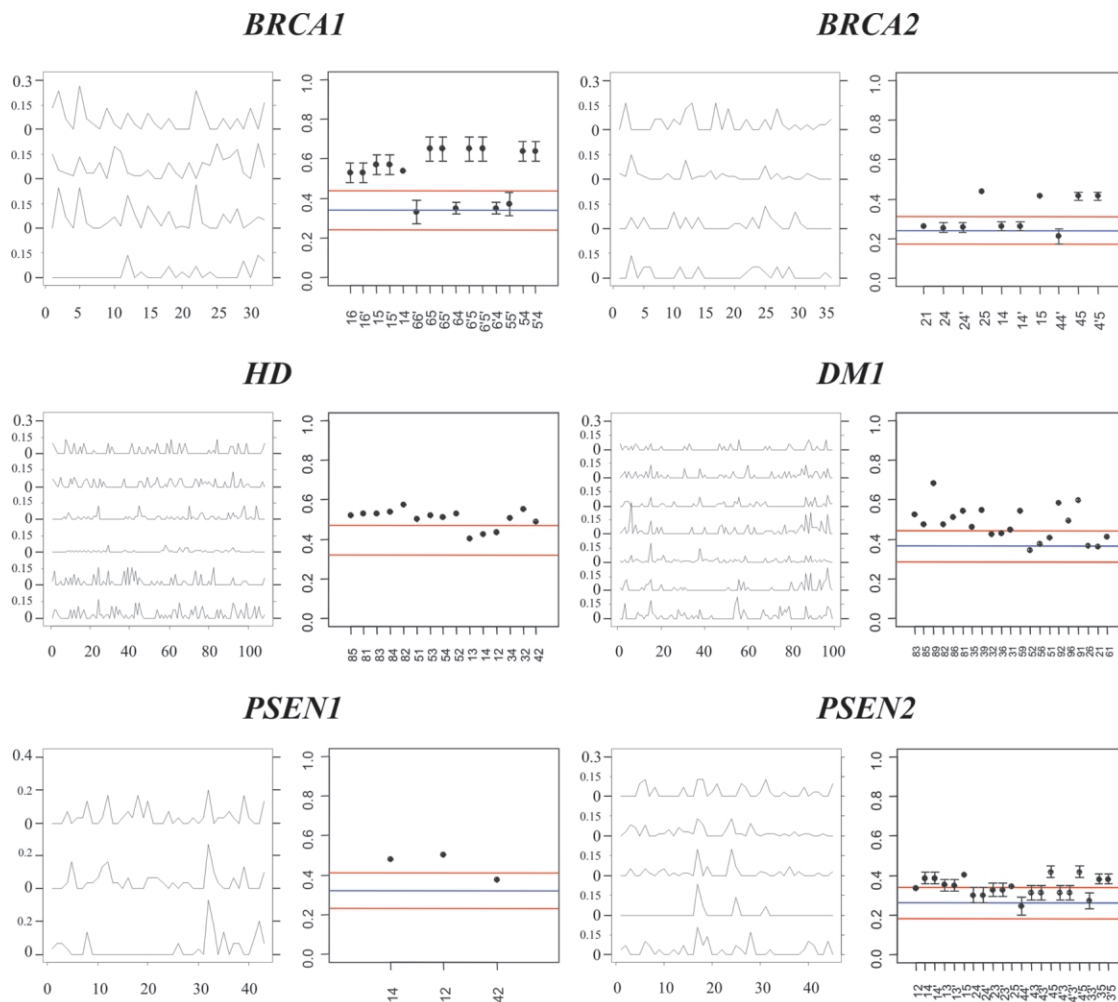


Figure 2. Interindividual variability of DNA methylation in six human disease genes. Bisulphite modification–based mapping of methylated cytosines in *BRCA1*, *BRCA2*, *HD*, *DM1*, *PSEN1*, and *PSEN2*. Thirty individual clones were sequenced from three to seven individuals. Analysis for each gene is represented in two panels. *Left panels*, graphical profile of the percentage of methylation (Y-axis, ranging from 0% to 40%) for every CpG dinucleotide (X-axis, ranging from 32 to 108 CpG dinucleotides), out of the total number of clones for each individual. *Right panels*, Euclidean distances (Y-axis) of pairwise comparisons between individual methylation profiles (X-axis). The blue line is the mean distance, and red lines are ± 2 SD from the mean, both obtained for each gene from the permutation study (see the “Material and Methods” section). Pairwise comparisons are annotated—for example, as “16”—for the comparison of the Euclidean distance of individual 1 with that of individual 6. Primed individual numbers (e.g., 4’) represent a second set of 30 clones from those individuals. The error bars on some data points represent SDs from 100,000 permutations of 30 clone groups from the individuals from whom 60 clones were sequenced.

Table 2. Statistical Analysis of Microarray Data from the Sperm DNA-HHA, Sperm DNA-HpaII, and Brain DNA-HpaII Data Sets

Descriptive Statistics	Sperm DNA-HHA (n = 25)	Sperm DNA-HpaII ^a (n = 21)	Brain DNA-HpaII (n = 22)
Mean CV (\pm SD) ^b	6.72 (2.48)	9.23 (4.92)	10.89 (5.40)
Loci count ^c	4,969	4,947	4,952
90th percentile (%) ^d	>9.53	>13.71	>16.33
10th percentile (%) ^d	<4.33	<5.16	<6.44
SNP analysis ^e :			
90th percentile:			
No. with SNPs	78	32	ND
No. with no SNPs	72	118	ND
10th percentile:			
No. with SNPs	74	22	ND
No. with no SNPs	76	128	ND
χ^2	.12	1.829	ND
P value	.729	.176	ND
CGI analysis ^f :			
Count:			
CGI	2,523	2,512	2,478
Non-CGI	2,446	2,435	2,401
Mean CV (\pm SD):			
CGI	6.69 (2.57)	9.6 (5.28)	11.06 (4.94)
Non-CGI	6.79 (2.32)	8.97 (4.39)	11.05 (5.60)
t Test P value	.14815	4.92 $\times 10^{-6}$.93995
CGI χ^2 test ^g :			
90th percentile:			
No. CGI	235	296	256
No. non-CGI	217	198	238
10th percentile:			
No. CGI	226	218	255
No. non-CGI	209	277	238
χ^2	.003	24.34	.001
P value	.955	5.81 $\times 10^{-7}$.974
Promoter χ^2 test ^g :			
90th percentile:			
No. promoter CGI	NA	245	NA
No. non-promoter CGI	NA	52	NA
10th percentile:			
No. promoter CGI	NA	152	NA
No. non-promoter CGI	NA	67	NA
χ^2	NA	11.44	NA
P value	NA	4.87 $\times 10^{-4}$	NA

^a Values in bold type are statistically significant.

^b The mean (\pm SD) of the CV in ratio Cy5/Cy3 across the individuals (n) for each data set was calculated.

^c The count represents the number of unique loci remaining after data trimming (see the "Material and Methods" section).

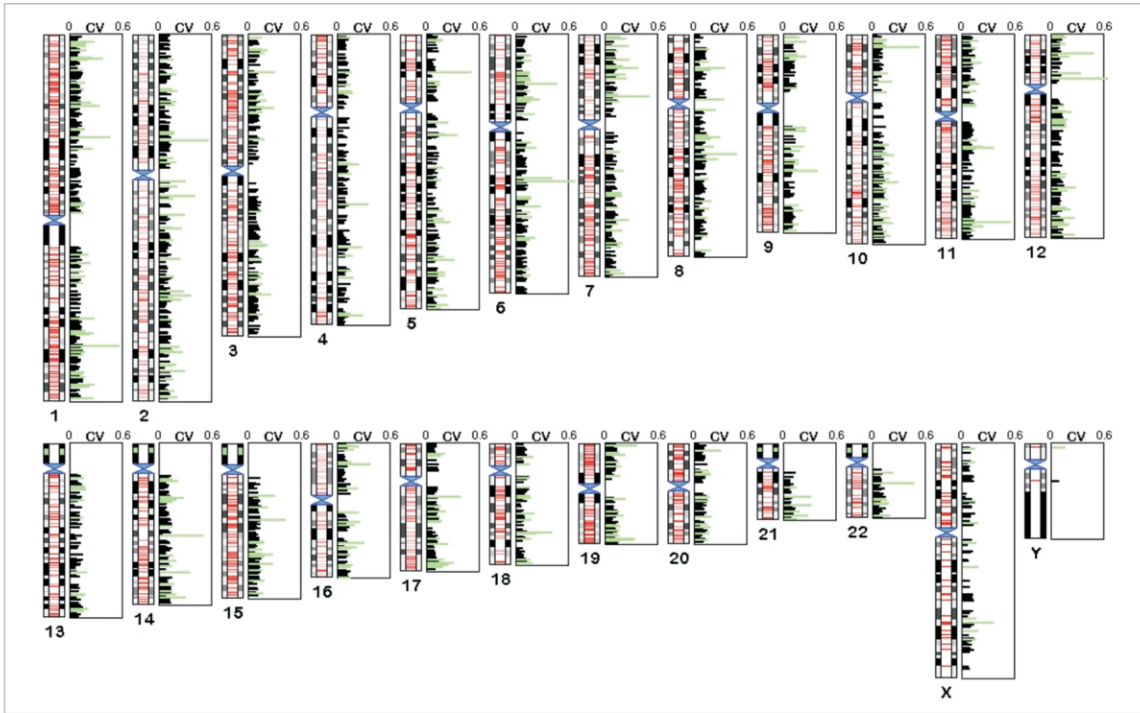
^d Loci with CVs >90th percentile are in the top 10% of methylation-variable regions, and loci with CVs <10th percentile are the least variable loci.

^e The SNP analysis was performed to test for the effects of SNPs from our DNA methylation analysis. We randomly selected 300 loci from the 10th and 90th percentile loci, and the clone sequence plus 1-kb flanking regions were screened for the presence of SNPs—in particular, SNPs that create or disrupt *HpaII*, *Hin6I*, and *AciI* restriction sites. ND = not done.

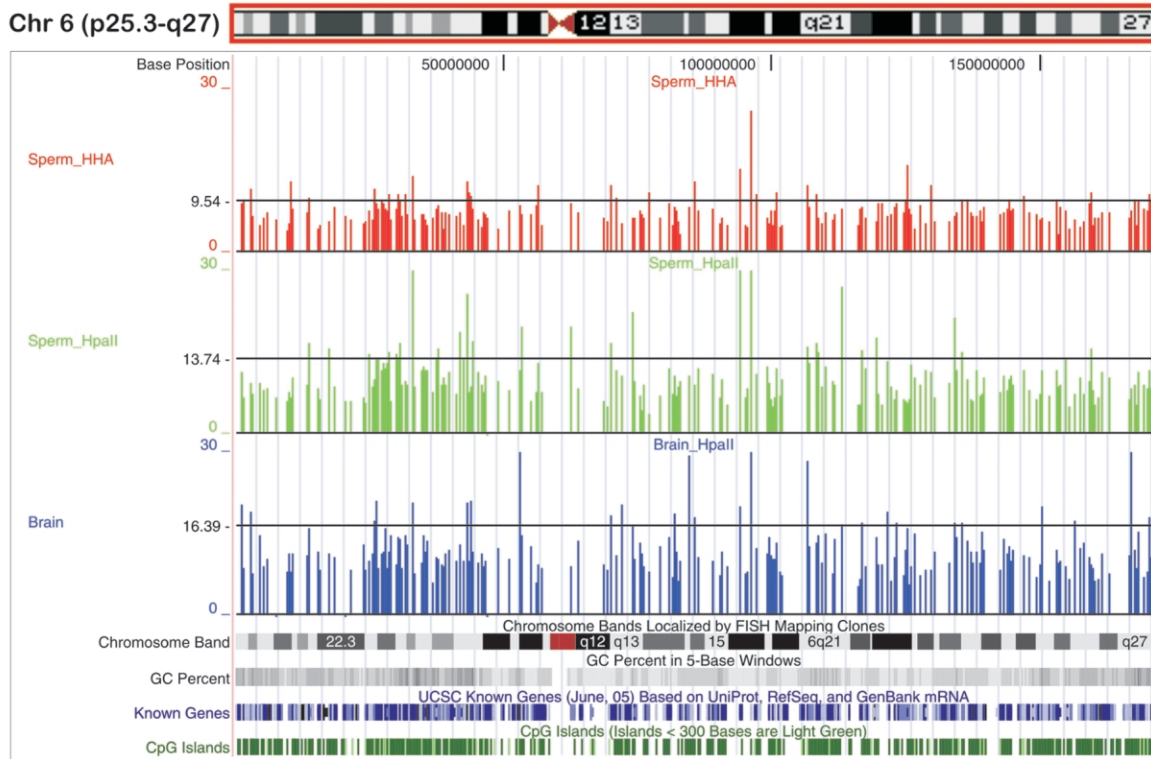
^f The CpG island (CGI) analyses were performed by separating loci into either CGI or non-CGI categories. A Student's *t* test was performed to analyze the difference in mean CV. The numbers of loci within each group in the 90th and 10th percentile loci were counted, and χ^2 analysis was performed.

^g The CGI loci were further subdivided into loci within promoter regions of genes or loci not within promoters, the numbers of loci in the 90th and 10th percentiles were counted, and χ^2 analysis was performed. NA = not applicable.

A



B



cal method, the additional sets of 30 clones representing *BRCA1*, *BRCA2*, and *PSEN2* were compared with the primary sets of 30 clones of the same individuals. In all cases, the results (compare pairs 55' and 66' in *BRCA1* in fig. 2) showed that their profiles were not different, which is to be expected since they are from the same individual.

DNA Methylation Differences Detected by the CpG Island Microarrays

This CpG island microarray contains 12,192 DNA fragments; however, unique sequences are represented by 4,970 distinct loci, of which only about half meet the commonly used criteria for CpG islands: GC content of $\geq 50\%$, length >200 bp, and observed/expected CG dinucleotide ratio >0.6 .²⁸ As described in the "Material and Methods" section, we used two strategies to increase the informativeness of our microarray analysis. The first strategy was to use an enzyme cocktail of *HpaII*, *Hin6I*, and *Acil* (sperm DNA-HHA data set), which is more informative for lower GC content loci, such as those that do not meet the CpG island criteria. The second strategy was to use *HpaII* alone (sperm DNA-HpaII data set), which would be more informative for the higher GC-containing loci. Lastly, we analyzed a brain DNA microarray data set (brain DNA-HpaII data set), to compare tissue-specific differences. As a measure of methylation variation, we have calculated the CV across individuals for each array set. The CV is calculated by dividing the SD in the Cy5/Cy3 ratio by the mean of the Cy5/Cy3 ratio, and it is expressed as a percentage. The variation in CV among individuals across the genome ranged from 2.1% to 30.5% (mean = 6.7%), from 0.8% to 66.2% (mean = 9.2%), and from 2.1% to 97.4% (mean = 10.9%) for the sperm DNA-HHA, sperm DNA-HpaII, and brain DNA-HpaII data sets, respectively (table 2). We considered the loci within the top 10% of CVs (90th percentile) as highly variable regions.

The data for each locus were plotted on the genome (figs. 3 and 4) and are also available online as a custom annotation track (Center for Addiction and Mental Health) with use of the UCSC Genome Browser (fig. 3B). Figure 3A depicts the sperm DNA-HpaII data set and highlights the highly variable regions on the genome. To assess whether this distribution of highly variable spots is non-random, we performed an autocorrelation analysis; how-

The figure is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Figure 4. Genomewide view of brain DNA-HpaII and sperm DNA-HHA data sets. The legend is available in its entirety in the online version of *The American Journal of Human Genetics*.

ever, this analysis did not identify any evidence for autocorrelation, most likely because of the large genomic distance between microarray clones (average 0.6 Mb). Other analyses included testing if the detected variability is confounded by DNA sequence variation; comparison of DNA methylation variation in CpG islands and in non-CpG islands, as well as across different classes of repetitive elements; and assessing if DNA methylation variation correlates with the GC content, clone length, or particular chromosomal cytobands.

Exclusion of Genetic Confounding Effects: SNPs and Copy-Number Polymorphisms

Any method that relies on restriction-enzyme digestion to differentiate between methylated and unmethylated DNA can be influenced by SNPs within the enzyme-restriction sites. Therefore, from each of the sperm DNA-HHA and sperm DNA-HpaII data sets, we selected 150 highly variable loci and 150 conserved loci and performed *in silico* screening, to identify all known SNPs within a 2-kb region of the selected clone that disrupt or create *HpaII*, *Hin6I*, or *Acil* enzyme sites for the sperm DNA-HHA data set or just *HpaII* sites for the sperm DNA-HpaII data set (SNP annotation of the UCSC Genome Browser). If SNPs were a significant confounding factor in DNA methylation variation, we would expect a higher proportion of SNPs in highly variable loci (e.g., 90th percentile) compared with the lowest variable loci (e.g., 10th percentile). The χ^2 analysis revealed no association between the number of potentially disruptive enzyme-restriction sites and the degree of variability in either data set (sperm DNA-HHA $\chi^2 = 0.12$, $P = .729$; sperm DNA-HpaII $\chi^2 = 1.83$, $P = .176$). This finding suggests that the degree of variability in the sperm DNA microarray analysis is more dependent

Figure 3. Chromosomal view of methylation variability by CpG island microarray analysis. *A*, Unmethylated fraction of genomic DNA extracted from sperm samples ($n = 21$) hybridized individually (Cy5), in contrast to the pooled reference control (Cy3). The CV of the Cy5/Cy3 ratio was calculated for each spot across the 21 individuals and was mapped to the corresponding genomic location. Each chromosome ideogram is overlaid with red bars that represent the position of each clone on the CpG island microarray. The bars highlighted in green are the loci that showed variance in the 90th percentile (the top 10% of loci exhibiting the largest degree of DNA methylation variation). *B*, Screenshot of the custom annotation track on the UCSC Genome Browser (available from the Center for Addiction and Mental Health). Shown is chromosome 6, which includes the major histocompatibility complex locus that was screened for epigenetic variability by the Human Epigenome Project pilot study.⁸

on DNA methylation differences than on DNA sequence differences.

Recent reports have identified >200 copy-number polymorphisms (CNPs) that represent large duplications and deletions that contribute significantly to genomic variation between individuals.^{29–31} Like SNPs, CNPs could simulate DNA methylation variability in the microarray analysis. We have cross-referenced the CNPs identified in these studies with the CpG island microarray loci and have identified 25 microarray loci that occur within known CNP regions. These include large CNPs in chromosomes 3 (covering the genes *OST α* , *ABO18337*, *UNQ3030*, *BC015560*, and *DLG1*), 16 (*BC008967*, *XYLT1*, *ARL61P*, *MIR16*, *MGC16943*, and *CDR2*), and 17 (*AY302137*, *BHD*, *RAI1*, *FLJ20308*, *TOP3A*, and *SMCR8*) and smaller CNPs on chromosomes 1 (*NEGR1*), 2 (*AK024244*), 6 (*RDBP*), 8 (*TSTA3*), 9 (*LHX2*), 11 (*TNNT3*), and 14 (*AK090461*). Microarray results for these genes listed could, therefore, be influenced by deletions or duplications as much as by methylation variability; however, none of these loci appear in the list of highly variable (>90th percentile) loci.

CpG Island Analysis

Not all DNA fragments on the CpG island microarray met the criteria for CpG islands. The list of loci were divided into “CpG islands” or “non-CpG islands,” and a Student’s *t* test was performed to test for any statistically significant difference in the mean CV (table 2). A significantly increased DNA methylation variability was found in loci defined as CpG islands in the sperm DNA-*HpaII* data set (*t* test $P = 4.92 \times 10^{-6}$), and this variability was exemplified by a bias towards CpG islands in the 90th percentile (highly variable regions) ($\chi^2 = 24.34$; $P = 5.81 \times 10^{-7}$). In addition, when the CpG islands were split into promoter CpG islands and CpG islands not associated with known gene promoters, significantly higher variability in promoter CpG islands ($\chi^2 = 11.44$; $P = 4.87 \times 10^{-4}$) was detected. However, analyses of methylation variability with other measures, including GC percent alone and clone length, did not reveal any association. No evidence for higher DNA methylation variation was detected in the promoter CpG islands in the brain DNA-*HpaII* data set, and there also was no association with SNPs. Therefore, this sperm DNA-*HpaII* experiment appears to have revealed genuine increased methylation differences in the promoter CpG islands.

Cytoband Analysis

It has been well described that different cytobands could have evolved in different ways and that the genes within each band could have evolutionary similarities.^{32,33} Since these bands are based on, among other things, GC content and *Alu* content, we sought to identify whether methylation variability was one of the aspects that showed similarities within bands. The CpG island microarray annotation includes the division of loci into different cyto-

bands, including the G bands (Giemsa negative: gneg) and the four classes of R bands (Giemsa positive: gpos25, gpos50, gpos75, and gpos100). These bands are defined as follows. The darkest R bands, gpos100, are very rich in GC and *Alu*; the next darkest, gpos75, are very rich in GC but not in *Alu*; gpos50 are not rich in GC but are rich in *Alu*; and gpos25 are Giemsa-dark bands that are rich in neither GC nor *Alu*.³⁴ Mean CVs for all the loci within each of these cytobands were calculated, and a Student’s *t* test was performed to identify any statistically significant differences. In each of the data sets, marginally significant associations with certain cytobands were identified. In the sperm DNA-*HHA* data set, significant decreases in variability between gpos75 band loci (CV = 6.51) and the other three R bands—gpos25, gpos50, and gpos100 (average CV = 6.83; average $P = .023$)—were detected. In the sperm DNA-*HpaII* data set, gpos25 exhibited a lower degree of methylation compared with gpos50 (CV = 8.97 and CV = 9.50, respectively; $P = .041$). Although the significance of these statistical tests diminished when corrected for multiple testing, the result is suggestive of an increase in variability in the *Alu*-rich cytobands, such as the gpos50 and gpos100 cytobands, compared with the *Alu*-poorer bands gpos25 and gpos75.

Age-Dependent DNA Methylation Changes in the Sperm

Methylation dynamics with age (sperm DNA-*HHA* age range 22–35 years; sperm DNA-*HpaII* age range 24–56 years) as a covariate were investigated. Individuals were ordered by increasing age, and the Pearson correlation between the age and relative methylation-signal intensity (ratio of case to reference) was calculated for each locus. In the sperm DNA-*HpaII* and sperm DNA-*HHA* data sets, 105 and 8 loci, respectively, were found, whose absolute correlation coefficients were >0.5 and whose *P* values were <.05. Numerous genes were identified in the germ cell data that corresponded to genes involved in spermatogenesis and development (e.g., *INSM1*, *TZFP*, and *EED*) and neurogenesis (e.g., *CALM1*, *STMN2*, *ARHGEF9*, and *ARX*) or to disease-related genes (e.g., *MAF*, *DCC*, and *CDH13* [MIM 601364]). A number of examples are shown in figure 5. The lists of genes for each data set are available in tables 3 and 4.

DNA Methylation in the Repetitive Elements

All the above analyses were performed on unique DNA sequences; however, the CpG island microarray also contains a large number of clones containing repetitive elements, which, as a rule, are heavily methylated.³⁵ Although it is difficult to directly distinguish between methylation and copy-number differences, one possible approach is to compare methylation of repetitive elements in the sperm to that in other tissues. For this reason, the sperm DNA-*HpaII* data set was analyzed in comparison with the brain DNA-*HpaII* data set. The microarray loci that contain a single repetitive element were separated

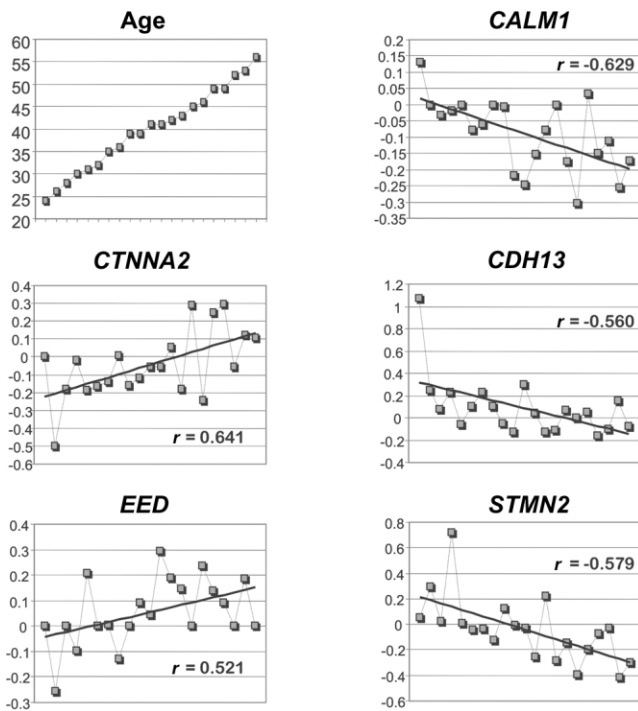


Figure 5. Age-related DNA methylation changes in the sperm. Individuals were ordered by increasing age (*top left panel*), and gene-specific DNA methylation dynamics were investigated using the individual ages (sperm DNA-*HpaII* age range 24–56 years) as a covariate. Pearson correlation was calculated for each locus, and the one-tailed *P* value of the coefficient was obtained. In the sperm DNA-*HpaII* data set, 105 loci were identified as significantly ($P < .05$) correlated ($r > 0.5$) or inversely correlated ($r < -0.5$) with age. Since the unmethylated fraction of DNA was interrogated, positive correlation indicates decreasing DNA methylation with age, whereas inverse correlation reflects increasing methylation with age. The genes *CTNNA2*, *EED*, *CALM1*, *CDH13*, and *STMN2* are shown as examples. Other genes for the sperm DNA-*HpaII*, sperm DNA-*HHA*, and brain DNA-*HpaII* data sets are available in tables 3 and 4.

into each repeat class, and the mean CV (\pm SD) was calculated. If the repetitive elements were influencing the methylation variability, one would expect that those loci containing repetitive elements would display significantly different mean CVs than those of nonrepetitive loci. This analysis revealed the overall average repetitive-element CV of 10.5 in the sperm, compared with the overall average CV in nonrepetitive elements of 9.6. The breakdown of CV for each type of repetitive element represented on the microarray is shown in figure 6. This analysis identified that satellite DNA repeats were statistically more variable than other repetitive elements in the sperm DNA-*HpaII* data set ($P = 6.12 \times 10^{-17}$). In comparison, this effect was far less pronounced in the brain DNA-*HpaII* data set ($P = .0027$) (fig. 6A). When the satellite repeats were further separated into specific repeat classes, a number of repeat classes, predominantly centromeric or peri-

Table 3. Age-Related Correlation in Sperm DNA-*HpaII* Data Set

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

centromeric satellite repeats, were identified as responsible for this increase in interindividual variability, including (GAATTC)_n (CV = 18.5), ALR/ α (human α -repetitive DNA [CV = 25.0]), *CER* (human *D22Z3*-centromeric-repetitive DNA [CV = 18.7]), and *HSATII* repeats (human satellite II DNA [CV = 34.8]) (fig. 6B).

Validation of the Microarray Data with Use of Bisulphite Modification-Based Methylated/Unmethylated Cytosine Analysis

For validation of the microarray data, 12 loci that were detected as variable in the CpG island microarray analysis (table 5) were analyzed using the MS-SNuPE reaction on the ABI SNaPshot platform²⁷ at the CpG dinucleotides in the *HpaII* and the *Hin6I* or *AcI* restriction sites. Initially, such loci were selected on the basis of increased variability (>90th percentile) in the sperm DNA-*HHA* data set; in addition, a number of these loci were also highly variable in the sperm DNA-*HpaII* data set (*CDH13*, *SCAM1*, *MKL2*, and *DIRAS3*). Each of the 12 loci selected were initially resequenced to confirm the identity of the sequence. DNA samples from 11 individuals were treated with sodium bisulphite and were PCR amplified, and primer extension reactions were performed to interrogate 65 CpG dinucleotides within the 12 sequences. Examples of six loci are presented in figure 7. This analysis revealed variable levels of methylation differences in at least one enzyme-restriction site in 11 of 12 loci tested. It should be noted here that DNA methylation differences in a single restriction site may be sufficient to generate significant differences in the microarray analysis. Only one locus (*DIRAS3*) showed no methylation differences between the 11 individuals; however, we were able to test only 5 of 20 CpG sites at this locus, so methylation variation in the untested CpG sites cannot be ruled out. To assess the replicability of the assay, we repeated the MS-SNuPE/SNaPshot experiment on five loci in five individuals. Consistent with published data,²⁷ the results in this second round of experiments were within 5% of the first experiment, on average (range 1.7%–9.9%).

Finally, for further validation of the MS-SNuPE method and microarray results, we performed bisulphite genomic sequencing of 30 clones from five individuals on a locus within the gene that encodes cadherin 13 (*CDH13* [University Health Network accession number

Table 4. Age-Related Correlation in Sperm DNA-*HHA* Data Set

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

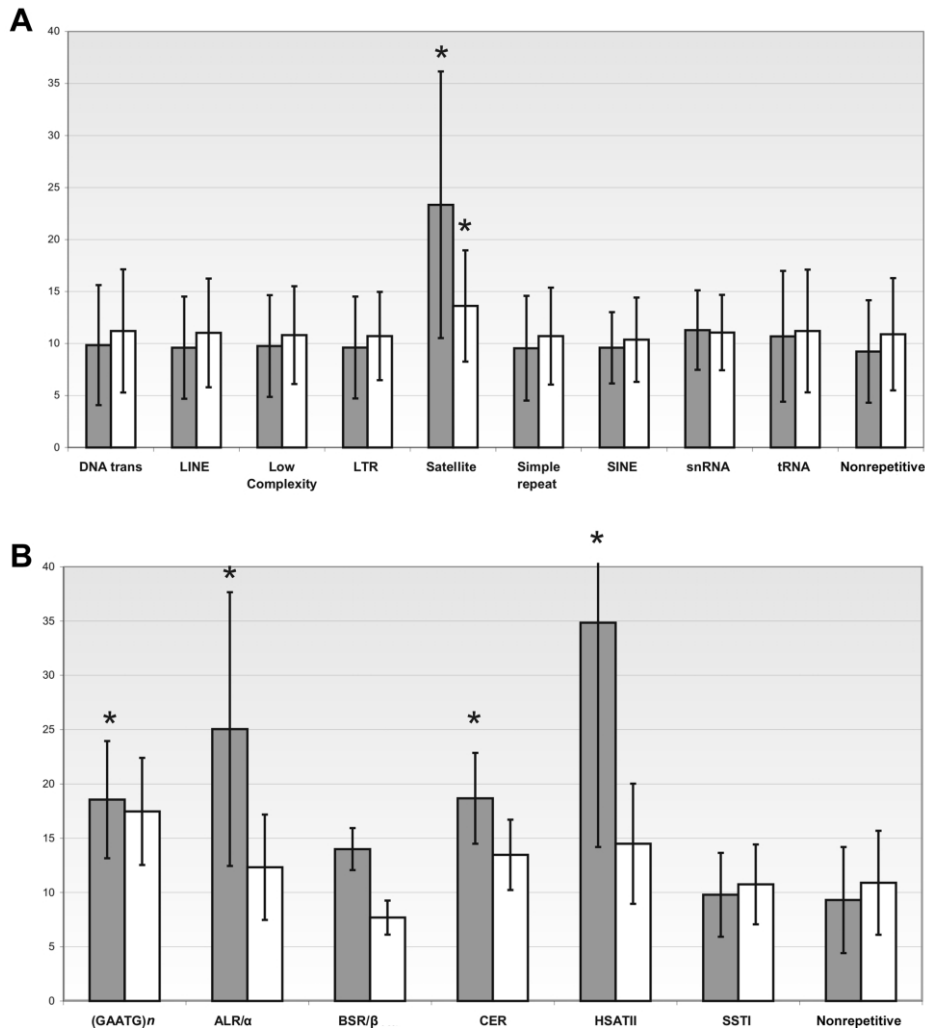


Figure 6. Repetitive-element analysis in sperm DNA-*HpaII* and brain DNA-*HpaII* data sets. The microarray loci that contain a single repetitive element were separated into each repeat class, and the mean CV (\pm SD) was calculated (A). The repeat classes include DNA transposons ($n = 209$), long interspersed transposable elements (LINEs [$n = 771$]), low-complexity repeats ($n = 461$), long terminal repeats (LTRs [$n = 360$]), satellites ($n = 208$), simple repeats ($n = 346$), SINEs ($n = 1,058$), small nuclear RNA (snRNA [$n = 30$]), and tRNA ($n = 40$), and the nonrepetitive loci ($n = 6,976$) are presented for comparison. The satellite repeats were the only class to show significantly increased variability in the sperm DNA-*HpaII* ($P = 6.12 \times 10^{-17}$) and less-significantly increased variability in the brain DNA-*HpaII* ($P = .0027$) data sets. B, Breakdown of the satellite repeats into specific satellite-repeat classes, which reveals a number of repeat classes with increased variability—predominantly, the centromeric satellite repeats, including (GAATTC)_n ($P = 8.44 \times 10^{-17}$; $n = 55$), ALR/α ($P = 4.08 \times 10^{-25}$; $n = 119$), CER ($P = .0026$; $n = 6$), and HSATII repeats ($P = 3.91 \times 10^{-5}$; $n = 19$) but not BSR/β repeats ($P > .05$; $n = 7$).

UHNhscpg0004063]). This analysis revealed a clear-cut bimodal distribution of epialleles, with the majority of clone sequences being either mostly methylated across all 16 CpG dinucleotides tested or predominantly unmethylated. In addition, this sequencing analysis identified a SNP, C/G—also identified in dbSNP as *rs16961372*—with a rare C allele frequency of 0.396 in whites. Of the five individuals sequenced, one was homozygous C, one was homozygous G, and the other three were C/G heterozygous (fig. 8). Of particular interest is the substantially higher density of methylated cytosines on the G allele,

whereas the C alleles predominantly exhibit a low degree of methylation. Counting all clones across all five individuals together, we found that 67 (77%) of 87 of the sequences with the G allele were methylated, whereas only 14 (22%) of 63 sequences containing the C allele were methylated ($\chi^2 = 40.4$; $P = 2.08 \times 10^{-10}$).

Given that the microarray analysis suggested that promoter CpG islands were significantly more variable, we also performed bisulphite genomic sequencing of the promoter CpG island of *CDH13*, which was not represented on the CpG microarray. This analysis, however, found that

Table 5. List of Clones Selected for Bisulfite-Modification MS-SNuPE Analysis

University Health Network Accession Number	Gene	Gene Description	Chromosome Location	Total Enzyme CpGs ^a	CpGs Tested ^b	Variable Methylated CpGs ^c
UHNhscpg0004931	<i>OLR1</i>	Oxidized low-density lipoprotein receptor 1	12p13.2	12	11	9
UHNhscpg0004063	<i>CDH13</i>	Cadherin 13 preproprotein	16q23.3	8	8	8
UHNhscpg0002847	<i>SCAM1</i>	Sorbin and SH3 domain-containing 3	8p21.3	13	6	5
UHNhscpg0003990	<i>NELL2</i>	NEL-like 2 (chicken)	12q12	2	2	2
UHNhscpg0003907	<i>NEIL2</i>	Nei-like 2 (<i>Escherichia coli</i>)	8p23.1	7	6	4
UHNhscpg0001947	<i>MKL2</i>	Megakaryoblastic leukemia 2	16p13.12	10	5	2
UHNhscpg0000823	<i>2-PDE</i>	2'-Phosphodiesterase	3p14.3	20	6	1
UHNhscpg0004641	<i>RHOQ</i>	Ras-related GTP-binding protein TC10	2p21	12	6	1
UHNhscpg0002006	<i>DIRAS3</i>	Ras homolog gene family, member I	1p31.2	20	5	0
UHNhscpg0005090	<i>AHR</i>	RWD domain-containing 3	1p21.3	3	3	2
UHNhscpg0009548	<i>DSCAM</i>	Down syndrome cell-adhesion molecule	21q22.2	4	3	3
UHNhscpg0004745	<i>FBN1</i>	Fibrillin 1 (Marfan syndrome)	15q21.1	2	2	2

^a The number of CpG sites within the recognition sequence of restriction enzymes *HpaII*, *Hin6I*, and *AclI*.

^b The number of CpG sites tested by MS-SNuPE.

^c The number of CpG sites that showed variable methylation between individuals.

the promoter CpG island of *CDH13* is predominantly unmethylated in all individuals, with only solitary methylation sites present in one to three clones for each of the individuals.

Discussion

In this study, we performed an in-depth analysis to address the question of epigenetic variability in the germline. The main conclusions are that (1) the male germline exhibits locus-, cell-, and age-dependent DNA methylation differences and that (2) DNA methylation variation is significant across unrelated individuals, at a level that, by far, exceeds DNA sequence variation. These findings are interesting from both basic molecular biological and biomedical points of view.

First, our study contributes to the understanding of epigenetic peculiarities of gene regulatory regions in the germline. It has been generally accepted that CpG islands are predominantly unmethylated,³⁶ which implies that DNA methylation differences would not be expected there. From our studies, we find that even relatively low densities of methylated cytosines in the CpG islands are sufficient to generate unique epigenetic profiles in DNA regions that do not exhibit any DNA sequence variation, both in different cells of the same individual and also across individuals. Fine-mapping of methylated cytosines of relatively short DNA fragments of *BRCA1*, *BRCA2*, *PSEN1*, *PSEN2*, *DM1*, and *HD* suggest that each sperm cell is unique not only in terms of DNA sequence but also in epigenomic profile, and variation of the latter by far exceeds the former.

At the genomewide level, unexpectedly, promoter CpG islands exhibited larger interindividual variation compared with other single-copy DNA sequences, including the non-promoter CpG islands. This epigenetic phenomenon seems to be discordant, with a general rule that functionally important loci exhibit a low degree of DNA var-

iation, as is seen in the case of SNPs being less common in promoters and exonic sequences than in introns and intergenic regions. In addition, promoter CpG-rich regions are often highly conserved between species; for instance, the mouse genome contains 15,500 CpG islands, of which ~10,000 are highly conserved.³⁷ Therefore, if the epigenetic variability were just “noise” of little functional relevance, one would expect more variability in these less biologically important regions, such as introns and intergenic sequences. Evidence of the opposite—increased epigenetic variability in the regions that directly control gene activity—may indicate some peculiarities of DNA methylation machinery during gametogenesis that may or may not be of functional importance in the somatic cells (see below for discussion of the postzygotic [in]stability of inherited epigenetic profiles).

Our study has also identified a larger degree of interindividual variability of centromeric satellite repeats. Although we cannot strictly rule out the possibility of DNA copy differences, which are common in centromeric satellite repeats,³⁸ the fact that the germ cell data set showed substantially larger CVs in comparison with the brain DNA data set suggests that germline satellite methylation differences in the germ cells could be a genuine biological phenomenon. Interindividual methylation variability in satellite repeats is consistent with current knowledge³⁹ and may contribute to phenotypic variability in immunodeficiency-centromeric instability-facial anomalies syndrome (ICF [MIM 242860]), a disease that is associated with methylation defects in pericentromeric satellites.⁴⁰ In addition, microRNAs (or siRNAs) regulate gene expression, heterochromatin formation, and genome stability and often arise from demethylation of tandem repeats that are common in pericentromeric sequences.⁴¹ Therefore, interindividual methylation variability in tandem repeats that give rise to microRNAs could also be involved in the variability in gene expression that results in inherited phenotypic variation. A recent study has described increased

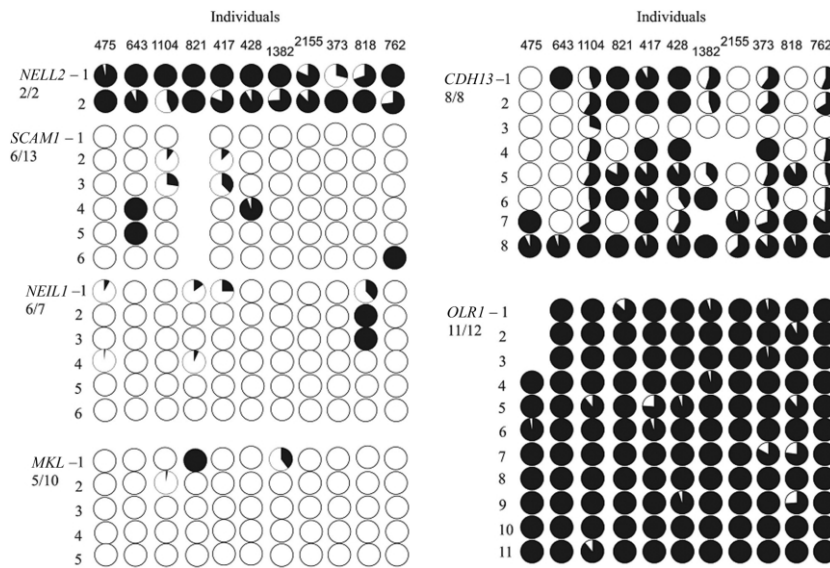


Figure 7. MS-SNuPE analysis of densities of methylated cytosines in CpG dinucleotides of selected genes. Genomic DNA from 11 individuals was treated with sodium bisulphite and then was PCR amplified for each gene. The genes *NELL2*, *SCAM1*, *NEIL1*, *MKL1*, *CDH13*, and *OLR1* are represented. The methylation status of CpG dinucleotides within each of the restriction-enzyme sites was interrogated using the primer-extension reactions. Methylation of each of the CpG dinucleotides is represented as a percentage of methylated PCR products: completely unmethylated (white circles), partially methylated (partially black circles), or completely methylated (black circles).

interindividual variability in the methylation of *Alu* repeats⁴² in whole-blood DNA, a finding that was not obvious in our analysis of germ cells, in which short interspersed transposable elements (SINEs) were not statistically different from nonrepetitive elements. However, Sandovici et al. noted that the parental-origin differences in methylation were identified only for *Alu* elements in pericentromeric chromosomal bands,⁴² which is consistent with our results.

Second, epigenetic variation within and across germline samples could be of significant interest in human morbid genetics, which, thus far, has nearly exclusively concentrated on DNA sequence differences. Inherited epigenetic variation may provide the basis for new hypotheses and experimental designs in the studies of various human diseases, where the traditional DNA sequence-based studies are reaching the limit of explanatory power. For example, although Huntington disease (HD) is caused by trinucleotide-repeat expansion in the *HD* gene, the correlation between the number of trinucleotide repeats and age at onset for patients with later-onset HD (at age >50 years) is low.⁴³ The epigenetic status of the *HD* promoter region may contribute to the steady-state *HD* mRNA levels and, therefore, to the production of toxic polyglutamine-containing proteins. HD genes containing identical trinucleotide-repeat expansion but differential DNA methylation and chromatin compaction in the promoter region may exhibit significant differences, in terms of their pathogenic potential reflected in the age at disease onset and severity of disease.

The role of differential germline epigenetic modification

in complex non-Mendelian disease may be even more critical. Despite significant efforts over the past several decades, DNA sequence-based risk factors have been uncovered in only a small fraction of complex diseases, such as familial breast cancer and early-onset Alzheimer disease. For a number of complex diseases, genetic epidemiological studies showed that DNA sequence differences account for only a small proportion of phenotypic variance among relatives, whereas the substantial remaining fraction of phenotypic differences (in some cancers, 58%–82%⁴⁴) are typically attributed to environment. Identification of causal environmental factors is very difficult because methodologically impeccable designs in epidemiological studies, as a rule, cannot be applied to humans.⁴⁵ At the same time, there is an increasing body of evidence that environmental factors play a minimal role in a number of complex traits and disease conditions.⁴⁶ In this context, epigenetic variation in the germline arises as a new molecular mechanism that may help the understanding of complex phenotypes that are not the outcome of DNA sequence variation or differential environment. The recent finding of germline epimutations of *MLH1* in three individuals affected with multiple cancers^{11,47} provides a good starting point for a systematic search for disease-specific epimutations in the germline.¹²

In our bisulphite modification-based analyses, the overwhelming majority of loci exhibited rather subtle DNA methylation differences (“shades of gray” type), whereas methylation of the locus within *CDH13* is clearly bimodal (“black or white” type). The cadherin gene is a putative mediator of cell-cell interaction in the heart and may act

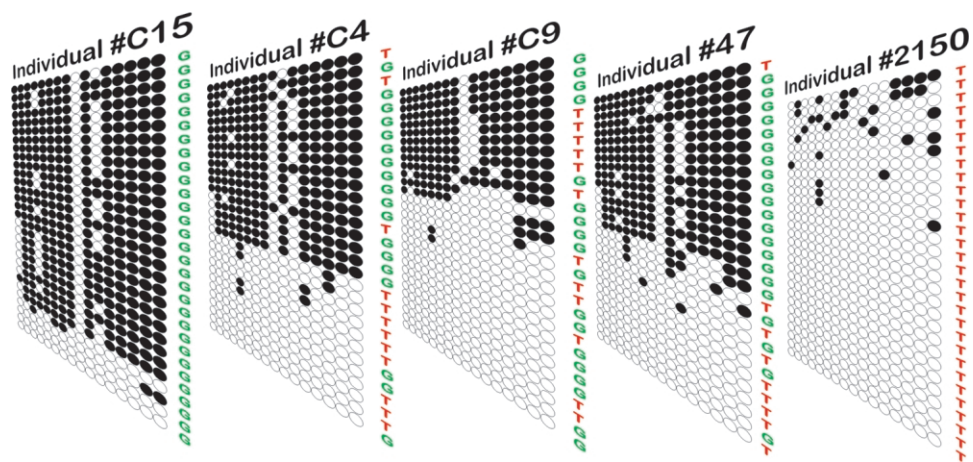


Figure 8. Methylation profiles of *CDH13*. Methylation status of 16 CpG sites surrounding the *CDH13* C/G SNP across 30 clones sequenced in each of five tested individuals. Seventy-seven percent (67/87) of the G alleles are methylated (four or more methylated CpGs), whereas 78% (49/63) of the C (bisulphite-converted to T) alleles are unmethylated. The first seven CpG dinucleotides interrogated by MS-SNuPE in figure 7 are represented in this figure as CpGs 5, 6, 9, 10, 13, 15, and 16. CpG 9 is the third MS-SNuPE primer that was predominantly unmethylated in all individuals. Each individual is represented, with single CpG dinucleotides from left to right (*black* = methylated; *white* = unmethylated) and with individual clones from top to bottom.

as a negative regulator of neural cell growth. This gene is not imprinted; however, the promoter is hypermethylated in numerous cancers.^{48–53} Of particular interest is the finding that DNA methylation profiles are associated with DNA alleles; the C allele of *CDH13* is predominantly unmethylated, whereas the G allele is predominantly methylated. To our knowledge, only a few human studies have identified associations between DNA sequence and epigenetic profiles. In Beckwith-Wiedemann syndrome, loss of maternal allele-specific methylation was more common on the G allele at the T382G SNP (CAGA haplotype) of the differentially methylated region *KvDMR1*.⁵⁴ A common variant, 677C→T, at the 5' 10-methylenetetrahydrofolate reductase gene (*MTHFR*), is associated with an increased risk of imprinting defects in the Prader-Willi syndrome/Angelman syndrome region of 15q.⁵⁵ A comprehensive screen of chromosome 21q has identified a single CpG island with a C/G SNP that was methylated in peripheral blood DNA on the C allele, regardless of the parent of origin.⁵⁶ Finally, the C102T polymorphism in the serotonin 5-HT_{2A} receptor gene (*SHT2AR*), which has been associated with several psychiatric disorders, was methylated specifically on the C allele.⁵⁷ In mice, a recent study identified a number of genes that, on in vitro mutation, affect epigenetic reprogramming during gametogenesis and early development on a genomewide level, suggesting a further mechanism by which DNA sequence (mutations) *in trans* can affect the epigenetic state.⁵⁸ A comprehensive epigenetic analysis of SNPs is warranted, and this effort may shed a new light on rather inconsistent genetic association studies in complex disease. Epialleles and epihaplotypes that combine both DNA sequence and epigenetic information may be better predictors of the risk

for a disease than any of the two components analyzed separately.

A number of genes in the sperm exhibited DNA methylation changes that correlate with age (fig. 5 and tables 3 and 4). This finding is particularly interesting in light of the evidence that older paternal age is associated with risk for schizophrenia in the offspring.^{59,60} Although it has been hypothesized that such effects could be due to epigenetic changes in the paternal genome, no locus-specific and age-dependent epigenetic changes in the human male germline have been identified thus far. In this study, a number of genes that show age-related changes in their DNA methylation have been detected, including a number of important developmental genes. The embryonic ectoderm development gene, *EED*, is a polycomb group gene involved in maintaining the epigenetically regulated repressive state of developmental genes over successive cell generations.⁶¹ *CTNNA2*, or catenin, is a neuronal cadherin-associated protein and may play a major role in the folding and lamination of the cerebral cortex.⁶² *CALM1*, or calmodulin, is a key calcium-modulated protein that functions in growth and in the cell cycle, as well as in signal transduction and in the synthesis and release of neurotransmitters. *STMN2*, or stathmin-like 2, is a neuronal growth-associated protein that shares significant amino acid-sequence similarity with the phosphoprotein stathmin, and *CDH13*, as described above, is the heart cadherin and is hypermethylated in a number of cancers.

All the above phenotype-related aspects were discussed under the assumption that the epigenetic peculiarities of the germline are, at least to some extent, reflected in the somatic cells after birth. What proportion and to what extent these inherited epigenetic signals can “survive” the

reprogramming that immediately follows fertilization, as well as during the later stages of embryogenesis,^{13,63,64} remain to be investigated. The methylation clearing is not complete and, on a global DNA level, is reduced to ~10%.^{65,66} That could represent 90% of all methylation for each gene being erased or could mean that 90% of methylated genes are completely cleared and that 10% of genes retain their methylation, or there could be numerous combinations of the two. It is also unknown what happens to the histone modifications through these phases of loss of DNA methylation signals. Since modifications of DNA and histones are codependent, even if the DNA methylation signals are erased, the histones may be able to carry on specific epigenetic messages to the next stage until the DNA gets remethylated. This concept of cellular memory through histone modifications has been demonstrated for polycomb group proteins through H3K27 trimethylation.⁶⁷ A combined analysis of both histone modifications and DNA methylation dynamics, from zygote to postnatal stage, is required for the understanding of the importance of germline epigenetics to phenotypic outcomes.

The second aspect that will determine biological importance of the epigenetic variation in the germline is transgenerational epigenetic inheritance: can complex DNA methylation patterns, at least to some extent, be inherited from the parents and transmitted to the offspring? There is already experimental evidence demonstrating epigenetic meiotic inheritance across different species, such as yeast,⁶⁸ Arabidopsis,⁶⁹ Drosophila,^{70,71} and mice.^{20,21} Although there is no doubt that transgenerational epigenetic inheritance does exist, it is not clear if this is limited to a few loci or if it is a common genome-wide phenomenon.

Acknowledgments

This research has been supported by the Special Initiative grant from the Ontario Mental Health Foundation, Canadian Institutes for Health and Research, National Institute of Mental Health, and by the National Alliance for Research on Schizophrenia and Depression, the Stanley Foundation, and the Crohn's and Colitis Foundation of Canada. We acknowledge Sigrid Ziegler for technical assistance.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Center for Addiction and Mental Health: Epigenomics, <http://www.epigenomics.ca> (for the online data linking the germline epigenetic variation to the genome, by use of the UCSC Genome Browser)

dbSNP, <http://www.ncbi.nlm.nih.gov/SNP/> (for *rs16961372*)

Entrez Gene, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene> (for *PSEN1* [accession number 5663], *PSEN2* [accession number 5664], *BRCA1* [accession number 672], *BRCA2* [accession number 675], *HD* [accession number 3064], *DM1* [accession number 1760], *CDH13* [accession number 1012], and genes in tables 3 and 4)

HapMap, <http://www.hapmap.org/>

Human Epigenome Project, <http://www.epigenome.org/index.php>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for *PSEN1*, *PSEN2*, *BRCA1*, *BRCA2*, *DM1*, *HD*, *CDH13*, and ICF)

UCSC Genome Browser, <http://genome.ucsc.edu/>

University Health Network CpG Island Microarray Database, <http://data.microarrays.ca/cpg/index.htm>

References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al (2001) The sequence of the human genome. *Science* 291:1304–1351
- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. *Nature* 437:1299–1320
- Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87
- Henikoff S, Matzke MA (1997) Exploring and explaining epigenetic effects. *Trends Genet* 13:293–295
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915–926
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33:S245–S254
- Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A, Fischer J, Gut IG, Berlin K, Beck S (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2:e405
- Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3:662–673
- Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2:21–32
- Suter CM, Martin DI, Ward RL (2004) Germline epimutation of *MLH1* in individuals with multiple cancers. *Nat Genet* 36:497–501
- Martin DI, Ward R, Suter CM (2005) Germline epimutation: a basis for epigenetic disease in humans. *Ann N Y Acad Sci* 1054:68–77
- Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293:1089–1093
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23:314–318
- Allegrucci C, Thurston A, Lucas E, Young L (2005) Epigenetics and the germline. *Reproduction* 129:137–149
- Kimmins S, Sassone-Corsi P (2005) Chromatin remodelling and epigenetic features of germ cells. *Nature* 434:583–589
- Zalensky AO, Siino JS, Gineitis AA, Zalenskaya IA, Tomilin NV, Yau P, Bradbury EM (2002) Human testis/sperm-specific histone H2B (hTSH2B): molecular cloning and characterization. *J Biol Chem* 277:43474–43480
- Churikov D, Zalenskaya IA, Zalensky AO (2004) Male germ-

- line-specific histones in mouse and man. *Cytogenet Genome Res* 105:203–214
19. Churikov D, Siino J, Svetlova M, Zhang K, Gineitis A, Morton Bradbury E, Zalensky A (2004) Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics* 84:745–756
 20. Rakyán V, Whitelaw E (2003) Transgenerational epigenetic inheritance. *Curr Biol* 13:R6
 21. Chong S, Whitelaw E (2004) Epigenetic germline inheritance. *Curr Opin Genet Dev* 14:692–696
 22. Hajkova P, el-Maarri O, Engemann S, Oswald J, Olek A, Walter J (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol Biol* 200:143–154
 23. Yatabe Y, Tavaré S, Shibata D (2001) Investigating stem cells in human colon by using methylation patterns. *Proc Natl Acad Sci USA* 98:10839–10844
 24. Heisler LE, Torti D, Boutros PC, Watson J, Chan C, Winegarden N, Takahashi M, Yau P, Huang TH, Farnham PJ, Jurisica I, Woodgett JR, Bremner R, Penn LZ, Der SD (2005) CpG island microarray probe sequences derived from a physical library are representative of CpG islands annotated on the human genome. *Nucleic Acids Res* 33:2952–2961
 25. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, Virtanen C, Winegarden N, Cheng J, Gingeras T, Petronis A (2006) Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res* 34:528–542
 26. Dudoit S, Fridlyand J (2002) A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biol* 3:RESEARCH0036
 27. Kaminsky ZA, Assadzadeh A, Flanagan J, Petronis A (2005) Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. *Nucleic Acids Res* 33:e95
 28. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
 29. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528
 30. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37:727–732
 31. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951
 32. Craig JM, Bickmore WA (1993) Chromosome bands: flavours to savour. *Bioessays* 15:349–354
 33. Furey TS, Haussler D (2003) Integration of the cytogenetic map with the draft human genome sequence. *Hum Mol Genet* 12:1037–1044
 34. Holmquist GP (1992) Chromosome bands, their chromatin flavors, and their functional features. *Am J Hum Genet* 51:17–37
 35. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335–340
 36. Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321:209–213
 37. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562
 38. Jabs EW, Goble CA, Cutting GR (1989) Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. *Proc Natl Acad Sci USA* 86:202–206
 39. Miniou P, Jeanpierre M, Bourc'his D, Coutinho Barbosa AC, Blanquet V, Viegas-Pequignot E (1997) α -Satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. *Hum Genet* 99:738–745
 40. Gisselsson D, Shao C, Tuck-Muller CM, Sogorovic S, Palsson E, Smeets D, Ehrlich M (2005) Interphase chromosomal abnormalities and mitotic missegregation of hypomethylated sequences in ICF syndrome cells. *Chromosoma* 114:118–126
 41. Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431:364–370
 42. Sandovici I, Kassovska-Bratinova S, Loredó-Ostí JC, Leppert M, Suarez A, Stewart R, Bautista FD, Schiraldi M, Sapienza C (2005) Interindividual variability and parent of origin DNA methylation differences at specific human *Alu* elements. *Hum Mol Genet* 14:2135–2143
 43. Wexler NS, Lorimer J, Porter J, Gomez F, Moskowitz C, Shackell E, Marder K, et al (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci USA* 101:3498–3503
 44. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343:78–85
 45. Taubes G (1995) Epidemiology faces its limits. *Science* 269:164–169
 46. Wong AH, Gottesman II, Petronis A (2005) Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum Mol Genet* 14:R11–R18
 47. Hitchins M, Williams R, Cheong K, Halani N, Lin VA, Packham D, Ku S, Buckle A, Hawkins N, Burn J, Gallinger S, Goldblatt J, Kirk J, Tomlinson I, Scott R, Spigelman A, Suter C, Martin D, Suthers G, Ward R (2005) MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* 129:1392–1399
 48. Hibi K, Kodera Y, Ito K, Akiyama S, Nakao A (2004) Methylation pattern of *CDH13* gene in digestive tract cancers. *Br J Cancer* 91:1139–1142
 49. Ogama Y, Ouchida M, Yoshino T, Ito S, Takimoto H, Shiote Y, Ishimaru F, Harada M, Tanimoto M, Shimizu K (2004) Prevalent hyper-methylation of the *CDH13* gene promoter in malignant B cell lymphomas. *Int J Oncol* 25:685–691
 50. Sakai M, Hibi K, Koshikawa K, Inoue S, Takeda S, Kaneko T, Nakao A (2004) Frequent promoter methylation and gene silencing of *CDH13* in pancreatic cancer. *Cancer Sci* 95:588–591
 51. Hibi K, Nakayama H, Kodera Y, Ito K, Akiyama S, Nakao A (2004) *CDH13* promoter region is specifically methylated in poorly differentiated colorectal cancer. *Br J Cancer* 90:1030–1033
 52. Widschwendter A, Ivarsson L, Blassnig A, Müller HM, Fiegl H, Wiedemair A, Müller-Holzner E, Goebel G, Marth C, Widschwendter M (2004) CDH1 and CDH13 methylation in se-

- rum is an independent prognostic marker in cervical cancer patients. *Int J Cancer* 109:163–166
53. Roman-Gomez J, Castillejo JA, Jimenez A, Cervantes F, Boque C, Hermosin L, Leon A, Granena A, Colomer D, Heiniger A, Torres A (2003) Cadherin-13, a mediator of calcium-dependent cell-cell adhesion, is silenced by methylation in chronic myeloid leukemia and correlates with pretreatment risk profile and cytogenetic response to interferon alfa. *J Clin Oncol* 21:1472–1479
 54. Murrell A, Heeson S, Cooper WN, Douglas E, Apostolidou S, Moore GE, Maher ER, Reik W (2004) An association between variants in the *IGF2* gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum Mol Genet* 13:247–255
 55. Zogel C, Bohringer S, Gross S, Varon R, Buiting K, Horsthemke B (2006) Identification of cis- and trans-acting factors possibly modifying the risk of epimutations on chromosome 15. *Eur J Hum Genet* (<http://www.nature.com/ejhg/journal/vaop/ncurrent/full/5201602a.html>) (electronically published April 5, 2006; accessed May 22, 2006)
 56. Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, Mukai T, Sakaki Y, Ito T (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res* 14:247–266
 57. Poleskaya OO, Aston C, Sokolov BP (2006) Allele C-specific methylation of the 5-HT2A receptor gene: evidence for correlation with its expression and expression of DNA methylase *DNMT1*. *J Neurosci Res* 83:362–373
 58. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JJ, Arkell R, Whitelaw E (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci USA* 102:7629–7634
 59. Malaspina D, Harlap S, Fennig S, Heiman D, Nahon D, Feldman D, Susser ES (2001) Advancing paternal age and the risk of schizophrenia. *Arch Gen Psychiatry* 58:361–367
 60. Byrne M, Agerbo E, Ewald H, Eaton WW, Mortensen PB (2003) Parental age and risk of schizophrenia: a case-control study. *Arch Gen Psychiatry* 60:673–678
 61. Montgomery ND, Yee D, Chen A, Kalantry S, Chamberlain SJ, Otte AP, Magnuson T (2005) The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* 15:942–947
 62. Smith A, Bourdeau I, Wang J, Bondy CA (2005) Expression of catenin family members CTNNA1, CTNNA2, CTNNB1 and JUP in the primate prefrontal cortex and hippocampus. *Brain Res Mol Brain Res* 135:225–231
 63. Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J (2000) Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* 10:475–478
 64. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T (2000) Demethylation of the zygotic paternal genome. *Nature* 403:501–502
 65. Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20:116–117
 66. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 117:15–23
 67. Czermin B, Imhof A (2003) The sounds of silence: histone deacetylation meets histone methylation. *Genetica* 117:159–164
 68. Klar AJ (1998) Propagating epigenetic states through meiosis: where Mendel's gene is more than a DNA moiety. *Trends Genet* 14:299–301
 69. Mittelsten Scheid O, Afsar K, Paszkowski J (2003) Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. *Nat Genet* 34:450–454
 70. Cavalli G, Paro R (1998) The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93:505–518
 71. Cavalli G, Paro R (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* 286:955–958