

A systematic search for DNA methyltransferase polymorphisms reveals a rare *DNMT3L* variant associated with subtelomeric hypomethylation

Osman El-Maarri^{1,*}, Michael S. Kareta^{4,†}, Thomas Mikeska^{2,†‡}, Tim Becker^{3,†},
Amalia Diaz-Lacava³, Judith Junen¹, Nicole Nüsgen¹, Frank Behne¹, Thomas Wienker³,
Andreas Waha², Johannes Oldenburg¹ and Frédéric Chédin⁴

¹Institute of Experimental Hematology and Transfusion Medicine, ²Department of Neuropathology and ³Institute of Medical Biometry, Informatics and Epidemiology (IMBIE), University of Bonn, Sigmund-Freud-Str. 25, Bonn 53127, Germany and ⁴Department of Molecular and Cellular Biology, University of California, Davis, CA, USA

Received November 18, 2008; Revised and Accepted February 20, 2009

Causes underlying inter-individual variations in DNA methylation profiles among normal healthy populations are not thoroughly understood. To investigate the contribution of genetic variation in DNA methyltransferase (DNMT) genes to such epigenetic variation, we performed a systematic search for polymorphisms in all known human DNMT genes [*DNMT1*, *DNMT3A*, *DNMT3B*, *DNMT3L* and *DNMT2 (TRDMT1)*] in 192 healthy males and females. One hundred and eleven different polymorphisms were detected. Of these, 24 were located in coding regions and 10 resulted in an amino acid change that may affect the corresponding DNMT protein structure or function. Association analysis between all major polymorphisms (frequency > 1%) and quantitative DNA methylation profiles did not return significant results after correction for multiple testing. Polymorphisms leading to an amino acid change were further investigated for changes in global DNA methylation by differential methylation hybridization. This analysis revealed that a rare change at *DNMT3L* (R271Q) was associated with significant DNA hypomethylation. Biochemical characterization confirmed that *DNMT3L*^{R271Q} is impaired in its ability to stimulate *de novo* DNA methylation by *DNMT3A*. Methylated DNA immunoprecipitation based analysis using CpG island microarrays revealed that the hypomethylation in this sample preferentially clustered to subtelomeric genomic regions with affected loci corresponding to a subset of repetitive CpG islands with low predicted promoter potential located outside of genes.

INTRODUCTION

In mammals, DNA and chromatin modifications represent a key layer of heritable biological information superimposed onto the primary DNA sequence. Such epigenetic information plays critical roles in the way that mammalian genomes are structurally organized, functionally regulated and stably maintained. DNA methylation, in particular, is required for proper embryonic development (1,2) and for the formation of mature functional germ cells (3–5). Likewise, defects in DNA methylation are increasingly associated with a series

of human conditions (6) including imprinting disorders (7,8), infertility (9–11), autoimmune disorders (12) and most strikingly, cancer (13,14). The links between DNA methylation deregulation and human health have resulted in a renewed attention to the mechanisms by which DNA methylation profiles are established and maintained and to the factors, genetic and environmental, which might influence these profiles.

In mammalian cells, a family of DNA methyltransferase (DNMT) proteins carries the primary responsibility for the deposition of genomic DNA methylation patterns and for their perpetuation through time. The *DNMT3A* and *DNMT3B* enzymes

*To whom correspondence should be addressed. Tel: +49 22828716737; Fax: +49 22828714320; Email: osman.elmaarri@ukb.uni-bonn.de

†These authors have equally contributed to this study.

‡Present address: Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, Victoria 8006, Australia.

are *de novo* DNMTs which are responsible for establishing DNA methylation patterns during gametogenesis and early embryogenesis (15). The DNMT1 enzyme acts primarily as a maintenance DNMT ensuring the faithful replication of DNA methylation patterns at each cell division (16,17). In addition to these active enzymes, the DNMT3L protein, while catalytically inactive by itself, also contributes to *de novo* methylation by interacting with the catalytic domains of DNMT3A and DNMT3B and enhancing their enzymatic activity (18–21). DNMT3L, through its ability to bind to histone H3 and to sense the modification status of the H3 histone tail, might also contribute to guiding the *de novo* methylation machinery (22,23). The DNMT2 enzyme was considered a member of the DNMT family based on its high sequence conservation with other DNMT proteins. However, the absence of a nuclear localization signal and recent findings that DNMT2 methylates a specific tRNA species suggest that this enzyme is in fact an aspartic acid tRNA methyltransferase (24,25). This enzyme was therefore renamed as tRNA aspartic acid methyltransferase 1 (TRDMT1). Nonetheless, this gene was included in our study, because DNMT2 was previously reported to show residual DNMT activity (26,27).

In humans, variations in the DNA methylation profiles are widely observed in non-pathological samples. These include gender-related, age-related and tissue-specific variations (28–30). Even within the same age group, the same gender and the same cell types, inter-individual variations could still be observed, and these variations were influenced by environmental factors, nutritional elements and to some extent by genetic factors (31–34). A clear example in which genetic variation impacts levels of DNA methylation is provided by the C677T polymorphism in the methylenetetrahydrofolate reductase gene (*MTHFR*), where TT homozygous individuals are more likely to carry hypomethylated DNA (35,36). In this study, we investigated whether genetic polymorphisms in the *DNMT* genes themselves could influence DNA methylation profiles in a well-defined group of healthy individuals of matched age and gender. For this, we performed a systematic search for polymorphisms in all coding regions of the five known human DNMT genes. The association between detected genetic variants and possible DNA methylation changes was analyzed using our previously published data reporting quantitative levels of methylation at selected single loci (*F8*, *H19*, *PEG3*, *NESP55* and a locus at 19q13.4), as well as at repetitive DNA elements (*Alu* and *Line1*). Association with global DNA methylation at 5'-CCGG-3' sites as measured by Luminometric Methylation Assay (LUMA) was also investigated (37). This study revealed that the major polymorphisms with >1% frequency have little effect on the DNA methylation levels at the studied loci. However, differential methylation hybridization (DMH) analysis on samples carrying polymorphisms that induce an amino acid change identified two DNA samples showing significant epigenetic variations from controls. A sample with a common *DNMT2* variant (Y101H; rs11254413) was associated with hypermethylation at a small subset of loci but it is unclear whether that SNP is causative in itself. A rare *DNMT3L* variant (R271Q) was associated with significant subtelomeric hypomethylation. Moreover, biochemical analysis of the mutant DNMT3L^{R271Q} protein was consistent with the

hypomethylated phenotype in that the modified protein is partially deficient in its ability to stimulate the *de novo* methylation activity of DNMT3A.

RESULTS

A systematic search for polymorphisms in *DNMT* genes

Detailed distribution of the detected polymorphisms in the five *DNMT* genes that were analyzed are given in Supplementary Material, Table S1.

DNMT1 gene at 19p13.2. We screened all 40 exons of the *DNMT1* gene. A total number of 25 different single nucleotide polymorphisms (SNPs) and a small fraction of deletions of an 'AG' dinucleotide at intron 21 were detected (Supplementary Material, Table S1). Three of these SNPs were found to induce amino acid changes (Table 1 and Fig. 1). The first one was at codon 131 (G to A change), resulting in an arginine to histidine substitution (both are positively charged) and had a frequency of 10 out of 384 alleles (2.6%). Sequence alignments between DNMT1 proteins from eight species showed that this position is not absolutely conserved (Fig. 1). The second SNP at codon 251 (C to T change) resulted in a serine to leucine change (uncharged amino acid to non-polar aliphatic amino acid). It was detected at only one allele out of 384 (0.26%) and occurred at a non-conserved position (Fig. 1). The third SNP (A to G) was observed at codon 374 and introduced an isoleucine to valine substitution (both are non-polar aliphatic amino acid) with a frequency of 18 out of 384 (4.69%). This position is also not conserved among species (Fig. 1). All three polymorphisms were detected in a heterozygous state.

DNMT2 (TRDMT1) gene at 10p15.1. The *DNMT2* gene is constituted of 11 exons. A total number of 29 SNPs were detected in addition to three insertion/deletion (Supplementary Material, Table S1). Two SNPs at exon 4 were found to induce an amino acid change (Table 1 and Fig. 1). The first (G to T) induced a glycine to valine change (both amino acids are non-polar aliphatic) at a conserved position among six different species at codon 86. It was detected at only one allele out of 384 (0.26%). The second SNP (T to C) introduced a tyrosine (aromatic amino acid) to histidine (positively charged amino acid) change at codon 101 and was frequent (185 occurrences out of 384 or 51.82%). This position is not conserved among different species (Fig. 1).

DNMT3A gene at 2p23.3. The *DNMT3A* gene comprises 23 exons and additional alternatively spliced exons, these include the upper exon 1, a 3'-untranslated region upstream of exon 4 and exons 1B and 2B located upstream of exon 7. A total number of 11 SNPs were detected, none of which induced an amino acid change (Supplementary Material, Table S1).

DNMT3B gene at 20q11.2. Compound heterozygous or homozygous mutations in *DNMT3B* can cause the human Immunodeficiency, Centromeric region instability and Facial abnormality syndrome (ICF syndrome) (38,39). The *DNMT3B*

Table 1. List of all non-synonymous DNMT polymorphisms in the sample cohort

	Number	Rs	Position	Exon	From	To	Freq 1	Freq 2	AA change	Codon
DNMT1 NT_011295	2	New	1554275	Ex 3	C	T	0.9740	0.0260	Arg cgt>His cat	131
	3	New	1549051	Ex 6	G	A	0.9974	0.0026	Ser tcg>Leu ttg	251
	7	rs8111085	1536174	Ex 12	T	C	0.9531	0.0469	Ile att>Val gtt	374
DNMT3B AL035071	9	New	55332	Ex 11	C	T	0.9974	0.0026	Arg cgc>Cys tgc	382
	10	New	55338	Ex 11	G	A	0.9948	0.0052	Ala gct>Thr act	384
DNMT3L NT_011515	6	New	991281	Ex 6	C	T	0.9974	0.0026	Gly ggc>Ser agc	127
	15	New	983528	Ex 10	C	T	0.9974	0.0026	Arg cgg>Gln cag	271
	16	rs7354779	983508	Ex 10	T	C	0.7188	0.2813	Arg agg>Gly ggg	278
DNMT2 NT_077569	8	New	11567127	Ex 4	C	A	0.9974	0.0026	Gly ggc>Val gtc	86
	9	rs11254413	11567083	Ex 4	A	G	0.4818	0.5182	Tyr tat>His cat	101

For each gene, the Table indicates the position of each SNP and the type of base change observed with respect to the annotated contig (given as an accession number under each gene name). In some instances, the indicated change may be complementary to the cDNA coding strand. The number of the exon and codon affected by each change as well as the type of amino acid (AA) change resulting from each SNP are also shown. The frequencies of each genotype in the sample cohort are indicated.

gene is constituted of 23 exons and an alternative transcript use of exon 1P. A total of 21 different polymorphisms were detected (Supplementary Material, Table S1). Among the exonic variations, two were non-synonymous (Table 1 and Fig. 1). The first (C to T change) triggered an arginine (positively charged amino acid) to cysteine change (polar uncharged) at codon 382 and was observed at only one allele out of 384 (0.26%). The second (G to A change) was observed at codon 384 on only two alleles out of 384 and induced an alanine (non-polar aliphatic amino acid) to threonine (polar uncharged) change, which is the common amino acid at this position in mice, rats and chickens. All non-synonymous SNPs were observed as heterozygous combinations and therefore will not lead to ICF syndrome. In addition, the amino acid substitutions at codons 382 and 384 have not been reported so far in ICF syndrome patients and occur in the N-terminal non-catalytic portion of DNMT3B located between the PWWP (proline–tryptophane–tryptophane–proline) domain and the zinc finger ADD (ATR–DNMT3–DNMT3L) type domain.

DNMT3L gene at 21q22.3. A total number of 20 polymorphisms were detected upon screening all 12 exons and flanking intronic sequences of *DNMT3L* (Supplementary Material, Table S1). Three introduced an amino acid change (Table 1 and Fig. 1). The first (G to A change) caused a glycine (non-polar aliphatic) to serine (polar uncharged) change at codon 127 and was observed in only one allele out of 384 (0.26%). This variation is located in the ADD zinc finger domain and occurs at a conserved position between humans, mice, rats and cows. The second polymorphism (G to A change) caused the exchange of an arginine (positively charged) to glutamine (polar, uncharged) at codon 271. It was also observed at only one allele. Arginine 271 is conserved in multiple species including human, rat as well as cow, dog, horse and macaca, although this position features a leucine in mice (Fig. 1 and data not shown). The last *DNMT3L* SNP (A to G change) occurred at codon 278 and imposed an arginine (positively charged) to glycine (non-polar aliphatic) substitution. This position features an arginine residue in multiple species and a glutamine in rats and cows. The last two SNPs occur in the C-terminal portion of DNMT3L which interacts with

the active catalytic methyltransferase domain of DNMT3A and DNMT3B (19,20). Particularly, the R271Q variant maps to one of three helices located at the interface with DNMT3A (residues 226–234, 258–274, 293–302 in DNMT3L).

Association analysis between polymorphisms in *DNMT* genes and quantitative DNA methylation analysis

Previously, we reported on quantitative DNA methylation analysis at several loci, from the same DNA sample cohort used in this study. The analyzed loci included two repetitive sequences (*Line1* and *Alu*), three imprinted loci (*PEG3*, *H19* and *NESP55*) and two single loci (*F8* and a locus at 19q13.4 located between *PEG3* and *USP29*). No clear association between polymorphisms and DNA methylation values at these loci were observed. This was true for both allele-based and genotype-based association analysis (Supplementary Material, Tables S2 and 3). We then asked whether a given extended haplotype (Supplementary Material, Table S4) over an entire gene is associated with an increase or a decrease in DNA methylation at one of the studied CpG sites or at *HpaII* sites as studied by the LUMA. For this analysis, all polymorphisms with <1% of frequency had to be excluded. As 1% correspond to a small number (less than four observations), this procedure has no impact on the power of the analysis. However, the analysis did not reveal any significant association (Supplementary Material, Table S2). We note, however, that the lack of significant association could be explained by the fact that only few CpG sites were analyzed in only one tissue, which may not reflect the complex tissue- and developmental-specific variation of the methylome. More global studies across multiple tissues are needed to fully address this point.

Genome-wide methylation scans reveal significant DNA methylation changes in DNA samples with SNPs in the *DNMT2* and *DNMT3L* genes

Among the 111 detected polymorphisms, ten caused an amino acid change (Table 1). Genome-wide DNA methylation patterns were further investigated among this subset

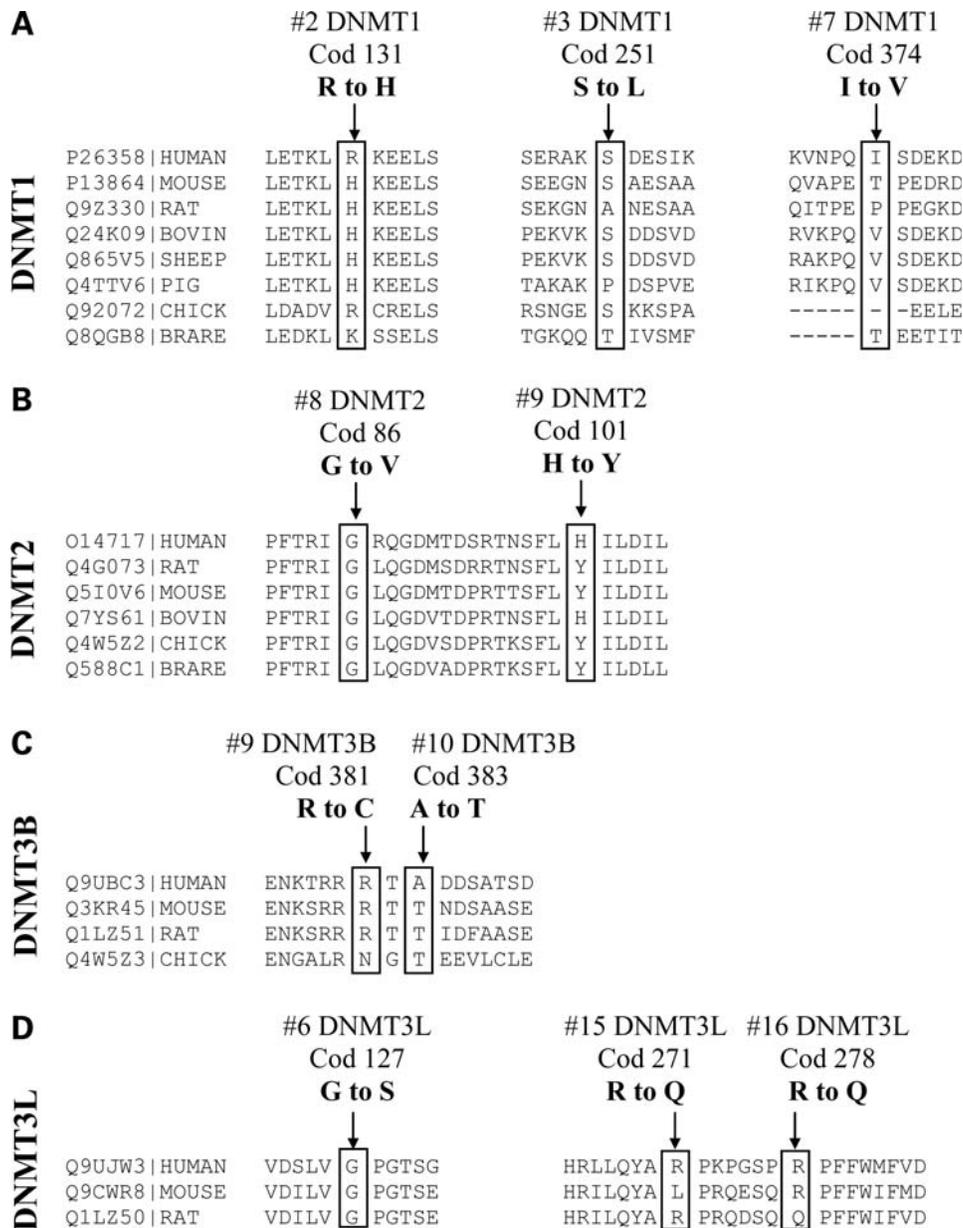


Figure 1. Identification of SNPs leading to amino acid changes in human DNMT genes. Each panel shows the alignment of amino acid sequences for a given DNMT protein from different species indicated at left along with corresponding accession numbers (SwissProt or NCBI). Regions surrounding the positions corresponding to detected amino acid changes are shown. (A) DNMT1, (B) DNMT2, (C) DNMT3B and (D) DNMT3L.

using DMH. Two samples showed significant variations compared with control samples (Table 2; DMH data are available upon request). Sample 109, corresponding to the *DNMT2* Y101H SNP, was characterized by hypermethylation at a small number of loci (0.7% of studied regions). Four regions (A5:2, A5:8, G24:15 and A4:40) (Supplementary Material, Fig. S1) predicted by DMH to be hypermethylated in sample 109 were selected, and their methylation status was determined using bisulfite methylation sequencing. This confirmed that two out of four regions (clone numbers: G24:15 and A5:8) show a 3- to 20-fold increase in DNA methylation in sample 109 relative to a normal control (Supplementary Material, Fig. S1). In contrast, sample 156, corresponding to the rare

DNMT3L R271Q SNP, showed hypomethylation at a considerable number of loci (6.1% of studied regions) in addition to hypermethylation at a few loci (0.9% of studied regions) (Table 2).

The *DNMT3L*^{R271Q} variant is associated with DNA hypomethylation preferentially at subtelomeric regions

In order to confirm and expand on the DMH findings, both 109 and 156 DNA samples were re-analyzed using a methylated DNA immunoprecipitation (MeDIP)-based microarray approach (40). The platform used for the analysis was a NimbleGen human promoter plus CpG island array carrying

Table 2. A statistical summary of the number of regions that were classified by DMH either as hypermethylated or hypomethylated (thresholds were set empirically)

Sample	Gene	SNP #	Gender	Cod #	Genotype	No. of Normal	No. of Hyper	No. of Hypo	% Normal	% Hyper	% Hypo	% 5mC
145	<i>DNMT1</i>	2	M	131	G/A	14204	0	0	100	0	0	3.33
200	<i>DNMT1</i>	3	F	251	C/T	14210	1	32	99.8	<0.1	0.2	3.32
35	<i>DNMT1</i>	7	M	374	C/T	14233	0	4	100	0	<0.1	3.58
37	<i>DNMT2</i>	8	M	86	C/A	14290	0	0	100	0	0	3.41
61	<i>DNMT2</i>	9	M	101	A/G	14216	0	3	100	0	<0.1	4.06
102	<i>DNMT2</i>	9	M	101	A/G	14257	0	4	100	0	<0.1	4.21
109	<i>DNMT2</i>	9	M	101	A/G	14158	105	8	99.2	0.7	0.1	3.37
115	<i>DNMT3B</i>	9	M	381	C/T	14236	0	0	100	0	0	3.31
302	<i>DNMT3B</i>	10	M	383	C/T	14314	0	0	100	0	0	3.27
84	<i>DNMT3L</i>	6	M	127	C/T	14282	0	1	100	0	<0.1	3.47
156	<i>DNMT3L</i>	15	F	271	C/T	13166	133	864	93.0	0.9	6.1	3.32
73	<i>DNMT3L</i>	16	M	278	T/T	14246	0	0	100	0	0	3.57
195	<i>DNMT3L</i>	16	M	278	C/C	14198	0	0	100	0	0	3.17
197	<i>DNMT3L</i>	16	M	278	T/C	14272	3	15	99.9	<0.1	0.1	3.80

The percentage average values of three measurements of total methyl cytosine content are given in the last column (M, Male; F, Female; 5mC, 5 methyl cytosine; Normal, no significant differences between test sample and control sample; Hyper, hypermethylated; Hypo, hypomethylated; Cod, codon). The two lines with bold entries correspond to the samples that shows considerable variations.

28 226 CpG islands tiled over 1 kb. MeDIP results (MeDIP data are available upon request) were found to be in good agreement with our previously published quantitative measurements of DNA methylation levels at 13 CpG-rich regions at 19q13.4 (41) (Supplementary Material, Fig. S2). In addition, we verified that methylated loci were distributed as expected along the X and Y chromosomes, taking advantage of the fact that the individuals 109 and 156 are of opposite gender (data not shown). The MeDIP analysis confirmed that many regions in sample 156 were hypomethylated in comparison with sample 109. Figure 2A shows a scatter plot of methylation scores (\log_2 ratios) for both samples across ~26 000 autosomal CpG islands. As is clearly evident, the methylation scores for the vast majority of loci tested are in good agreement between the two samples (78% of loci are within 1 standard deviation (SD) of each other with an overall correlation coefficient of 0.83). As expected, loci that are unchanged and remain unmethylated in both samples correspond to strong CpG islands located primarily at gene promoters. In contrast, loci that are unchanged and remain methylated in both samples correspond to weak, repetitive CpG islands located outside of promoters regions (data not shown). In addition to these mostly unchanged loci, a significant portion of loci showed a trend towards hypomethylation in sample 156. Hence, 1056 loci (3.9% of total) show hypomethylation above 2 SDs, whereas 381 loci (1.4% of total) show hypomethylation above 3 SDs, which is considered highly significant. In contrast, <0.1% of loci showed a gain of methylation above 3 SDs in sample 156 showing that the overall trend was largely biased towards loss of DNA methylation. In order to further validate this data, we performed an additional and independent MeDIP-chip analysis on a control sample corresponding to a healthy male individual (that carry no polymorphism in the analyzed DNMT genes) and replotted the dataset from samples 156 and 109 against this control. In the case of sample 156 (*DNMT3L* R271Q), 97% of the loci initially identified as significantly hypomethylated remained hypomethylated above 1 SD (83% above 2 SDs; Fig. 2B) when compared with the control

MeDIP dataset. This strongly suggests that the majority of the hypomethylated loci were correctly identified by our initial comparison. Replotting of MeDIP data from sample 109 (*DNMT2* Y101H) against the control dataset showed that the majority of loci originally identified as hypomethylated in sample 156 had high methylation scores both in sample 109 and in the control dataset, which is as expected (Fig. 2C). A slight trend towards hypermethylation in sample 109 was also detected at a subset of these loci, which is consistent with our DMH observations. Altogether, these two additional replots validate the identification of hypomethylated loci in the genome of the individual carrying the *DNMT3L* R271Q SNP. The MeDIP dataset was also replotted against the independently derived dataset of Rauch *et al.* (42), who recently reported the complete methylome of a normal human B-cell with a similar conclusion (Supplementary Material, Fig. S3). Finally, we performed direct bisulfite methylation sequencing on a selection of the six hypomethylated loci and confirmed that in three regions sample 156 is associated with a reduction in DNA methylation compared with sample 109 and most controls (Supplementary Material, Fig. S4); we note here that the repetitive nature of the affected regions (see below) make them technically difficult to analyze by bisulfite sequencing.

We then analyzed the genomic distribution of hypomethylated loci and observed that they were not uniformly distributed but rather, were significantly enriched at subtelomeric regions. Figure 3A shows that although the subtelomeric regions (arbitrarily defined as the last 5% of each chromosome) carry 29% of CpG islands on the array, these regions encompassed 50.3% of hypomethylated loci (including all loci above 2 SDs—this value goes to 53% for the loci above 3 SDs). The distribution of all hypomethylated loci above 2 SDs along a composite model human chromosome has shown in Figure 3B shows a clear, statistically significant, bias towards telomeric regions. This trend held true for all autosomes (data not shown). In contrast, the few loci associated with a gain in DNA methylation were equally and randomly distributed along the chromosomal arms (data not shown).

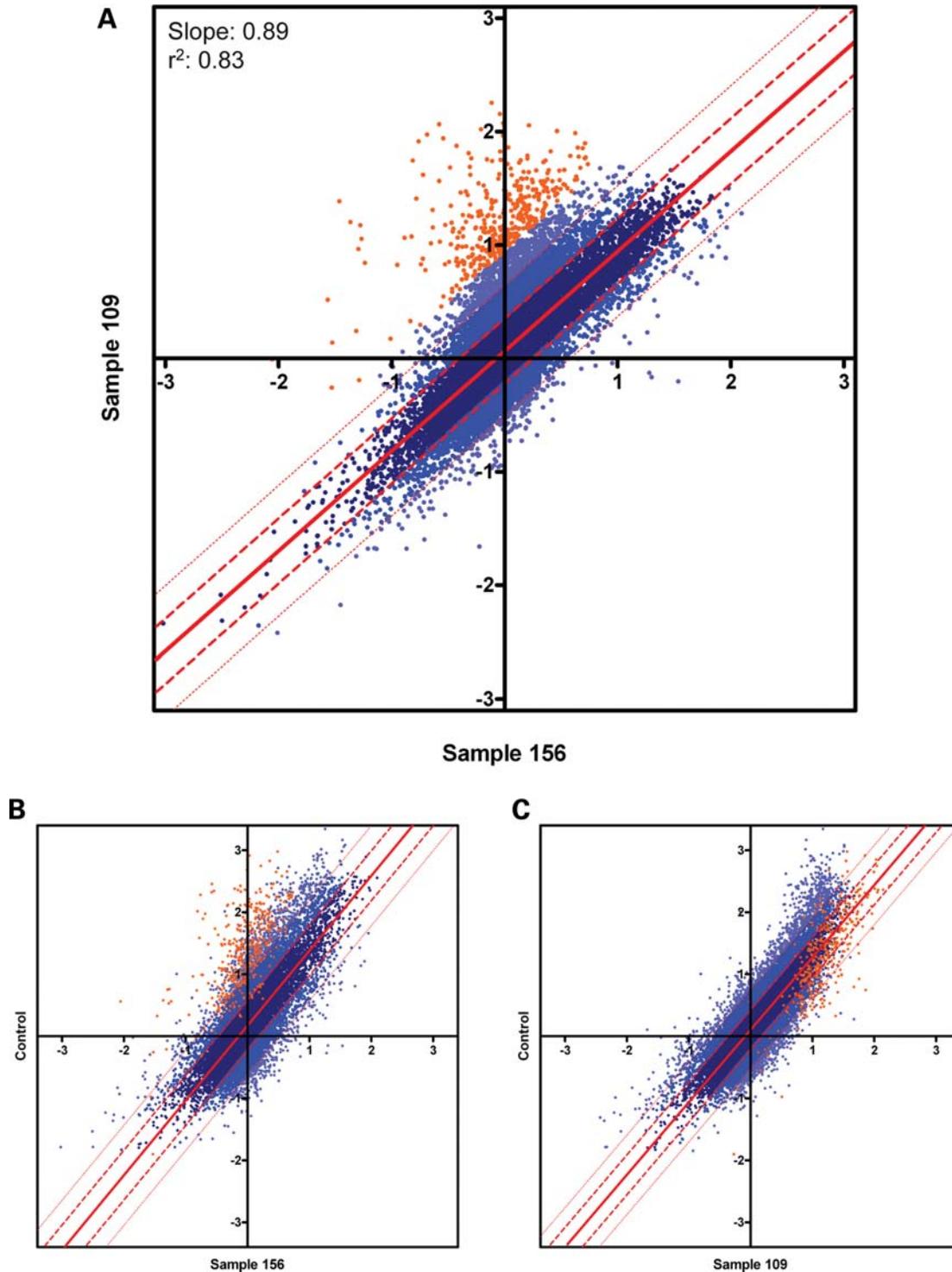


Figure 2. MeDIP analysis reveals significant DNA hypomethylation in sample 156. (A) The average \log_2 scores for sample 156 [*DNMT3L*(R271Q), x-axis] and sample 109 [*DNMT2*(Y101H), y-axis] are plotted against each other at ~26 000 autosomal CpG islands on the NimbleGen CpG island plus promoter array. Data points are shaded differently depending on whether they fall within one standard deviation (SD) (dark blue); within one and 2 SDs (medium blue); beyond 2 SDs (light blue) and beyond 3 SDs (orange) of the overall regression analysis (indicated by a solid straight line). The slope and correlation coefficient for the regression analysis are given in the upper left corner. (B and C) Replots of the sample 156 dataset (B) and of the sample 109 dataset (C) against an independently determined dataset derived from a healthy male control. Symbols are as above. Orange datapoints correspond to the previously identified hypomethylated loci.

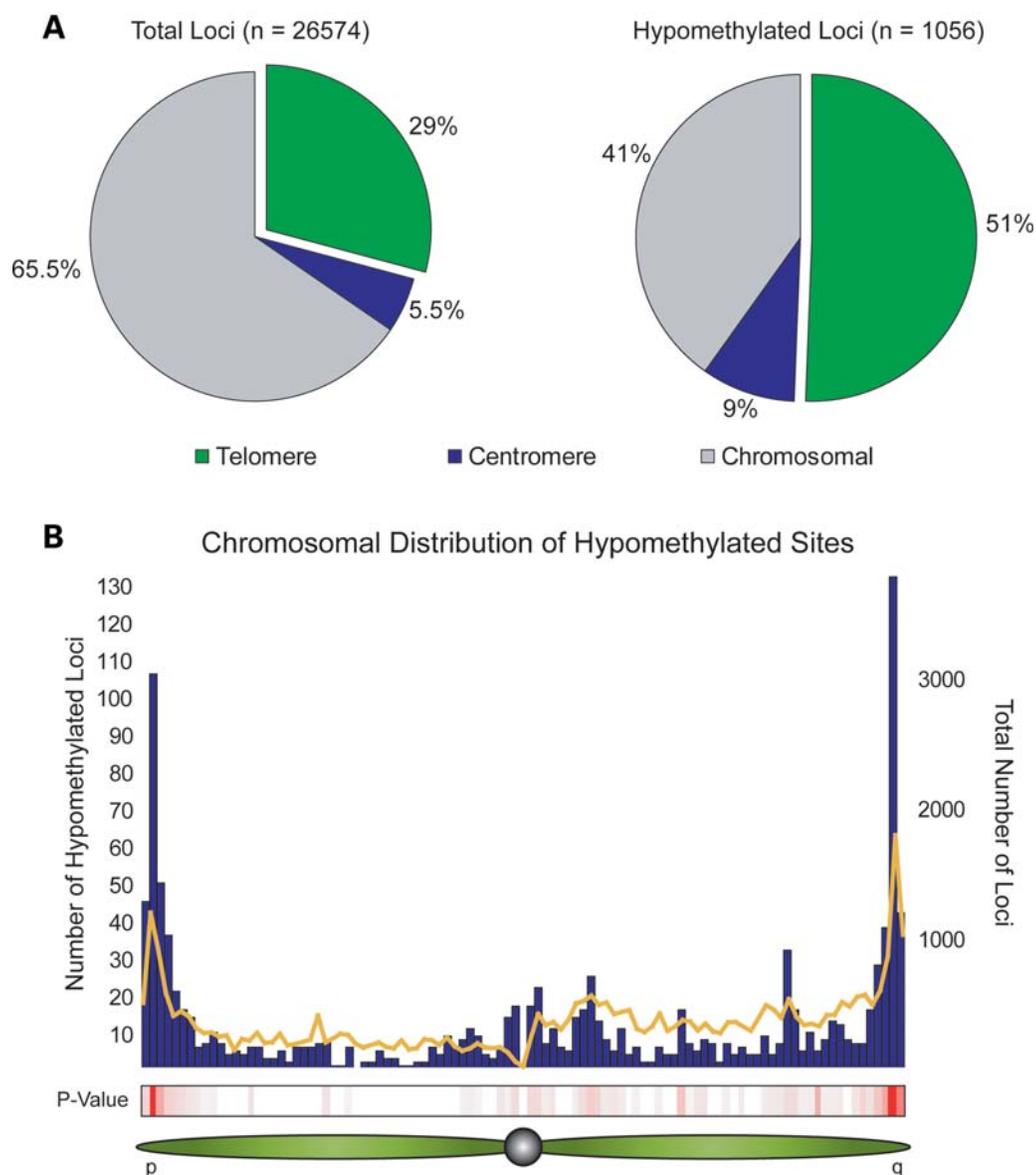


Figure 3. Hypomethylated loci are preferentially distributed at subtelomeric regions. (A) Pie charts representing the distribution of loci at telomeres (defined as the 5% of DNA sequences located at the tip of each chromosome arm), centromeres (defined as the 10% of DNA sequences flanking the p and q transition) or chromosome bodies. The chart at left shows all probes on the array; the chart at right shows loci hypomethylated above 2 SD. (B) The number of hypomethylated loci (blue bars, left axis) and total loci on the array (gold line, right axis) were determined for each autosome along a collection of 100 sub-regions each corresponding to 1% of each chromosome's length. These numbers were then combined across autosomes and shown along a composite model human chromosome displayed at bottom. *P*-values representing whether hypomethylated loci were enriched over the total number of loci in each chromosomal sub-region (corresponding to 1% of the total sequence length) were determined by a *t*-test and are displayed along the chromosome as a 'heat map' in which the shade of red color is indicative of the statistical significance (the two largest peaks corresponding to hypomethylated loci have a *P*-value < 10^{-52}).

Hypomethylated loci define a subset of repetitive CpG islands with low promoter potential located predominantly in intergenic regions

Further analysis of the distribution of hypomethylated loci in sample 156 revealed that very few of these loci (~2.8%, Fig. 4A) were present at promoter regions. In contrast, a control set of 200 CpG islands selected by virtue of the fact that the difference in their methylation scores between samples 109 and 156 was closest to zero are for the most part (40%) associated with promoter regions, as expected.

In addition, hypomethylated loci were found predominantly in intergenic regions (51%) and within introns of coding regions (20%). Finally, detailed analysis of the sequence content of the hypomethylated CpG loci revealed that these regions tend to have on average a lower CpG index score (43) when compared with the set of 200 control CpG islands or, even more markedly, with a set of ~10 000 highly specific 'bona fide' CpG islands associated with promoter regions (Fig. 4B). Likewise the hypomethylated loci were 2–3.5 times more likely to contain simple tandem DNA repeats than the control set of 200 unchanged CpG islands or the

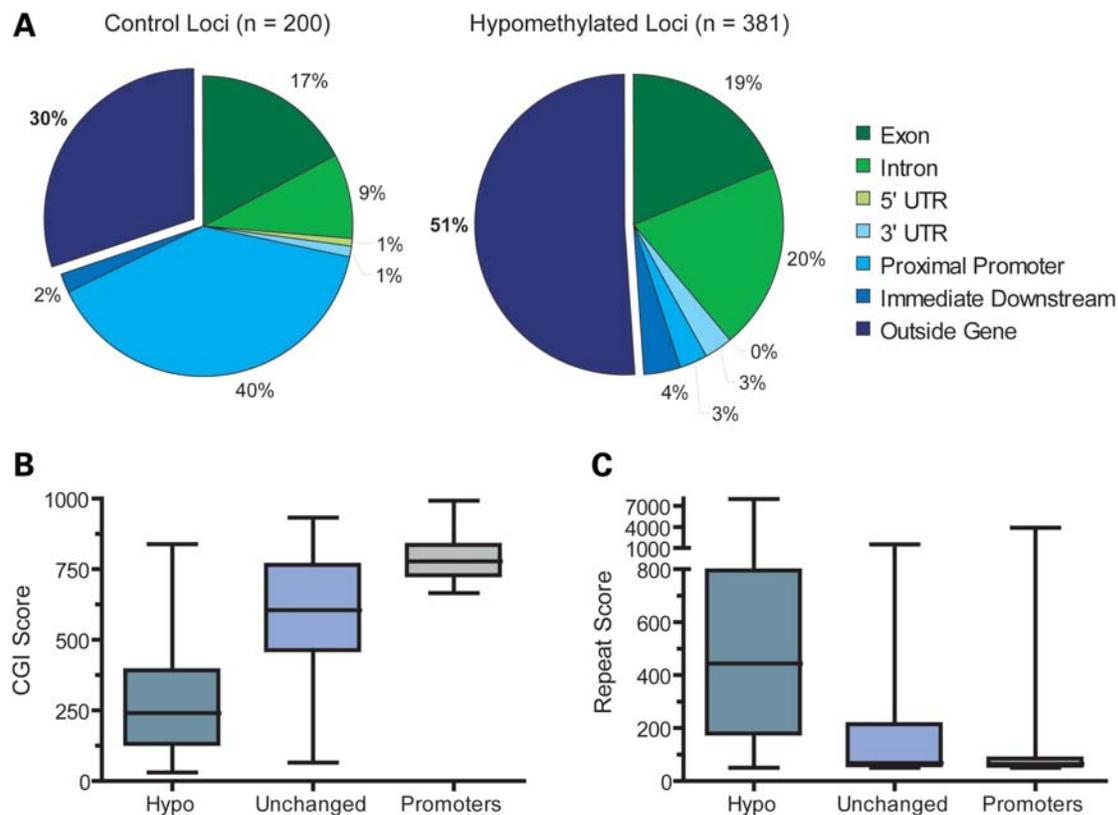


Figure 4. Hypomethylated loci are not associated with promoters. (A) Pie charts representing the distribution of 200 unchanged control CpG islands (left) and of the 381 hypomethylated loci (>3 SD) (right) among various gene regions, as indicated. The analysis was performed using the *cis*-regulatory element annotation system (CEAS) website (61). (B) Whisker plot showing the CpG island (CGI) scores of the top 381 hypomethylated loci (left), 200 control loci (middle) and ~10 000 highly specific CpG island promoters. CGI scores were extracted from the UCSC Genome Browser and as were defined by Bock *et al.* (43). (C) Whisker plot showing the repeat content scores of the top 381 hypomethylated loci (left), 200 control loci (middle) and ~10 000 highly specific CpG island promoters. Repeat scores were extracted from the UCSC Genome Browser and were compiled from the Tandem Repeat Finder program (44).

highly specific CpG islands associated with promoters (63% of hypomethylated loci contained repeats compared with 33 and 18% for specific CpG islands and control loci, respectively). Furthermore, the repeats were much more significant in both length and count [as measured by their tandem repeat score (44)] in the hypomethylated sample than in the other two samples (Fig. 4C). This indicates that the loci found to be hypomethylated in association with the DNMT3L R271Q SNP correspond to a specific subset of CpG islands that (i) are enriched at subtelomeric regions, (ii) are predominantly located outside of genes or to a lesser extent, within introns, (iii) have a low potential to function as promoters and (iv) are repetitive in nature.

The DNMT3L^{R271Q} protein is impaired in its ability to interact with DNMT3A and stimulate *de novo* methylation

In order to determine whether the changes in DNA methylation patterns observed in sample 156 could be directly attributed to a change in the normal activity of DNMT3L, we investigated the ability of the purified DNMT3L^{R271Q} protein to stimulate *de novo* methylation by the full-length DNMT3A2 protein *in vitro*. For this, the purified DNMT3L^{R271Q} protein was pre-incubated with DNMT3A2 in a 1:1 molar stoichiometric ratio for 60 min at 37°C in the

presence of *S*-adenosyl-L- $\{methyl-^3H\}$ methionine (3H -SAM) before the reaction was initiated by the addition of DNA. DNA methylation was measured by the incorporation of tritiated methyl groups into DNA as described previously (21). The wild-type DNMT3L protein led to a robust stimulation of *de novo* methylation (Fig. 5A) characterized by a ~6-fold increase of the initial rate of the reaction, in agreement with prior data (21). In contrast, the DNMT3L^{R271Q} protein only led to a ~3.5-fold increase of the initial rate. Similarly, the extent of the reaction after 2 h was also lower for the mutant protein than for the wild-type control (Fig. 5A). This suggested that the DNMT3L^{R271Q} protein is impaired in its ability to stimulate *de novo* methylation.

Crystallographic data indicates that arginine 271 is localized at the interaction surface between DNMT3L and DNMT3A (Fig. 5B) (20). This suggests that the Arg 271 to Gln change might result in a defect in the ability of the mutated protein to interact with the DNMT3A catalytic domain. One measure of the ability of DNMT3L to interact with DNMT3A is reflected in the stoichiometry of the interaction. The wild-type DNMT3L protein interacts in a 1:1 molar ratio with DNMT3A2 at which point maximal stimulation of DNA methylation is reached (21). We therefore expected that a mutant protein with a reduced interaction might require more protein to reach maximal saturation. To test

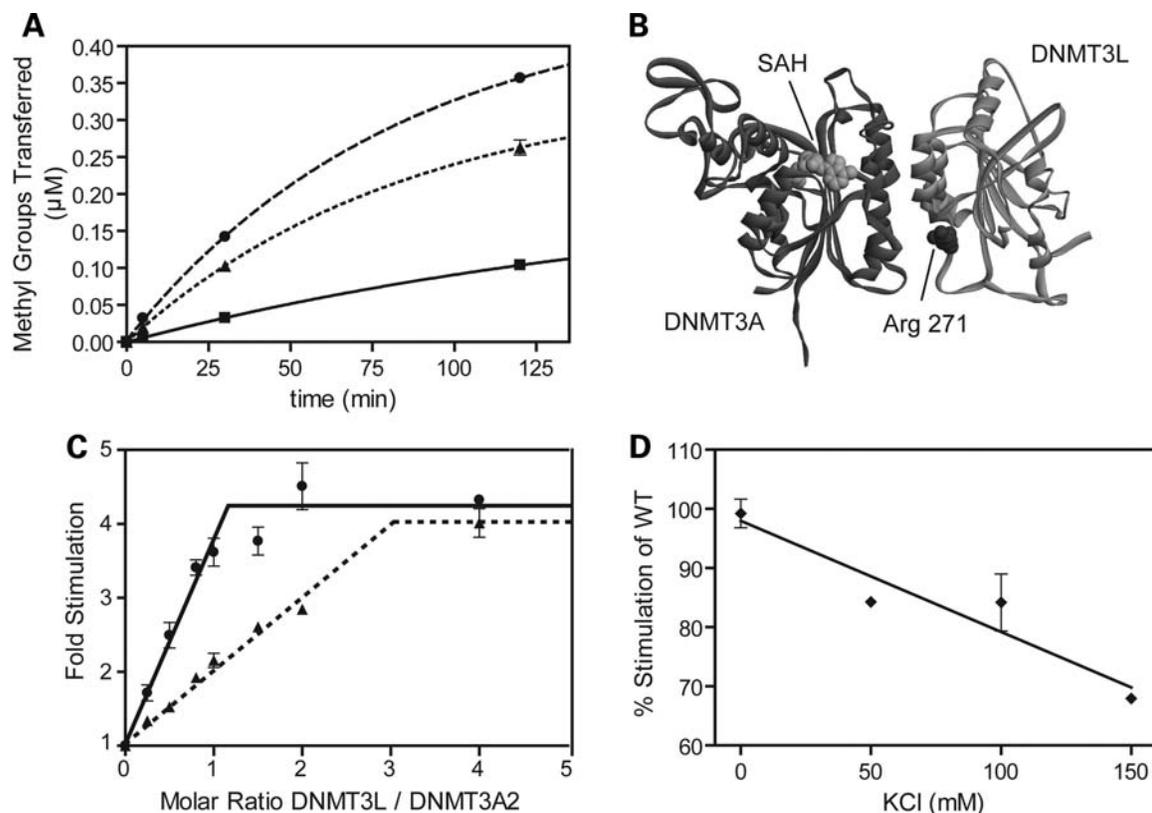


Figure 5. The DNMT3L^{R271Q} variant is impaired in its stimulation of DNMT3A. (A) DNA methylation activity is plotted as the molar unit of methyl groups transferred into DNA against reaction time. Reactions were performed with DNMT3A2 alone (■, solid line), DNMT3A2:DNMT3L (●, dashed line) and DNMT3A2:DNMT3L^{R271Q} (▲, dotted line). (B) Crystallographic structure of the C-terminal domains of DNMT3A (dark gray) and DNMT3L (light gray) [adapted from Jia *et al.* (20)]. The location of Arginine 271 of DNMT3L is highlighted. S-adenosyl-L-homocysteine (SAH) bound to the DNMT3A active site is indicated. (C) The fold stimulation of DNA methylation mediated by the wild-type DNMT3L (●, solid line) or the mutant DNMT3L^{R271Q} (▲, dashed line) is shown as a function of the molar ratio of DNMT3L to DNMT3A2 (DNMT3L was titrated against a constant concentration of DNMT3A2). (D) The activity of the DNMT3L^{R271Q} mutant was measured under increasing KCl concentrations and is reported as a fraction of the activity of the wild-type DNMT3L measured in the same condition. All reactions in (A, C and D) were performed in duplicate and are shown as averages with error bars.

this, we titrated DNMT3L against a constant amount of DNMT3A2 and measured the resulting incorporation of methyl groups 30 min after initiation of the reaction. As expected, maximal stimulation was reached for the wild-type protein around a 1:1 molar ratio (Fig. 5C). Interestingly, the DNMT3L^{R271Q} protein was able to stimulate *de novo* methylation to the same extent as the wild-type protein at saturation but reaching maximal stimulation required a 3:1 molar ratio of DNMT3L^{R271Q} to DNMT3A. This is consistent with the notion that DNMT3L^{R271Q}, while fully capable of stimulating DNMT3A2 when added in excess, is defective in its ability to form proper complexes with DNMT3A. In order to further determine whether the R271Q mutation had weakened the interaction between DNMT3L and DNMT3A, we measured the ability of both wild-type and mutant proteins to stimulate DNA methylation when challenged by increasing concentrations of monovalent salt. Salt is indeed expected to disrupt any electrostatic interaction involved in the formation of DNMT3A:DNMT3L complexes. As expected, overall methyl group incorporation by DNMT3A on its own and by DNMT3A:DNMT3L complexes dropped with increasing KCl. However, when the DNA methylation output mediated by the DNMT3L^{R271Q} protein was compared in each condition

to the output associated with the wild-type protein in the same condition, it was clear that the mutant protein showed increased sensitivity to the salt challenge (Fig. 5D). This further validates the notion that the DNMT3L^{R271Q} protein is affected in its ability to physically interact with DNMT3A, in agreement with the localization of the Arg 271 at the interface between these two proteins.

DISCUSSION

Our study is the first one that combines the systematic search for polymorphisms in the coding region of the *DNMT* genes and the measurement of DNA methylation levels in a 'healthy' population. The advantage of our study is that we actively screened, among healthy individuals, for variations in all coding regions of the examined genes and thus we were able to identify common known polymorphisms as well as the rare unknown ones.

The large number of 111 different detected polymorphisms were used to generate haplotypes and their association with the DNA methylation levels was studied. Detailed analysis based on alleles, genotypes or haplotypes did not reveal

highly significant associations with the measured DNA methylation values (Supplementary Material, Tables S2–S4). We note here that non-synonymous polymorphisms were not observed in the catalytic domains of DNMT3A/B or DNMT1 probably because such variations would have strong deleterious effects that would lead to abnormal development and thus cause early spontaneous abortions; therefore such cases will go undetected. We then focused our analysis on the 10 non-synonymous polymorphisms for two reasons: (1) these polymorphisms are more likely to have an effect on the corresponding protein activities or interactions and (2) since most of these polymorphisms are rare, their impact on DNA methylation levels or patterns cannot be estimated using statistical analysis. Using the DMH technique, we therefore studied global DNA methylation patterns associated with these non-synonymous polymorphisms, and showed that while most SNPs were undistinguishable from controls, two samples gave rise to global changes in DNA methylation profiles that clearly deviated from controls. Sample 109 carries a common tyrosine to histidine change at codon 101 of the *DNMT2* gene (*TRDMT1*) that was found in 51% of the alleles in the studied samples (Table 1). For this sample, a small but significant portion (0.74%, corresponding to 105 loci out of 14 271 tested) of the genomic loci under study showed hypermethylation (Table 2). This hypermethylation was confirmed independently by bisulfite methylation sequencing at four regions, two out of four showing increase in DNA methylation (Supplementary Material, Fig. S1). However, based on the fact that two other samples carrying the same polymorphism were essentially normal (samples 61 and 102, Table 2) and because this gene is now widely recognized as a tRNA methyltransferase rather than a DNMT, we conclude that this SNP in itself is not causative for the methylation variations that we observed. However, larger population studies combined with biochemical analysis of the variants are needed to provide conclusive answers. Likewise, environmental or nutritional effects, as observed earlier in animal models (32,33) or even other genetic factors, can not be ruled out.

Of all the variants uncovered in our search, the rare heterozygous DNMT3L^{R271Q} variant observed in sample 156 was associated with a considerable reduction in genomic methylation levels as measured by DMH and by MeDIP-based microarray analysis (Table 2 and Fig. 2). The R271Q polymorphism is located immediately at the end of an α -helix located at the interface between the DNMT3L and the DNMT3A proteins (20), suggesting that the substitution might affect the ability of DNMT3L to physically bind to DNMT3A and to functionally stimulate its activity. Our biochemical analysis is entirely consistent with this hypothesis (Fig. 5). We indeed report that the mutated protein is less efficient than the wild-type DNMT3L at stimulating DNMT3A when both proteins are present at equimolar concentrations. Furthermore, we show that although the mutated protein is capable of stimulating DNMT3A to levels similar as those observed for the wild-type DNMT3L, achieving this stimulation requires a 3-fold molar excess of the mutant protein. Finally, we show that the complexes formed by the mutated protein are more sensitive to increasing ionic conditions, consistent with the notion that

the interaction between the mutated protein and DNMT3A is weaker.

From this, we conclude that the DNMT3L^{R271Q} allele should lead to reduced DNA methylation activity which fits well with our genomic analysis. One complicating factor in translating our biochemical analysis to real life situation is that the DNMT3L^{R271Q} allele is found in heterozygous combination with the wild-type allele. One might expect that the wild-type allele should be sufficient to provide full DNMT3L function, as observed in *Dnmt3L*[±] mice (45,46). This expectation depends in part on the assumption that the expression of DNMT3L from one allele is sufficient to provide full stimulation of DNMT3A which requires at least equimolar amounts of each protein (20,21). Although this assumption appears correct in mouse, it is clear that the expression profile of *Dnmt3L* is different for human (47), thus raising the possibility that methylation levels in human might be more sensitive to dosage effects. It is also possible that the presence of the mutated variant is not neutral owing to the fact that the mutated protein can compete with the wild-type protein and promotes the formation of mixed complexes with altered catalytic characteristics. Altogether, the structural and biochemical data predict that the DNMT3L^{R271Q} polymorphism should lead to a reduced DNA methylation output.

Assigning a direct causal relationship between the mutated allele and the observed genomic hypomethylation is complicated by the fact that we know little of other genetic or environmental factors that could also impact global DNA methylation patterns in this individual. Such determination would require further analysis of additional samples carrying this rare polymorphism. In addition, our current study design did not allow us to go back to the family of the carrier in order to follow the segregation of the mutated allele through the family pedigree. We therefore could not trace the parental origin of the allele. Likewise, the possibility that such a base change could have originated *de novo* during gametogenesis in either parent or early during development can be neither excluded nor confirmed. We note, however, that of all the SNPs that induced an amino acid change in this cohort, the DNMT3L^{R271Q} allele is the only one associated with such pronounced changes in genomic methylation levels. Therefore, although we cannot formally rule out the involvement of extrinsic factors, like the influence of environmental or additional genetic factors, we believe that on the basis of our data, it is reasonable to propose that the genomic hypomethylation observed in individual 156 results from the DNMT3L R271Q SNP. Together with ICF syndrome, this would represent the second example of an association between a mutation/polymorphism in a gene from the *DNMT3* family and genomic hypomethylation in human.

Interestingly, the hypomethylation defect associated with the DNMT3L^{R271Q} allele was only manifest at 1–4% (depending on statistical significance thresholds) of all CpG islands analyzed by MeDIP. The mutated allele therefore did not cause a genome-wide loss of DNA methylation. Rather, our analysis shows that the hypomethylated loci correspond to a subset of CpG islands that show a strong telomeric bias in their distribution (Fig. 3). In that context, it is interesting to

note that telomeres are highly methylated regions under epigenetic control (48,49). Furthermore, recent data clearly implicated the DNMT3A and DNMT3B enzymes in telomeric epigenetic maintenance, because a combined deficiency in these genes led to telomeric hypomethylation in mouse ES cells (50). In humans, cells from ICF syndrome patients carrying mutations in *DNMT3B* also show subtelomeric hypomethylation and present with shortened telomere lengths (51). Interestingly, the DNMT3L^{R271Q} carrier also presented with shorter telomeres than age-matched controls (Supplementary Material, Fig. S5), although the measured length still remained within the normal range of variation observed in human populations (Dr Jue Lin, personal communication). Altogether, these data suggest that DNMT3L might be part of a DNMT complex that operates at telomeres. This notion is consistent with recent data showing that DNMT3L plays a much broader role in establishing genomic DNA methylation patterns than initially thought. For instance, *Dnmt3L*-deficiency led to hypomethylation at almost all types of repeated DNA elements in mouse male prospermatogonia (52). Furthermore, a recent study showed that hundreds of non-repetitive loci also fail to acquire DNA methylation in testis and to a lesser extent in somatic tissues of *Dnmt3L*-deficient mice (45,46). Interestingly, these loci corresponded in their majority to low CpG content sequences located away from 5' regions of genes. Likewise, two additional studies have shown that inter-individual variation in DNA methylation levels is often associated with repetitive intra- or inter-genic regions (53,54). These studies are broadly consistent with our observations that the hypomethylated loci we observed correspond to weak CpG islands characterized by a lower CpG content, a higher density in simple tandem repeats, and by the fact that they map away from known promoters or 5' regions. This, together with the preferential localization at telomeres, might suggest that DNMT3L in human is involved in setting a particular chromosomal organization at chromosome ends.

Altogether, our study reports the first systematic search for genetic polymorphisms in DNMT genes in a healthy human cohort. A total of 111 SNPs, including 10 leading to an amino acid change in the corresponding protein, were described, many of which were novel. Of all SNPs, the DNMT3L^{R271Q} variant, which leads to the production of a mutated protein with a reduced ability to stimulate DNA methylation, was associated with significant levels of DNA hypomethylation. The hypomethylated loci clustered at telomeric regions located away from genes, suggesting that DNMT3L might be involved in maintaining the particular epigenetic identity of telomeres, an important element in telomere length regulation (49). We note that the anonymous DNMT3L^{R271Q} carrier was asymptomatic upon entry into our study. This does not preclude the possibility that this individual might be predisposed to diseases in future life, in particular since our cohort was very young (average age 23 years). Possible fertility defects, cancer susceptibility or perhaps premature aging might be expected to occur in this individual based on our knowledge of DNMT3L function and the data presented here (access to the carrier individual for follow-up tests was not possible within the context of this study). It will be desirable in the future to screen for

mutations in the *DNMT3L* gene in other human cohorts predisposed to particular conditions in order to fully investigate the function of this gene on human health.

MATERIALS AND METHODS

DNA from blood donors

Blood samples collected from 96 healthy women and 96 healthy men were obtained from blood donors at Institute of Experimental Hematology and Transfusion Medicine, Bonn, Germany. The individuals were matched for their age with averages for males and females of 24.8 ± 3.4 and 24.3 ± 3.5 , respectively. The Ethics Committee of the medical faculty at University of Bonn has approved the use of this material for research purposes (Ethics Committee approval no. 106/05). The samples were previously used in the study by El-Maarri *et al.* (28).

Screening for polymorphisms by denaturing high performance liquid chromatography

The targeted sequence for polymorphism screening on genomic DNA corresponded to the mRNA of the five DNMT genes (*DNMT1*: NM_001379.1; *DNMT2* (TRDMT1): NM_004412.3; *DNMT3A*: AF067972.2; *DNMT3B*: AF331857.1; *DNMT3L*: AF194032.1). The amplicons were generated by standard PCR; primer sequences and annealing temperatures are listed in Supplementary Material, Table S5. PCR products were heated at 95°C for 10 min followed by incubation at 55°C for 10 min to allow the formation of heteroduplexes. The presence of homoduplexes and heteroduplexes indicate that the two alleles differ by the presence of polymorphism(s) at one of the alleles. Analysis of the heteroduplex and homoduplex mixture was performed by denaturing high performance liquid chromatography (dHPLC) on the WAVETM DNA Fragment Analysis System (Transgenomics, San Jose, USA). The presence of more than one peak in a given chromatogram indicates the existence of differences (polymorphisms) between the two alleles. Representatives, corresponding to all abnormal patterns, were sequenced from a newly generated PCR product and the polymorphisms were determined. Homozygous samples for allele 1 and allele 2 for each polymorphism were determined. To determine the genotype of the homozygous samples, equal volumes of PCR products were mixed with the same volume of a homozygous control sample of known genotype; heteroduplex formation was performed and the products were further analyzed by dHPLC. The presence of one peak in the chromatogram indicates that both products share the same genotype, whereas the presence of more than one peak indicates that the sample has an opposite genotype as the control sample.

PCR fragments with ambiguous peak pattern or which contain more than one polymorphism were directly sequenced. This included exons 3, 13, 17, 20 and 21 in *DNMT1*; exons 7 and 10 in *DNMT3L*; exons 3, 4 and 8 in *DNMT2* (*TRDMT1*) and the polymorphism at nucleotide 1298 in *MTHFR* corresponding to rs1801131. Furthermore, polymorphisms in *DNMT3B* exons 23-1, 23-3 and 23-4 were determined by single nucleotide primer extension and dHPLC (55).

Statistical analysis

Prior to statistical analysis, three SNPs (number 10-DNMT1, 4-DNMT3B and 11-DNMT3L in Supplementary Material, Table S1) had to be removed because of strong deviations from Hardy–Weinberg equilibrium. For association analysis, we performed haplotype trend regression on a quantitative phenotype as suggested (56). We used the respective implementation in FAMHAP (57). The method was applied to each single SNP, as well. Genotypes were analyzed, coding heterozygotes as homozygote for a dummy allele 3. As this coding does not lead to correct asymptotic *P*-values, the respective *P*-values were computed with Monte-Carlo simulations. We also computed a corrected *P*-value for the single-marker analysis of each gene and measurement. For this purpose, linkage disequilibrium between SNPs from one gene was accounted for via Monte-Carlo simulations (57).

Haplotypes extending over the entire gene regions were analyzed, as no block substructures could be identified based on the haplotype frequency estimates obtained with FAMHAP (58). SNPs with a minimum allele frequency < 1% were excluded for haplotype analysis. We performed allele-based single-marker analysis and genotype-based single-marker analysis, both corrected for the number of SNPs tested within each gene, and haplotype analysis. These three tests, which already involve correction for the number of SNPs analyzed within each gene region, were done for five genes and 17 measurements. The corresponding Bonferroni *P*-value at $\alpha = 0.05$ is 0.00019.

Differential methylation hybridization

The DMH protocol was performed as described previously (59,60) with slight modifications. To generate targets, 2 μ g of genomic DNA from 14 test samples and an appropriate pool of control samples were digested with *Mse*I (New England Biolabs) and purified. After ligation of linker H12/H24 (H24: 5'-AGGCAACTGTGCTATCCGAGGGAT-3' and H12: 5'-TAATCCCTCGGA-3') using T4 DNA ligase (New England Biolabs), the fragments were subsequently digested with two methylation sensitive restriction enzymes [*Hpa*II (5'-CCGG-3') and *Bst*UI (5'-CGCG-3')]. This was followed by 20 cycles of amplification with the H24 linker primer. Amplified products were indirectly labeled with the Alexa Fluor 647 (test samples) and the Alexa Fluor 555 dyes (pool of control samples) using the BioPrime Plus Array CGH Indirect Genomic Labeling Kit (Invitrogen). Equal amounts of the labeled amplicons of the test samples, pool of control samples and 10 μ g human Cot-1 DNA (Invitrogen) were co-hybridized to a microarray slide spotted with about 15 000 CpG-rich DNA fragments (each double spotted). Data from single-copy sequences were normalized and loci with a Alexa Fluor 647/Alexa Fluor 555 ratio greater or equal to 2.5 were scored as hypermethylated, whereas a ratio lesser or equal to 0.5 were scored as hypomethylated.

MeDIP and microarrays analysis

The MeDIP assay was done essentially as described (40) and according to Roche-NimbleGen (Madison, USA)

recommendations. Briefly, ~ 6 μ g of DNA was sheared by sonication to yield DNA fragments of ~ 200 – 1000 bp. The methylated portion of DNA was immunoprecipitated using an antibody against 5-methylcytosine (Eurogentec). The precipitated DNA was further purified according to NimbleGen procedure; and amplified using a Whole Genome Amplification kit (Sigma). The resulting WGA products were hybridized to a NimbleGen CpG island plus promoter array that includes over 28 000 CpG islands based on the human genome release 18. Labeling, hybridization and scanning were performed by Roche-NimbleGen on one technical replicate. Peak detection was performed by the NimbleScan v2.3 software by NimbleGen. Peak finding was performed using a sliding-window one-sided Kolmogorov–Smirnov test with a minimum *P*-value cutoff of 0 and calling peaks ensuring that peaks encompass a minimum of four consecutive probes.

Total methylation analysis

The measurements of total cytosine methylation were essentially done according to the method developed by Fraga *et al.* (31) using a capillary electrophoresis from Beckman Coulter (PA800).

Purification and biochemical analysis of the DNMT3L^{R271Q} protein

The DNMT3A2 protein was purified as previously described (21). The DNMT3L R271Q mutation was created by PCR mutagenesis and the resulting cDNA inserted into a modified pMAL-c2 \times vector (New England Biolabs) so that the protein was expressed as an N-terminal fusion with a six histidine tag. The presence of the appropriate mutation was verified by DNA sequencing. The DNMT3L^{R271Q} protein was expressed in RosettaTM 2(DE3) cells (Novagen) and purified by nickel affinity chromatography followed by fractionation through a heparin column (GE Healthcare). After concentration through an Amicon Ultra Centrifugal filter device (Millipore), the protein was dialyzed into storage buffer (25 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1 mM dithiothreitol, 0.1% Triton X-100, 20% glycerol) and snap frozen. Protein concentration was determined by the absorbance at 280 nm using the theoretical extinction coefficient ($\epsilon = 68\,610\text{ M}^{-1}\text{ cm}^{-1}$) and by Bradford assays (Bio-Rad). The DNMT3L^{R271Q} protein preparation was over 96% pure as determined by band densitometry.

Protein activity was monitored by the incorporation of tritiated methyl groups from ³H-SAM (15 Ci/mmol, Perkin-Elmer) into double-stranded poly(dIdC) substrates, as previously described (21). In particular, purified proteins were pre-incubated together with ³H-SAM for 1 h at 37°C to reconstitute DNMT2A2:DNMT3L complexes and the reaction initiated by the addition of DNA. Assays were performed in Activity Buffer (25 mM Tris–HCl, pH 7.5, 50 mM KCl, 0.5 mM MgCl₂, 100 μ g/ml BSA and 1 mM dithiothreitol), unless specified.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Drs Lifeng Xu and Jue Lin from Dr Elizabeth Blackburn's laboratory at UCSF for their generous help with telomere assays.

Conflicts of Interest statement. The authors declare no conflict of interest.

FUNDING

This work was supported in part by a grant from March of Dimes Birth Defects Foundation (F.C.). M.S.K. was funded in part by Stem Cell Training Grant from California Institute of Regenerative Medicine.

REFERENCES

- Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
- Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*, **99**, 247–257.
- Bourc'his, D. and Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, **431**, 96–99.
- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536–2539.
- Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. and Sasaki, H. (2004) Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, **429**, 900–903.
- Robertson, K.D. (2005) DNA methylation and human disease. *Nat. Rev. Genet.*, **6**, 597–610.
- Delaval, K., Wagschal, A. and Feil, R. (2006) Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *Bioessays*, **28**, 453–459.
- Nicholls, R.D. and Knepper, J.L. (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu. Rev. Genomics Hum. Genet.*, **2**, 153–175.
- Horsthemke, B. and Ludwig, M. (2005) Assisted reproduction: the epigenetic perspective. *Hum. Reprod. Update*, **11**, 473–482.
- Houshdaran, S., Cortessis, V.K., Siegmund, K., Yang, A., Laird, P.W. and Sokol, R.Z. (2007) Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS ONE*, **2**, e1289.
- Murdoch, S., Djuric, U., Mazhar, B., Seoud, M., Khan, R., Kuick, R., Bagga, R., Kircheisen, R., Ao, A., Ratti, B. *et al.* (2006) Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans. *Nat. Genet.*, **38**, 300–302.
- Richardson, B. (2003) DNA methylation and autoimmune disease. *Clin. Immunol.*, **109**, 72–79.
- Feinberg, A.P. and Tycko, B. (2004) The history of cancer epigenetics. *Nat. Rev. Cancer*, **4**, 143–153.
- Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
- Goll, M.G. and Bestor, T.H. (2004) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–574.
- Chen, Z.X. and Riggs, A.D. (2005) Maintenance and regulation of DNA methylation patterns in mammals. *Biochem. Cell Biol.*, **83**, 438–448.
- Chen, T. and Li, E. (2006) Establishment and maintenance of DNA methylation patterns in mammals. *Curr. Top. Microbiol. Immunol.*, **301**, 179–201.
- Chedin, F., Lieber, M.R. and Hsieh, C.L. (2002) The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proc. Natl Acad. Sci. USA*, **99**, 16916–16921.
- Chen, Z.X., Mann, J.R., Hsieh, C.L., Riggs, A.D. and Chedin, F. (2005) Physical and functional interactions between the human DNMT3L protein and members of the *de novo* methyltransferase family. *J. Cell. Biochem.*, **95**, 902–917.
- Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature*, **449**, 248–251.
- Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C. and Chedin, F. (2006) Reconstitution and mechanism of the stimulation of *de novo* methylation by human DNMT3L. *J. Biol. Chem.*, **281**, 25893–25902.
- Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D. *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature*, **448**, 714–717.
- Nady, N., Min, J., Kareta, M.S., Chedin, F. and Arrowsmith, C.H. (2008) A SPOT on the chromatin landscape? Histone peptide arrays as a tool for epigenetic research. *Trends Biochem. Sci.*, **33**, 305–313.
- Goll, M.G., Kirpekar, F., Magerl, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, **311**, 395–398.
- Rai, K., Chidester, S., Zavala, C.V., Manos, E.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2007) Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.*, **21**, 261–266.
- Hermann, A., Schmitt, S. and Jeltsch, A. (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.*, **278**, 31717–31721.
- Kunert, N., Marhold, J., Stanke, J., Stach, D. and Lyko, F. (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development*, **130**, 5083–5090.
- El-Maarri, O., Becker, T., Junen, J., Manzoor, S.S., Diaz-Lacava, A., Schwaab, R., Wienker, T. and Oldenburg, J. (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum. Genet.*, **122**, 505–514.
- Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.*, **38**, 1378–1385.
- Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y. and Jinno, Y. (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann. Hum. Genet.*, **68**, 196–204.
- Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Orioste, M., Benitez, J. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA*, **102**, 10604–10609.
- Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M. and Meaney, M.J. (2004) Epigenetic programming by maternal behavior. *Nat. Neurosci.*, **7**, 847–854.
- Sinclair, K.D., Allegrucci, C., Singh, R., Gardner, D.S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J.F., Rees, W.D., Maloney, C.A. *et al.* (2007) DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proc. Natl Acad. Sci. USA*, **104**, 19351–19356.
- Valenza-Schaerly, P., Pickard, B., Walter, J., Jung, M., Pourcel, L., Reik, W., Gauguier, D., Vergnaud, G. and Pourcel, C. (2001) A dominant modifier of transgene methylation is mapped by QTL analysis to mouse chromosome 13. *Genome Res.*, **11**, 382–388.
- Friso, S., Choi, S.W., Girelli, D., Mason, J.B., Dolnikowski, G.G., Bagley, P.J., Olivieri, O., Jacques, P.F., Rosenberg, I.H., Corrocher, R. *et al.* (2002) A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. *Proc. Natl Acad. Sci. USA*, **99**, 5606–5611.
- Stern, L.L., Mason, J.B., Selhub, J. and Choi, S.W. (2000) Genomic DNA hypomethylation, a characteristic of most cancers, is present in peripheral leukocytes of individuals who are homozygous for the C677T polymorphism in the methylenetetrahydrofolate reductase gene. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 849–853.
- Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grander, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C. and Ekstrom, T.J. (2006) LUMA (LUMinometric Methylation Assay)—a high throughput method to the analysis of genomic DNA methylation. *Exp. Cell Res.*, **312**, 1989–1995.

38. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M. and Gartler, S.M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl Acad. Sci. USA*, **96**, 14412–14417.
39. Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J. and Viegas-Pequignot, E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187–191.
40. 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.
41. Djuric, U., El-Maarri, O., Lamb, B., Kuick, R., Seoud, M., Coullin, P., Oldenburg, J., Hanash, S. and Slim, R. (2006) Familial molar tissues due to mutations in the inflammatory gene, NALP7, have normal postzygotic DNA methylation. *Hum. Genet.*, **120**, 390–395.
42. Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D. and Pfeifer, G.P. (2009) A human B cell methylome at 100-base pair resolution. *Proc. Natl Acad. Sci. USA*, **106**, 671–678.
43. Bock, C., Walter, J., Paulsen, M. and Lengauer, T. (2007) CpG island mapping by epigenome prediction. *PLoS Comput. Biol.*, **3**, e110.
44. Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.
45. Oakes, C.C., La Salle, S., Smiraglia, D.J., Robaire, B. and Trasler, J.M. (2007) A unique configuration of genome-wide DNA methylation patterns in the testis. *Proc. Natl Acad. Sci. USA*, **104**, 228–233.
46. La Salle, S., Oakes, C.C., Neaga, O.R., Bourc'his, D., Bestor, T.H. and Trasler, J.M. (2007) Loss of spermatogonia and wide-spread DNA methylation defects in newborn male mice deficient in DNMT3L. *BMC Dev. Biol.*, **7**, 104.
47. Huntriss, J., Hinkins, M., Oliver, B., Harris, S.E., Beazley, J.C., Rutherford, A.J., Gosden, R.G., Lanzendorf, S.E. and Picton, H.M. (2004) Expression of mRNAs for DNA methyltransferases and methyl-CpG-binding proteins in the human female germ line, preimplantation embryos, and embryonic stem cells. *Mol. Reprod. Dev.*, **67**, 323–336.
48. 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.
49. Blasco, M.A. (2007) The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.*, **8**, 299–309.
50. Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M. and Blasco, M.A. (2006) DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat. Cell Biol.*, **8**, 416–424.
51. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K. and Selig, S. (2008) Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum. Mol. Genet.*, **17**, 2776–2789.
52. Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M. and Sasaki, H. (2007) Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.*, **16**, 2272–2280.
53. Flanagan, J.M., Munoz-Alegre, M., Henderson, S., Tang, T., Sun, P., Johnson, N., Fletcher, O., Dos Santos Silva, I., Peto, J., Boshoff, C. *et al.* (2009) Gene body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum. Mol. Genet.*, January 19 [Epub ahead of print].
54. Rauch, T.A., Zhong, X., Wu, X., Wang, M., Kernstine, K.H., Wang, Z., Riggs, A.D. and Pfeifer, G.P. (2008) High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl Acad. Sci. USA*, **105**, 252–257.
55. El-Maarri, O., Herbiniaux, U., Walter, J. and Oldenburg, J. (2002) A rapid, quantitative, non-radioactive bisulfite-SNuPE-IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res.*, **30**, e25.
56. Zaykin, D.V., Westfall, P.H., Young, S.S., Karnoub, M.A., Wagner, M.J. and Ehm, M.G. (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum. Hered.*, **53**, 79–91.
57. Becker, T., Cichon, S., Jonson, E. and Knapp, M. (2005) Multiple testing in the context of haplotype analysis revisited: application to case-control data. *Ann. Hum. Genet.*, **69**, 747–756.
58. Becker, T. and Knapp, M. (2004) Maximum-likelihood estimation of haplotype frequencies in nuclear families. *Genet. Epidemiol.*, **27**, 21–32.
59. Waha, A., Guntner, S., Huang, T.H., Yan, P.S., Arslan, B., Pietsch, T., Wiestler, O.D. and Waha, A. (2005) Epigenetic silencing of the protocadherin family member PCDH-gamma-A11 in astrocytomas. *Neoplasia*, **7**, 193–199.
60. Huang, T.H., Perry, M.R. and Laux, D.E. (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.*, **8**, 459–470.
61. Ji, X., Li, W., Song, J., Wei, L. and Liu, X.S. (2006) CEAS: *cis*-regulatory element annotation system. *Nucleic Acids Res.*, **34**, W551–W554.