Published online January 20, 2006

528–542 *Nucleic Acids Research, 2006, Vol. 34, No. 2*
doi:10.1093/nar/gkj461

Microarray-based DNA methylation profiling: technology and applications

Axel Schumacher, Philipp Kapranov[1], Zachary Kaminsky, James Flanagan, Abbas Assadzadeh, Patrick Yau[2], Carl Virtanen[2], Neil Winegarden[2], Jill Cheng[1], Thomas Gingeras[1] and Arturas Petronis*

The Krembil Family Epigenetics Laboratory, Centre for Addiction and Mental Health, 250 College Street, Toronto, ON, Canada M5T 1R8, [1]Affymetrix, Santa Clara, USA and [2]The Microarray Centre, The University Health Network, 200 Elizabeth Street, Toronto, ON, Canada M5G 2C4

Received November 21, 2005; Revised December 20, 2005; Accepted January 5, 2006

ABSTRACT

This work is dedicated to the development of a technology for unbiased, high-throughput DNA methylation profiling of large genomic regions. In this method, unmethylated and methylated DNA fractions are enriched using a series of treatments with methylation sensitive restriction enzymes, and interrogated on microarrays. We have investigated various aspects of the technology including its replicability, informativeness, sensitivity and optimal PCR conditions using microarrays containing oligonucleotides representing 100 kb of genomic DNA derived from the chromosome 22 *COMT* region in addition to 12 192 element CpG island microarrays. Several new aspects of methylation profiling are provided, including the parallel identification of confounding effects of DNA sequence variation, the description of the principles of microarray design for epigenomic studies and the optimal choice of methylation sensitive restriction enzymes. We also demonstrate the advantages of using the unmethylated DNA fraction versus the methylated one, which substantially improve the chances of detecting DNA methylation differences. We applied this methodology for fine-mapping of methylation patterns of chromosomes 21 and 22 in eight individuals using tiling microarrays consisting of over 340 000 oligonucleotide probe pairs. The principles developed in this work will help to make epigenetic profiling of the entire human genome a routine procedure.

INTRODUCTION

Over the last decade the field of DNA methylation has grown dramatically and become one of the most dynamic and rapidly developing branches of molecular biology. The methyl group at the fifth position of the cytosine pyrimidine ring, that is present in about 80% of CpG-dinucleotides in the human genome, can be of major functional significance and is regarded as the 'fifth base' of the genome (1). DNA methylation, along with histone modifications (acetylation, methylation, phosphorylation and the like), are referred to as epigenetic phenomena that control various genomic functions without a change in nucleotide sequence (2). Such functions include meiotic and mitotic recombination, replication, control of 'parasitic' DNA elements, establishing and maintenance of gene expression profiles, X chromosome inactivation as well as regulation of developmental programming and cell differentiation (3–6). Aberrations in epigenetic regulation, or 'epimutations', cause several paediatric syndromes (Prader–Willi [OMIM #176270], Angelman [OMIM #105830], Beckwith–Wiedemann [OMIM #130650] and Rett [OMIM #312750]) (7) and may also predispose to cancer (8).

Our understanding of the peculiarities of DNA methylation in the human genome is still very superficial. Based on the review of available publications, our estimate is that <0.1% of the genome has been subjected to a detailed DNA modification analysis. The recently completed Human Genome sequencing project did not attempt to differentiate between methylated and unmethylated cytosines. To some extent our understanding of the dynamic state of genome-wide DNA methylation has been hampered by the lack of high-throughput technologies that would interrogate DNA methylation profiles over large genomic regions. A gold standard technique in DNA methylation studies, the bisulfite modification-based fine mapping of ^{met}C (9), although precise, is very labour intensive and in

*To whom correspondence should be addressed. The Krembil Family Epigenetics Laboratory, Room 28, Centre for Addiction and Mental Health, 250 College Street, Toronto, ON, Canada M4T 1R8. Tel: +1 416 5358501 4880; Fax: +1 416 979 4666; Email: Arturas_Petronis@camh.net

most cases limited to short DNA fragments, often less than a kilobase.

The advent of microarray technologies that enabled the interrogation of a large number of DNA/RNA fragments in a highly parallel fashion has opened new opportunities for epigenetic studies (10). A number of microarray-based technologies used for epigenetic analyses are already available (11–23). However, all of these methods have some limitations, which renders them unsuitable for some experimental setups. Additionally, many technological parameters, such as the influence of DNA sequence variations, amplification conditions and sensitivity of the methods have not been investigated before. Here we present a detailed analysis of various parameters of epigenetic profiling and provide a substantially improved microarray-based high-throughput technology for DNA methylation profiling of DNA regions that span from hundreds of kilobases to megabases. Eventually, this technology will be applied to the entire human genome, as exemplified by the methylation mapping of chromosomes 21 and 22 as reported here.

MATERIALS AND METHODS

Microarray fabrication and data processing

COMT and CpG island microarrays were printed on Corning CMT-GAPSII slides (Corning Life Sciences, Acton, MA) using a VersArray ChipWriter Pro System (Bio-Rad Laboratories, Hercules, CA). For the *COMT* array, we designed 384 oligonucleotides (Operon/Qiagen, US), each 50 bases long, representing every restriction fragment flanked by HpaII, Hin6I and AciI restriction sites. In addition, control DNA fragments containing λ phage, pBR322, ΦX174 and pUC57 sequences were spotted on the slide. Each oligonucleotide was diluted to a 25 μM solution and spotted four times to give a total of 1536 elements. In addition, 192 blank spots consisted of SSC buffer and 48 spots contained *Arabidopsis* clones. The human CpG island array contains 12 192 sequenced CpG island clones derived from a CpG island library that was originally created with MeCP2 DNA binding columns (24,25).

Hybridized arrays were scanned on a GenePix 4000A scanner (Axon Instruments, Union City/CA) and analysed using the GenePix 6.0 software. The GenePix PMT voltage for Cy3 and Cy5 channels were balanced with the histogram feature of the scanner software to ensure a similar dynamic range for the two channels. Final scans were taken at 10 μm resolution, and images for each channel were saved as separate 16-bit TIFF files. The emission signals for each channel were determined by subtracting the local background from its corresponding median average intensity. These raw data were either exported into a custom Excel spreadsheet for subsequent data analysis or directly imported into the Acuity 4.0 software (Axon Instruments). The resulting datasets were normalized for the normalization features (spike-DNAs) and for signal intensity (Lowess normalization).

Profiling of unmethylated sites in the brain tissue of eight adults was carried out using a tiling array spanning ~12 Mb of non-repetitive sequence of chromosome 21 and 22 (q arms), with probes spaced on average every 35 bp center-to-center (26). The genomic DNA from these individuals was cut with HpaII and Hin6I, amplified and hybridized to the microarray as described previously (26,27). Unprocessed total genomic DNA from the same brain region (prefrontal cortex) was used as a control. Unmethylated sites were defined using a two-step analysis approach similar to the one used to determine transcription factor binding sites in the chromatin immunoprecipitation (ChIP)-chip assay (27). First, a smoothing-window Wilcoxon approach was applied to generate a *P*-value graph for each individual where probe signal from the enriched fraction was compared with the total genomic DNA in a one-sided upper paired test. The window used in this report was 501 bp. Second, three thresholds were applied to determine the boundaries of the unmethylated site: (i) an individual probe threshold of $P < 10^{-4}$ to determine if a probe is significantly enriched in the unmethylated fraction compared with the control total genomic DNA; (ii) the maximum distance between the two positive probes set to 250 bp and (iii) the minimal size of a site set to 1 bp. The graphs can be downloaded from the internet (see Web resources). All coordinates and annotation analysis were done on the April 2003 version of the genome.

Methylation -sensitive digestion of genomic DNA (gDNA)

Prior to treatment with restriction enzymes, gDNA was supplemented with 'spike'-DNAs (different concentrations of λ and *Arabidopsis* fragments), which were used as controls for signal normalization. For enrichment of the unmethylated fraction, depending on the number of CpG dinucleotides to be interrogated, several combinations of methylation-sensitive enzymes, HpaII, Hin6I, AciI and HpyCH4IV, were used. gDNA was cleaved with a cocktail of these enzymes (10 U/μl in 2xY+/Tango buffer, Fermentas Life Sciences/Lithuania) for 8 h at 37°C. For enrichment of the methylated fraction, gDNA was cleaved by TasI or Csp6I (10 U/μl in G^+-buffer, Fermentas) for 8 h at 65°C (TasI) or at 37°C (Csp6I). After the restriction reaction, TasI was inactivated by 0.5 M EDTA.

Adaptor–ligation

For the ligation step, gDNA was supplemented with 8 GE MspI-cleaved pBR322 plasmid (1 GE = 1.45 pg/ 1 μg gDNA), which was used as control for a potential ligation bias. The ends of the cleaved DNA fragments were ligated to the unphosphorylated adaptors. Our adaptors contained a sequence-specific protruding end, a non-target homologous core sequence, a specific antisense-overhang that prevents tandem repeat formation and blunt-end ligation, a 'disruptor' sequence that interrupts the original restriction sites after ligation, a new non-palindromic Alw26I (BsmAI) restriction site that enables the blunt-end cleavage of the adaptor from the target sequences (e.g. for library enrichment) and a non-5′-complementary end. The CpG-overhang specific universal adaptor 'U-CG1' for the unmethylated DNA fraction ligates to DNA fragments generated by 11 CpG-methylation-sensitive restriction enzymes HpaII, Hin6I (Hinp1I), HpyCH4IV, Bsu15I (ClaI, BspDI), AciI (SsiI), Psp1406I (AclI), Bsp119I (AsuII), Hin1I (AcyI, BsaHI), XmiI (AccI), NarI, BstBI (FspII) and also TaqI and MspI, which are not affected by methylation of the internal cytosine. The adaptor represents the annealing product of the two primers U-CG1a, 5′-CGTGGAGACTGACTACCAGAT-3′, and U-CG1b, 5′-AGTTACATCTGGTAGTCAGTCTCCA-3′.

The AATT-overhang specific adaptor 'AATT-1' for the methylated DNA fraction fits to DNA ends produced by the restriction enzyme TasI (TspEI), whereas the 'TA-1' adaptor fits to ends produced by Csp6I, BfaI or MseI, respectively:

AATT-1a, 5′-AATTGAGACTGACTACCAGAT-3′; AATT-1b, 5′-AGTTACATCTGGTAGTCAGTCTC-3′; TA-1a, 5′-TATGAGACTGACTACCAGAT-3′; and TA-1b: 5′-AGTTACATCTGGTAGTCAGTCTCA-3′.

All adapters were prepared by mixing equimolar amounts of the primer pairs, incubating the mixture at 80°C for 5 min, and then cooling it down to 4°C with 1°C/min. The double-stranded adaptors [200 pmol/µl] were added at 0.1 pmol per enzyme for each ng of the cleaved DNA (e.g. 0.3 pmol/ng in a triple-digest HpaII/Hin6I/AciI). The ligation-mixture with 400 ng template DNA was supplemented with 2 µl of 10× ligation buffer (Fermentas), 1 µl ATP [10 mM] and water to 18 µl. The reaction was started in a thermal-cycler at 45°C for 10 min, chilled on ice and 2 µl T4 ligase (Fermentas) was added. The ligation reaction was carried out at 22°C for 18 h, followed by a heat-inactivation step at 65°C for 5 min. The mixture was then cooled down to room temperature with 1°C/min and stored at 4°C for subsequent procedures.

PCR

To control for a potential PCR bias, the DNA mixture was supplemented with 2 GE ΦX174 plasmid (1 GE = 1.8 pg of ΦX174 corresponding to 1 µg gDNA) that was cut with -HpyCH4IV and ligated to the adaptor. PCR amplifications were conducted for up to 25 cycles. A standard aminoallyl-PCR mixture included 400 ng of the ligate, 40 µl of 10× reaction-buffer (Sigma), 42 µl $MgCl_2$ [25 mM], 3 µl aminoallyl-dNTP Mix [containing 15 mM aminoallyl-dUTP, 10 mM dTTP and 25 mM each dCTP, dGTP and dATP], 200 pmol primer (U-CG1a, AATT-1b or TA-1b, respectively), 3 µl *Taq* enzyme (5 U/µl, NEB) and water to a final volume of 400 µl. For PCR conditions and generation of dye-coupled adaptor products see Supplementary Data.

Array hybridizations

Each microarray slide was prehybridized with a mixture consisting of DIG Easy Hyb (Roche Diagnostics), 25 µg/ml tRNA and 200 µg/ml BSA. The printed area was covered with the prehybridization mixture under a coverslip for 1 h at 45°C. The microarray slides were then washed in two changes of water for 2 min at 45°C, followed by two wash-steps at room temperature and a final wash-step in isopropanol for 1 min. The slides were immediately blown dry with pressurized air and stored for hybridization. The hybridization mixtures were then pipetted onto the arrays and covered with Sigma Hybri-slips. The microarrays were placed in hybridization chambers (Corning Microarray Technologies, NY) and incubated on a level surface for 16 h at 42°C for the COMT-arrays and 44–52°C for the CpG island microarrays in a covered water bath. The coverslips were removed by immersion of the arrays in a wash solution containing 2× SSC and 0.5% SDS (washing buffer I). The array was washed twice for 15 min at 42–52°C in washing buffer I (low stringency), followed by two wash-steps in washing buffer II (0.5× SSC, 0.5% SDS), followed by 2 min of incubation in water. The slides were then rinsed quickly in isopropanol and finally dried with pressurized air. The hybridization method used for the chromosome 21 and 22 tiling arrays was described before (26,27).

Whole genome amplification

Genomic DNA was amplified using the GenomiPhi Kit (Amersham Biosciences) according to the manufacturer's protocol. Briefly, 10 ng of gDNA (1 µl) was mixed with 9 µl of sample buffer, denatured at 95°C for 3 min, cooled on ice and then added to 9 µl of reaction buffer and 1 µl of Phi29 DNA polymerase. The reaction was incubated at 30°C for 16 h and then inactivated at 65°C for 10 min.

Bisulfite sequencing

The methylation status of a number of CpG islands were analysed by direct sequencing of sodium bisulphite modified gDNA (9). gDNA samples were subjected to bisulfite modification using a standard protocol (28). The primer sequences, PCR conditions and cloning methods are provided in the Supplementary Data.

Genomic DNA

Genomic DNA from all tissues was purified using standard laboratory methods (Phenol–Chloroform or Qiagen Blood and Cell DNA Midi columns). To avoid cross reactivity of amine groups with the aminoallyl-labeling procedure, DNA samples were stored in 0.5 M POPSO buffer (pH 8.0) instead of Tris–EDTA. Male placental DNA was purchased from Sigma and the post mortem brain samples were provided by the Stanley Medical Research Institute. All parts of the study were approved by the CAMH review/ethics board.

Web resources

All chromosome 21/22 tiling array data can be viewed in the UCSC genome browser available via the methylation database at www.epigenomics.ca. Additionally, the complete tiling array source data plus graphs that can be viewed in the Integrated Genome Browser (Affymetrix; www.affymetrix.com/support/developer/downloads/TilingArrayTools/index.affx) and can be downloaded at http://transcriptome.affymetrix.com/download/DataMethPaper (case sensitive). All coordinates and annotation analysis was done on the April 2003 version of the genome. SNP data were derived from the SNP consortium, www.ncbi.nlm.nih.gov/SNP.

OMIM numbers are derived from Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM. Genome annotations were derived from the ReSeq database, http://www.ncbi.nlm.nih.gov/RefSeq/ and the UCSC database, http://genome.ucsc.edu/cgi-bin/hgGateway.

RESULTS

Enrichment of the unmethylated fraction of gDNA

The strategy for enrichment of unmethylated portions of the genome is presented in Figure 1. gDNA is digested with methylation-sensitive restriction enzymes (Figure 1, middle panel). Whereas methylated restriction sites remain unaltered, the sites containing unmethylated CpGs are cleaved by the enzymes, and DNA fragments with 5′-CpG protruding ends are generated. The proportion of interrogated

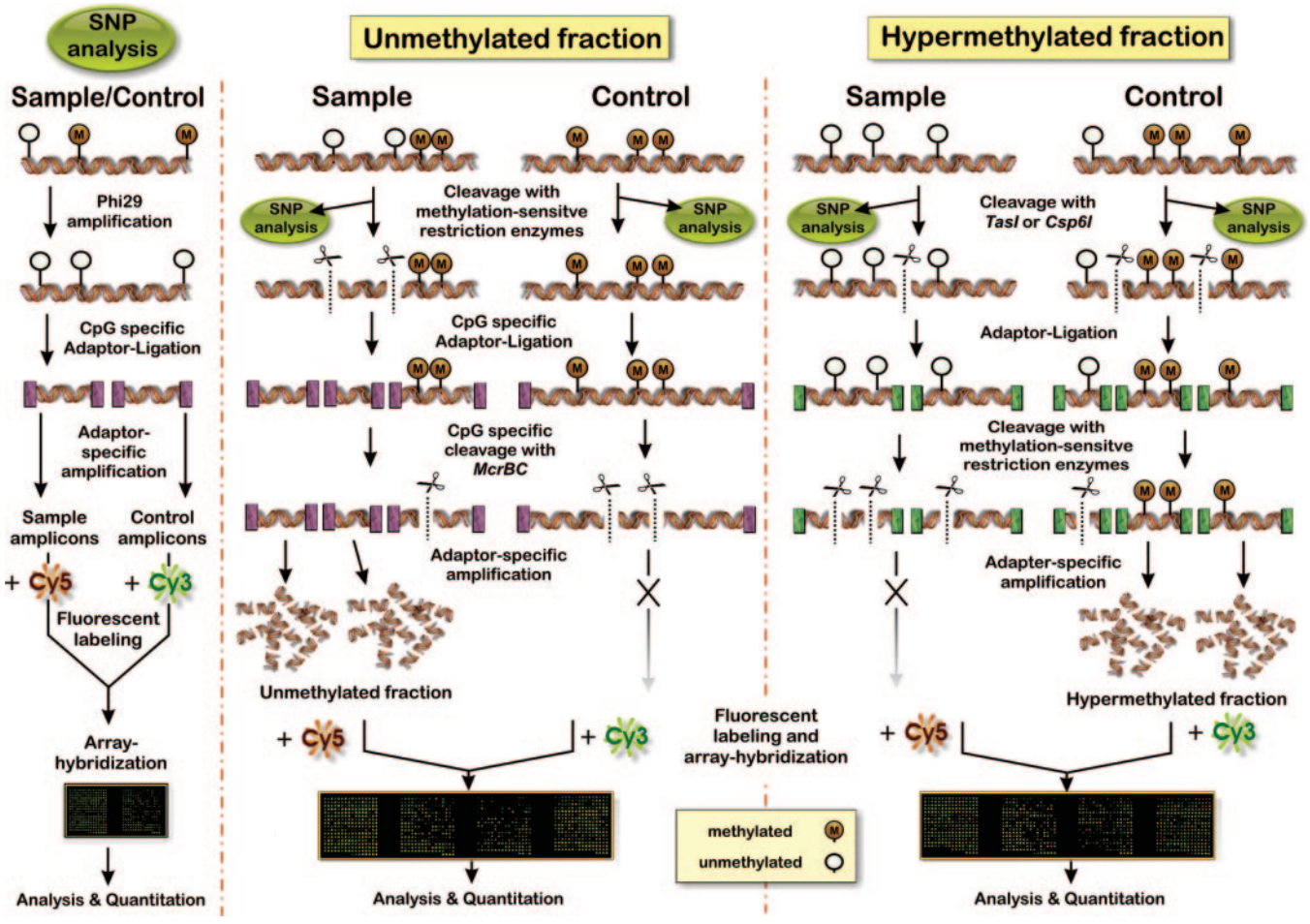


Figure 1. Schematic outline of the microarray-based method for identification of DNA methylation differences and DNA polymorphisms in genomic DNA. Left panel: analysis of DNA sequence variation. Middle panel: the main strategy of the method is based on enrichment of unmethylated DNA fragments. DNA samples are cleaved by methylation-sensitive restriction endonucleases, and the resulting DNA fragments are then selectively enriched by adaptor-specific aminoallyl-PCR's, labelled and hybridized to microarrays. Right panel: alternative procedure to enrich the hypermethylated DNA fraction.

CpG sites depends on the methylation-sensitive restriction enzymes used for the restriction of DNA. Based on our analysis of the CpG dinucleotides within the sites of methylation-sensitive restriction enzymes across several megabases of human gDNA, the combination of three enzymes, HpaII, Hin6I and AciI, should interrogate ~32% of all CpG dinucleotides in mammalian DNA (Table 1). The addition of two other relatively inexpensive methylation-sensitive CpG-overhang generating enzymes, HpyCH4IV and Hin1I, would theoretically increase the proportion of interrogated CpGs to ~41%. Depending on the microarray-type, in our experiments we usually use either a single enzyme or a 'cocktail' of up to three restriction enzymes. The application of a set of enzymes might be disadvantageous for the analysis of GC-rich regions as such a strategy would produce restriction fragments too short for an efficient hybridization. In the latter case, it is advisable to use a smaller number of restriction enzymes. Based on our experimental results and computer-based analysis of 100 randomly selected CpG islands, the most suitable restriction enzymes are Hin6I and HpaII, followed by AciI and Hin1I (Table 1). In contrast, for regular DNA sequences, double- or triple-digest combinations of AciI, HpaII, HpyCH4IV and Hin6I are recommended.

After the digestion of gDNA, the double-stranded adaptor U-CG1 is ligated to the CpG-overhangs. At this point, it is expected that most of the relatively short (<1.5 kb) and amplifiable DNA fragments derive from the unmethylated DNA regions. To some extent, the length of the amplified fragments depends on the primer annealing temperature of the PCR reaction (Figure 2A). Some ligation fragments, however, may still contain methylated cytosines. A proportion of such fragments can be eliminated by treatment with McrBC, which cleaves DNA containing ^{met}C and will not act upon unmethylated DNA. McrBC restriction sites consist of two half-sites of the form (G/A)^{met}C, which can be separated by up to 3 kb (29,30). Hence, as can be seen in Figure 2B, a proportion of DNA fragments with two or more (G/A)^{met}C within the restriction fragment are cleaved and therefore deleted from the subsequent enrichment steps. The remaining pool of unmethylated DNA fragments is then enriched by aminoallyl-PCR amplification that uses primers complementary to the adaptor U-CG1. One important advantage of using protruding ends in the adaptor–ligation step is that degraded

Table 1. Enzymes that generate protruding ends in the restriction fragments, which are complementary to the adaptors U-CG1, TA-1 and AATT-1

Enzymes	Recognition sequence	Percentage coverage of CpGs in human gDNA (%)	Number of fragments (per kb) in CpG islands*	Number of fragments (per kb) in non-CpG islands*
HpaII (BsiSI)	CCGG	8.6	**3.98**	1.18
Hin6I (HinP1I)	GCGC	6.4	**3.98**	0.61
AciI (SsiI)	CCGC	17.4	3.23	**1.79**
Hin1I (AcyI, BsaHI)	GPuCGPyC	2.0	1.92	0.11
HpyCH4IV	ACGT	6.6	1.31	1.08
Bsu15I (ClaI, BspDI)	ATCGAT	0.2	<0.01	0.02
NarI (MlyI)	GGCGCC	0.6	1.08	<0.01
Bsp119I (BstBI, AsuII)	TTCGAA	0.1	0.11	<0.01
Psp1406I (AclI, PspI)	AACGTT	0.3	<0.01	0.05
XmiI (AccI)	GTMKAC	0.1	0.19	0.34
TasI	AATT	**na**	0.80	**2.88**
Csp6I	GTAC	**na**	**2.23**	1.41
MseI	TTAA	**na**	0.80	**2.88**
BfaI	CTAG	**na**	1.56	1.55

Asterisk (*) indicates the number of 50 bp to 1.5 kb long ('informative') fragments, derived from several Mb of randomly selected CpG island and non-CpG island sequences on chromosomes 1, 2, 4, 5, 6, 9, 17, 19 and 20; bold numbers represent the most informative enzymes; na = not applicable; M = Adenine or Cytosine; K = Guanine or Thymine.

gDNA fragments (which are common in human post mortem tissues) will not be ligated and amplified, and therefore will not interfere with DNA methylation analysis.

Most previous microarray-based epigenetic studies target hypermethylated DNA sequences (15,17,31,32); however, interrogation of the unmethylated fraction is significantly more informative. For example, the 100 kb region of chromosome 22 interrogated by our COMT oligonucleotide array (*TXNRD2-COMT-ARVCF* region; Microarray Design), contains 2193 methylatable cytosines. Enrichment of the unmethylated fraction can generate up to 401 amplicons of sufficient size (50–1.5 kb), each representing the methylation status of at least one cytosine. In contrast, the combination of MseI (+BsuI, to remove unmethylated fragments), the most frequently used enzymes for enrichment of the hypermethylated fraction (15,17,31,32), would produce 227 amplicons. Seventy-seven amplicons would either contain no CpG dinucleotides or would be too short to stringently hybridize to a microarray. Of the remaining 150 fragments, 144 contain multiple CpGs; hence, they are not fully informative since a single unmethylated BsuI restriction site would eliminate the entire fragment from the eventual amplification. Overall, only 6 of the 2193 methylatable cytosines are truly informative, and none of these CpG dinucleotides are targeted by BsuI. Computer-based analysis of 50 randomly selected CpG island sequences revealed that the unmethylated fraction derived from HpaII cleavage results in ~22 times more fragments (19.9 fragments/kb) of the suitable size range (50 bp to 1.5 kb) than the hypermethylated fraction (0.9 fragments/kb) using MseI.

Nevertheless, analysis of the hypermethylated DNA fraction may also add some new information to the methylation profiles, especially in the case of hypermethylated CpG islands or when the overall level of methylation in the genome is low (e.g. in insects). Thus, we developed an additional, modified method to previously published methods of enrichment of methylated sequences to complement our data from the unmethylated fraction (Figure 1, right panel). This enrichment method relies on cleavage with the 4 bp frequent cutters TasI (AATT↓) and/or Csp6I (G↓TAC). Alternatively, BfaI or MseI can be used in combination with the Csp6I-specific adaptor. All four enzymes produce DNA fragments in mammalian genomes of an average length 400–750 bp. The recognition sequences of TasI and Csp6I are infrequent within GC-rich regions, leaving most CpG-islands intact. The analysis of 50 randomly selected CpG islands and several megabases of different chromosomes revealed that Csp6I would produce more informative fragments in CpG islands than a digest with MseI, whereas TasI and MseI produce informative fragments preferentially in DNA regions outside of CpG islands (Table 1). After ligation to the AATT- and TA-overhang specific adaptors 'AATT-1' and 'TA-1', the un- and hypomethylated ligation products are eliminated from the reaction by cleavage with a cocktail of methylation-sensitive restriction enzymes such as HpaII, HhaI (Hin6I), HpyCH4IV, Hin1I and AciI. Compared with a single digestion with BstUI (17), a cocktail of restriction enzymes will delete a higher percentage of unmethylated sequences from the DNA fraction. The remaining pool of mostly hypermethylated DNA fragments is subsequently enriched by the aminoallyl-PCR amplification as described for the unmethylated fraction, and then hybridized to a microarray (Figure 2C).

Microarray design

Various aspects of the microarray-based DNA modification profiling were investigated on the oligonucleotide microarray that interrogates ~100 kb fragment on 22q11.2 (Figure 3A). In addition to the catechol-*O*-methyltransferase (*COMT*, [MIM 116790]), this chromosomal region contains also the gene encoding the thioredoxin reductase 3 gene (*TXNRD2*, [MIM 606448]) and the armadillo repeat gene deleted in velocardiofacial syndrome (*ARVCF*, [MIM 602269]). For maximal informativeness, it is necessary to design oligonucleotides according to the restriction sites of the methylation sensitive endonucleases used for the treatment of gDNA (Figure 3B). For the *COMT* array, 384 oligonucleotides were designed, each 50 nucleotides long, representing every restriction fragment flanked by HpaII, Hin6I and AciI restriction sites. In addition, control DNA fragments containing λ phage, pBR322, ΦX174, pUC57 and *Arabidopsis* sequences were spotted on the array (Materials and Methods). Additionally, we used 12 192 element containing CpG island- and high-density chromosome 21/22-microarrays (Materials and Methods).

Detection of confounding effects of DNA sequence variation

Since restriction enzymes are used in the enrichment of differentially modified DNA fractions, DNA sequence variation may simulate epigenetic differences. However, until now, microarray methods used in epigenetic studies have not been

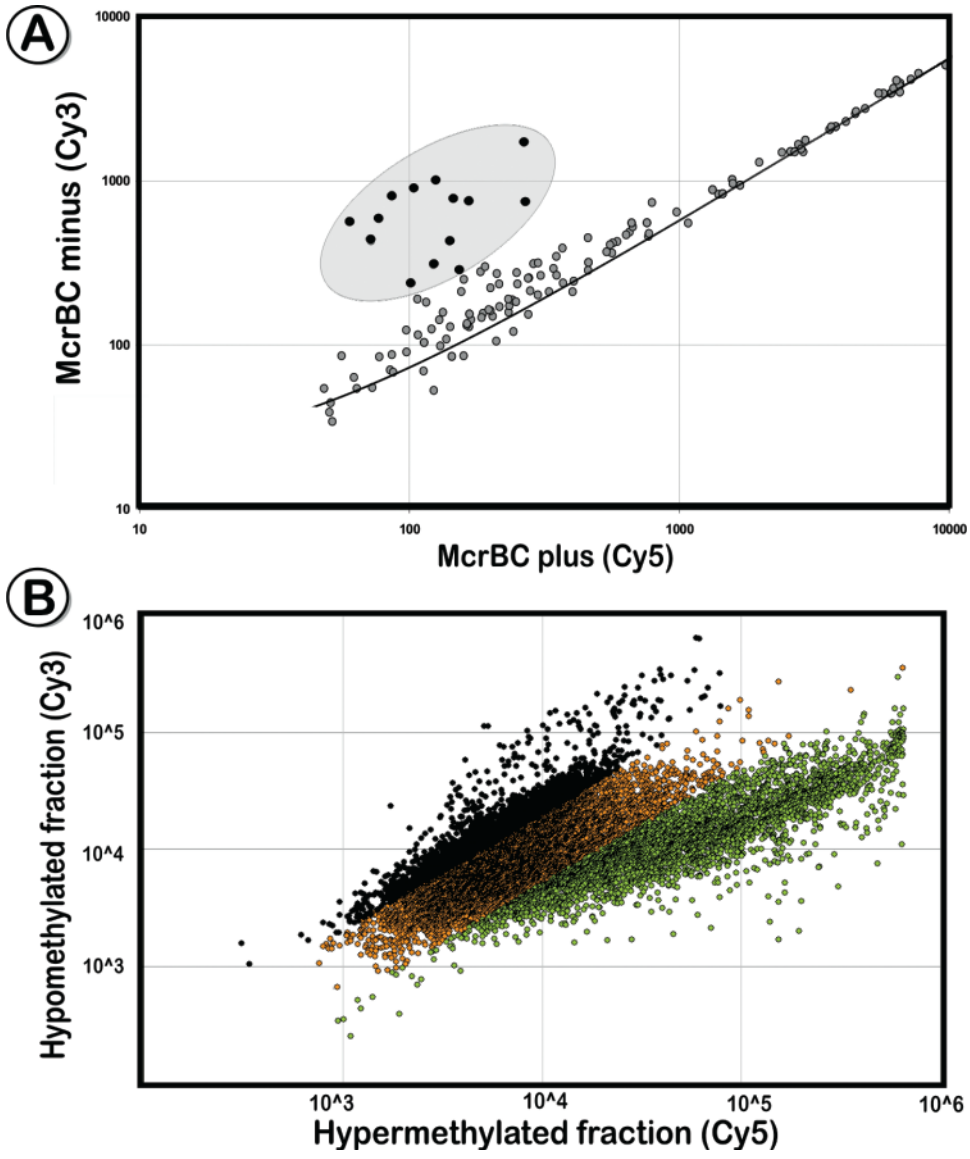


Figure 2. Selective enrichment of restriction fragments with the universal adaptor U-CG1. (**A**) Scatter plot that shows a comparison of ligation products treated with McrBC versus the untreated sample on the *COMT* array. McrBC treated fragments that contained at least two methylated cytosines were cleaved and could not be amplified in the following adaptor-PCR, resulting in reduced signal intensities in the Cy5 channel. (**B**) Co-hybridization of enriched unmethylated (Figure 1, middle panel) and hypermethylated (Figure 1, right panel) fragments derived from the same DNA source to a CpG island microarray. A large portion of amplicons is present only in one of the enriched fractions (marked black for log >0.3 black, green for log <−0.3). Although the hypermethylated fraction hybridized to ∼75% of the microarray spots, based on our DNA sequence analysis, only a small fraction of them provide epigenetic information in comparison with the unmethylated fraction.

differentiating between real DNA methylation differences and single nucleotide polymorphisms (SNPs) within the restriction sites of the applied restriction enzymes. This problem applies to some extent also to the ^{met}C antibody-based strategy (22), which does not differentiate unmethylated CpG and TpG dinucleotides. In order to exclude the impact of DNA sequence variation, two approaches are suggested. One is to check the available SNP databases in order to identify the DNA sequence variation within the restriction sites of the enzymes used. For example, our 100 kb *COMT* array contains a total of 273 SNPs (SNPper, http://snpper.chip.org/bio/snpper-enter), of which 101 (37%) reside within CpG dinucleotides and 55 (20%) are located within the restriction site of the four main enzymes used to interrogate methylation patterns, HpaII, Hin6I, AciI and HpyCH4IV. The majority of these CpG-SNPs were located in AciI and HpaII restriction sites, with Hin6I and HpyCh4IV sites containing fewer polymorphisms (data not shown). Another approach to test for DNA polymorphisms is the use of restriction endonuclease isoschizomers with different sensitivity to CpG methylation. However, this approach is currently only possible for HpaII/MspI as there are no isoschizomers for most other methylation sensitive restriction enzymes.

The third approach to differentiate the DNA sequence effects from the genuine epigenetic differences consists of performing an identical microarray experiment on the same DNA sample that has been stripped of all methylated cytosines. Our protocol utilizes the Phi29 DNA polymerase

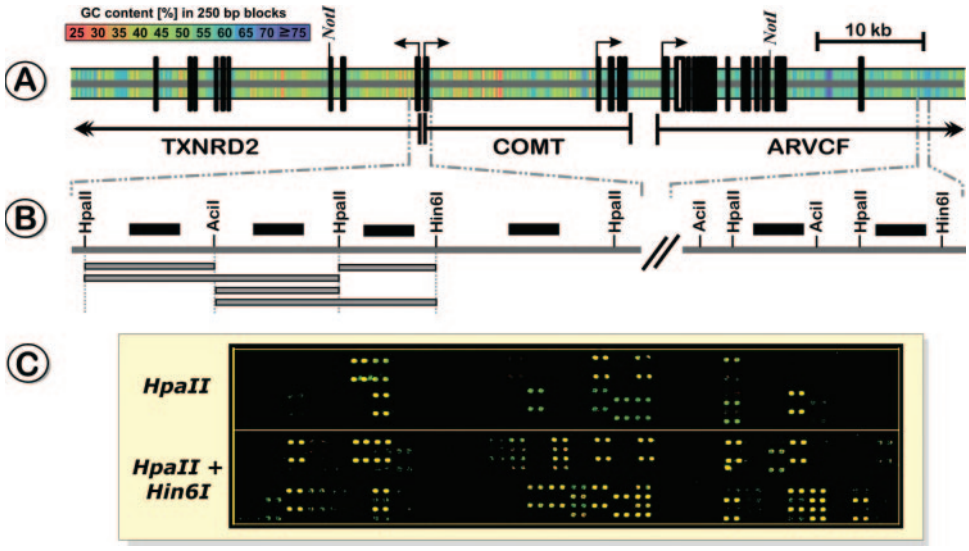


Figure 3. (**A**) Structure and GC-content of the chromosomal region on human chromosome 22q11.2 that spans the catechol-*o*-methyltransferase gene (*COMT*), the thioredoxin reductase 2 gene (*TXNRD2*) and the armadillo repeat gene deleted in VCFS (*ARVCF*). Vertical black bars represent exons. (**B**) To determine the methylation profile of the 100 kb *TXNRD2-COMT-ARVCF* region, 384 oligonucleotides (50mers, black horizontal bars) were designed based on the restriction sites for the methylation-sensitive endonucleases, HpaII, Hin6I and AciI (additional alternative enzymes are HpyCH4IV or Hin1I). Depending on the methylation status of the CpG-dinucleotides several combinations of amplicons (grey horizontal bars) can potentially hybridize to the oligonucleotides. (**C**) Typical hybridization patterns of the hypomethylated fraction of human gDNA on the *COMT* oligonucleotide-microarray. As discussed in Results, the complexity and informativeness of the hybridization signals increases with increasing number of methylation-sensitive restriction enzymes.

to amplify whole genomic DNA, which creates a copy of the genome with all methylated cytosines replaced by unmethylated cytosines. Amplified DNA samples are then subjected to the same steps as depicted in Figure 1 and hybridized on the microarrays. In this experiment all of the outliers must be a result of DNA sequence variations within the restriction sites of the enzymes used. These data can then be plotted against the DNA methylation data, which are assayed in parallel (Figure 4). In six experiments that used amplified genomic DNA, the number of SNP-based outliers (threshold log-ratio <-0.3, >0.3) ranged from 272 to 741 (432 ± 165, mean ± SD), or 2.2–6.1% of 12 192 CpG islands. Out of these, 72–234 (120 ± 66, mean ± SD) were initially identified as DNA methylation differences in microarray experiments using the unmethylated fraction derived from the triple-digest with HpaII, AciI and Hin6I. From the CpG island array studies, our estimate is that 10–30% of the outliers detected in DNA methylation experiment could be due to DNA sequence variation.

Reproducibility

To test the reproducibility of the method, a genomic DNA sample was split and subjected to the procedure of enrichment of the unmethylated fraction. The resulting amplification products were labelled with Cy5 and Cy3 and then co-hybridized on the *COMT* array, which contains probes that flank the HpaII, Hin6I and AciI restriction fragments around the *COMT* gene. The Cy3 and Cy5 hybridization intensities exhibited very similar values ($R^2 = 0.997$; Figure 5A). Analogous experiments, including switch dye hybridizations, were repeated several times also with the CpG island arrays and in all cases were highly reproducible ($R^2 > 0.97$).

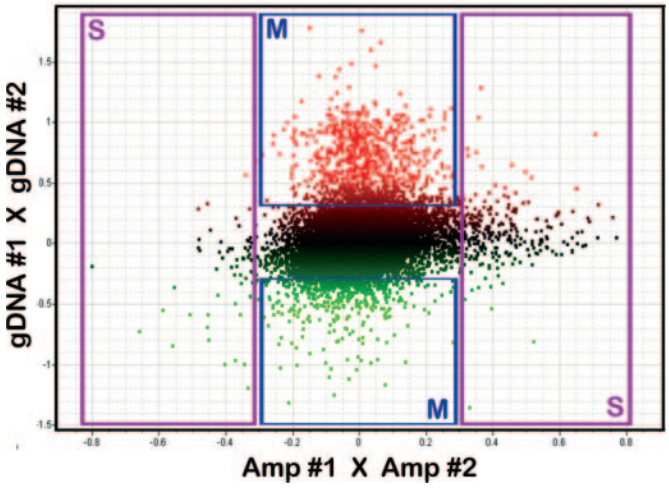
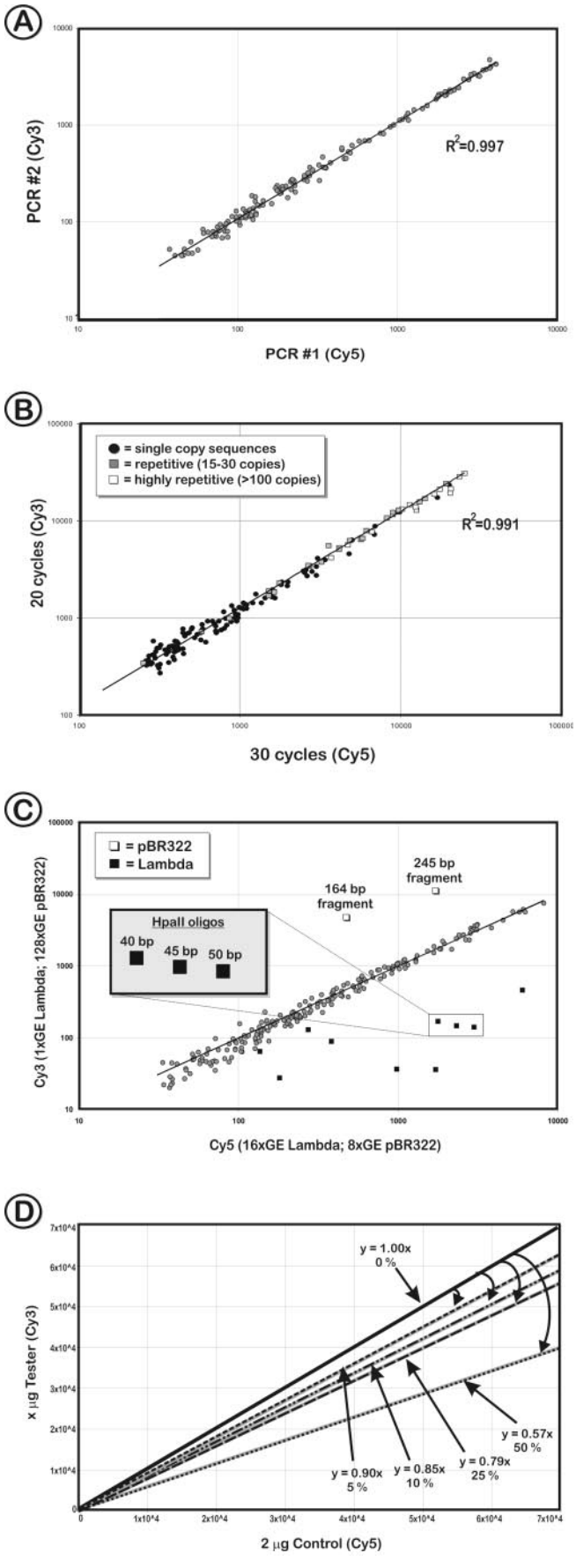


Figure 4. Combined methylation- and SNP-analysis on a CpG island microarray. The data of two separate hybridizations of DNA samples derived from post mortem brain of two individuals are plotted against each other. The *Y*-axis contains the data derived from a methylation analysis (triple-cleavage with HpaII, Hin6I and AciI), whereas the *X*-axis contains the SNP data derived from the hybridization of the same DNA samples, which were subjected to the entire genome amplification prior to cleavage by the methylation-sensitive restriction enzymes (Materials and Methods). Scale: log (Cy5/Cy3); an increased log-value on the *Y*-axis is indicated by red versus a decreased log-value represented by green. Significant outliers (log-ratio <-0.3, >0.3, 2-fold difference) can be classified into four clusters (S = SNPs, M = DNA methylation differences), enabling the differentiation of epigenetic differences and nucleotide polymorphisms between the test-samples. Amp = Whole-genome amplified sample.

Another critical factor in the amplification of unmethylated or hypermethylated DNA fragments is to ensure that no sequence specific bias is introduced. The rate of amplification of repetitive sequences generally declines faster than that of



less abundant fragments in the later cycles of PCR (33). With increasing amplification cycles, repetitive DNA strands reach relatively high concentration and begin re-annealing to each other during the steps below the DNA melting temperature. To avoid this, a two-temperature PCR that uses a combined high-temperature elongation–annealing step was applied. A series of experiments were performed investigating how the number of PCR cycles would affect the hybridization patterns. As can be seen in Figure 5B, the relative intensities of the hybridization signals of both single copy sequences and repetitive DNA fragments, were similar in the range of 20–30 amplification cycles ($R^2 = 0.991$). Only when increasing the cycle numbers beyond 40 cycles was a biased amplification of some DNA sequences observed (data not shown).

Sensitivity

To test if differentially represented DNA fragments in two different DNA samples can be detected by this method, prior to methylation-sensitive cleavage, human gDNA was 'spiked' with unmethylated heterologous DNA, λ phage and pBR322 plasmid (Figure 5C). Each sample was supplemented with a different amount of spike-DNA, therefore mimicking differentially methylated sequences. The exact amount of λ and pBR322 corresponded to increasing numbers of human genomic equivalents (1 GE of 'spike' DNA equals 16.28 pg λ/µg gDNA and 1.45 pg/µg gDNA of pBR322, respectively). Hence, each of the experiments compared the intensities generated by 1 GE of λ plus 128 GE of pBR322 (*Y*-axis) versus 16 GE of λ plus 8 GE of pBR322 (*X*-axis). While the plotted signal intensities of the human gDNA sequences are positioned on or close to the regression line (indicating no methylation difference), the λ and pBR322 fragments were identified as outliers. The average signal

Figure 5. Reproducibility and sensitivity of the method. (**A**) A *COMT* microarray scatter plot representing two sets of amplification products derived from the same DNA source but produced at different time points by different researchers. The high-correlation coefficient of signal intensities demonstrates a high reproducibility of the method. (**B**) Influence of the PCR cycle number. Scatter plot diagrams show hybridization signal intensities of the unmethylated fraction that was amplified using 20 PCR cycles (Cy3 channel) and 30 cycles (Cy5 channel). Amplification products of each PCR were co-hybridized to the *COMT* microarray that contained oligonucleotides representing single copy sequences (closed circles), partially repetitive sequences (grey squares; 15–99 copies/genome) and highly repetitive DNA fragments (open squares; >100 copies/genome), such as ALU and LINE repeats. (**C**) Scatter plot representing the unmethylated fraction of human gDNA 'spiked' with different amounts of control DNA. The test samples were hybridized to the *COMT* array and contained either a 16-fold excess of λ DNA (16 genome equivalents [GE] versus 1 GE; 10 fragments) or a 16-fold excess of pBR322 (128 GE versus 8 GE; 2 fragments), respectively. The amplicons of the spiked DNA (representing unmethylated DNA) can be easily distinguished as outliers; whereas the signals representing gDNA are located close to the regression line. Median signal intensities of different length oligonucleotides (40–50 bases) that target a specific HpaII restriction fragment in λ DNA reveal that the length of spotted sequences directly influences the spot intensity and therefore the sensitivity of the microarray. (**D**) Sensitivity of the CpG-island microarray hybridization. Control amplicon (2 µg) (post mortem brain, unmethylated fraction) was labelled with Cy5 and co-hybridized with 2 µg (0% difference), 1.9 µg (5% difference), 1.8 µg (10% difference), 1.5 µg (25% difference) or 1.0 µg (50% difference) of Cy3-labelled amplicon. For each hybridization to a *COMT* array, the regression lines represent the overall intensity that mimics methylation differences over the entire sample. The decrease of amount of DNA is reflected in the angle of the regression lines, which deviated by 5–7% from the expected values.

intensity ratio of λ oligonucleotides was 15.4, which is very close to the ratio of spiked-DNA (16:1). The intensity values for pBR322 were not as linear and exhibited a 6.5- to 10-fold difference (expected the same ratio of 1:16), most likely due to saturation effects of the hybridization.

In order to determine the sensitivity of the hybridization *per se*, a control amplicon DNA was compared with itself but by decreasing the amounts of DNA by 5, 10, 25 and 50%. On the global level, the regression lines [$y = f(x)$] reflected reproducible differences of the amount of amplicon DNA used in the hybridization and varied by 5–7% from the expected values (Figure 5D). Individual sites exhibited a lower accuracy, which depended on the signal intensity, i.e. the stronger the signal, the closer the observed spot intensity was to the expected one. The rate of false outliers (log-ratio <-0.3; >0.3; 2-fold difference) was on average 3%. Usually, replication of microarray experiments reduced the degree of aberration (log-ratio <-0.3; >0.3) below 2% for all types of microarrays.

Examples of DNA methylation profiles

Identification of DNA modification differences is provided in a series of examples below. The *COMT* oligonucleotide array was used to identify DNA methylation changes in a brain tumour (Figure 6A). In contrast to the pair of control brain DNA samples, where hybridization signals are close to the regression line (indicating similar DNA methylation patterns), a visible proportion of the hybridization signals originating from the unmethylated DNA fraction of the brain tumour deviates from the regression line. More subtle changes in DNA methylation patterns have been identified when post mortem brain tissues of healthy individuals were compared with the same tissues from schizophrenia patients (A. Schumacher, A. Petronis, manuscript in preparation; representative example is shown in Figure 6B). The differences of the cancer and psychosis studies show that diseases other than cancer may reveal more subtle epigenetic differences, and therefore, the informativeness and sensitivity of the epigenetic profiling method is of critical importance.

Another application of the technology includes epigenetic profiling of different tissues. One example of tissue specific effects is shown using the CpG island microarrays that contain 12 192 CpG island clones of whom 8025 represent unique sequences. CpG islands tend to be found in many promoter sequences and their methylation has profound effects on gene silencing in mammalian genomes. The scatter plot shows two distinct spot areas, which represent predominantly unmethylated fragments in placenta (yellow spots) and brain (orange spots), respectively (Figure 7A). About 11% of the CpG island fragments exhibited 2-fold or more signal intensity difference between the two tissues. Some of the strongest brain-specific signals could be identified for CpG islands associated with neuronal genes such as *DPYSL5*, *FABP7*, *DIRAS2*, *GRIN3A*, *SLC24A3* and *DSCAML1*, whereas strong placenta-specific outliers were associated with genes expressed in placenta, such as *PCM1*, *CCND1*, *HA-1* and *ADAMTSL1*. Overall, analysis revealed that brain DNA harboured notably more unmethylated CpG islands than placenta DNA.

Verification of detected methylation differences

Several loci that displayed methylation differences in our experiments were selected for verification by the sodium bisulfite modification mapping of methylated cytosines (Materials and Methods). The technique is based on the reaction of gDNA with sodium bisulfite under conditions such that cytosine is deaminated to uracil but 5-methylcytosine remains unaltered. In the sequencing of amplified products, all uracil and thymine residues are detected as thymine and only ^{met}C residues remain as cytosine. The sites for the methylation-sensitive restriction enzymes used in our experiments showed the expected methylation difference across the DNA samples, as exemplified for CpG island clones located in the promoter region of galectin-1 and in the promoter region of a brain-specific transcript CR606704 (Figure 7B and C).

Chromosome-wide mapping of DNA methylation differences

Analysis of the unmethylated fraction from brain specific DNA of eight adults using a chromosome 21/22 tiling array detected 488–747 unmethylated sites per sample (Table 2). This number increased to 977 in a merged map, showing that many sites were common between different individuals. The vast majority of the sites (~90%) were positioned outside of the 5′ ends and 5′ flanking regions of the genes consistent with abundant transcriptional activity and a significant fraction of transcription factor binding sites found outside of known annotations (26,27,34). The unmethylated sites outside of the 5′ ends of known genes were about equally distributed

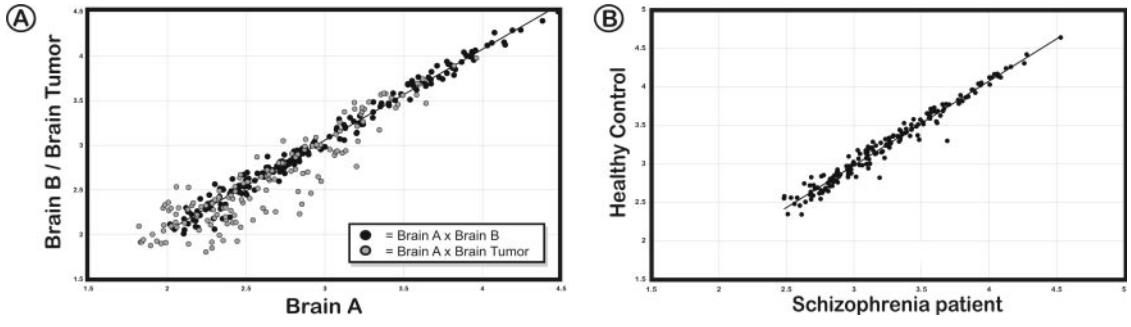


Figure 6. Applications of the epigenetic profiling technology. (**A**) Changes of methylation profiles at *TXNRD2-COMT-ARVCF* in a brain tumour. The data from two different microarrays experiments are superimposed over each other. The analysis of two post mortem brain samples (closed dots) reveals no major difference in methylation levels, whereas the signal intensities vary significantly in the brain tumour (grey dots) when compared with the normal brain. (**B**) The comparison of DNA methylation profiles using the COMT microarray in brain tissue of a healthy control and a schizophrenia patient displays subtle epigenetic differences.

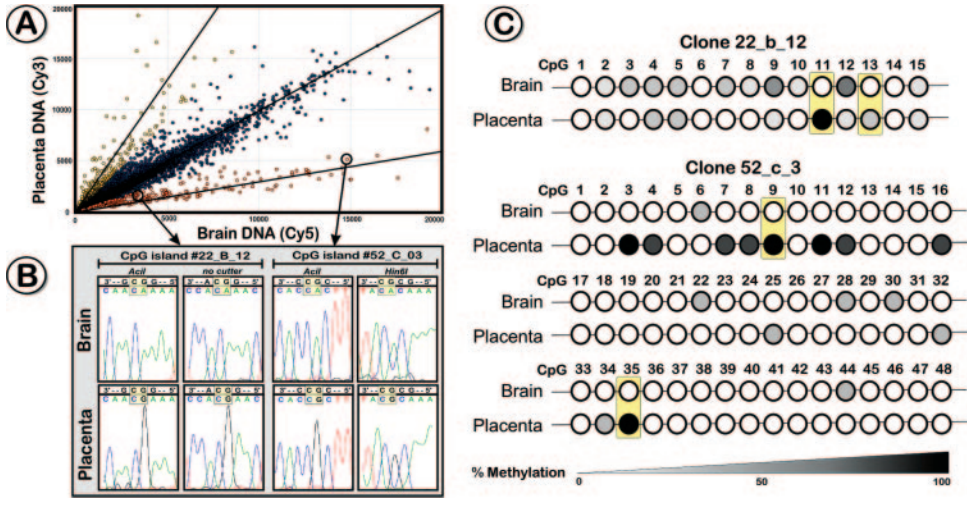


Figure 7. Examples of applications using a CpG island microarray. (**A**) Hybridization of the unmethylated fraction of placenta DNA and post mortem brain DNA to a CpG island array. Two pools of CpG island elements could be identified, which display extensively different methylation levels between these tissues (*Note*: some of the identified differences could be due to DNA sequence variation). (**B**) To validate the identified methylation differences, several CpG islands were subjected to bisulfite modification based mapping of methylated cytosines as exemplified for CpG island clones 22_B_12 (promoter region of Galectin-1) and 52_C_03 (promoter region of a brain-specific transcript, CR606704). The top sequence shows the reverse strand (−) of the original restriction sites, the bottom sequence displays the bisulfite-modified DNA. For each bisulfite-modified CpG-island, 8–10 clones were sequenced per tissue. Sequence 52_C_03 revealed several fully methylated CpG's in placenta, which were unmethylated in brain. In contrast, clone 22_B_12 showed subtler methylation differences (15–100%), depending on the position of CpG-dinucleotide. (**C**) Methylation patterns of clones 22B_12 and 52_C_03 derived from bisulfite sequencing of 10–12 clones per tissue. The yellow boxes indicate CpG dinucleotides that are shown in the sequenced graph (Figure 7B).

Table 2. Interindividual differences and distribution of the detected unmethylated sites with respect to the known genes as defined by the combined set of RefSeq and UCSC known genes for each brain DNA sample (M17–M25) and the merged map

Individual	3′-flanking	3′ter	5′-flanking	5′flanking–3′flanking	5′ter	Distal	Internal	Total	Site coverage (bp)
#M17 chr21/22	13/12	2/16	8/20	2/4	10/20	64/122	98/97	488	64943/134730
%Total	5.1	3.7	5.7	1.2	6.1	38.1	40.0		
#M18 chr21/22	17/22	9/15	13/29	3/3	16/28	95/191	134/152	727	98456/236797
%Total	5.4	3.3	5.8	0.8	6.1	39.3	39.3		
#M19 chr21/22	15/24	11/14	12/27	2/5	14/21	86/173	119/130	653	88290/221721
%Total	6.0	3.8	6.0	1.1	5.4	39.7	38.1		
#M21 chr21/22	20/24	12/18	15/29	2/5	14/22	102/184	143/157	747	109595/252347
%Total	5.9	4.0	5.9	0.9	4.8	38.3	40.2		
#M22 chr21/22	18/20	8/17	9/29	3/6	15/24	86/169	127/143	674	87604/213453
%Total	5.6	3.7	5.6	1.3	5.8	37.8	40.1		
#M23 chr21/22	12/15	4/13	10/25	2/3	10/21	68/150	101/111	545	70912/163322
%Total	5.0	3.1	6.4	0.9	5.7	40.0	38.9		
#M24 chr21/22	14/18	5/12	7/20	4/3	10/20	61/158	88/107	527	65639/187229
%Total	6.1	3.2	5.1	1.3	5.7	41.6	37.0		
#M25 chr21/22	17/15	7/13	10/18	3/3	9/22	65/171	102/97	552	69937/171073
%Total	5.8	3.6	5.1	1.1	5.6	42.8	36.1		
Merged chr21/22	26/28	13/22	19/36	4/9	19/34	142/237	187/201	977	152148/314374
%Total	5.5	3.6	5.6	1.3	5.4	38.8	39.7		

'5′ter' or '3′ter' refers to a 5′ or 3′ terminal site internal and within 1 kb of a gene boundary '5′flanking' or '3′flanking' refers to a site outside and within 5 kb of a gene boundary; 'internal' refers to an intronic site and 'distal' refers to an intergenic site outside of the −5 kb/+1 kb boundaries. A site can also be both 5′ and 3′ flanking in a gene rich region and referred as '5′flanking–3′flanking'.

between sites residing within introns of known genes and outside of the gene boundaries. Interestingly, while some genes, like *BCR*, showed a large number of sites inside the gene boundaries, some loci, like *C21ORF55* spanning ∼150 kb, were essentially devoid of internal unmethylated sites and in some cases, such as the *SIM2* locus, the unmethylated sites were limited to the first intron (Figure 8A–C). Such intragenic methylation may inhibit inappropriate transcriptional initiation at cryptic sites (35) or may serve as regulators of alternate transcripts as can be seen for *SIM2*. Overall, unmethylated sites detected in this study cover ∼0.47 Mb or ∼4% of the 12 Mb of non-repetitive sequences of chromosomes 21 and 22 interrogated in the combined map of all eight individuals with an average of 0.28 Mb (2.3%) in any given individual. Maps of the methylation patterns (average value of the eight tested individuals) of the q-arms of chromosome 21 and 22 are shown in Figure 9A–B. Detailed maps of all individuals for chromosome 21 and 22, linked to the UCSC Genome Browser (http://genome.ucsc.edu) are also available on our web-based methylation database (Web Recourses).

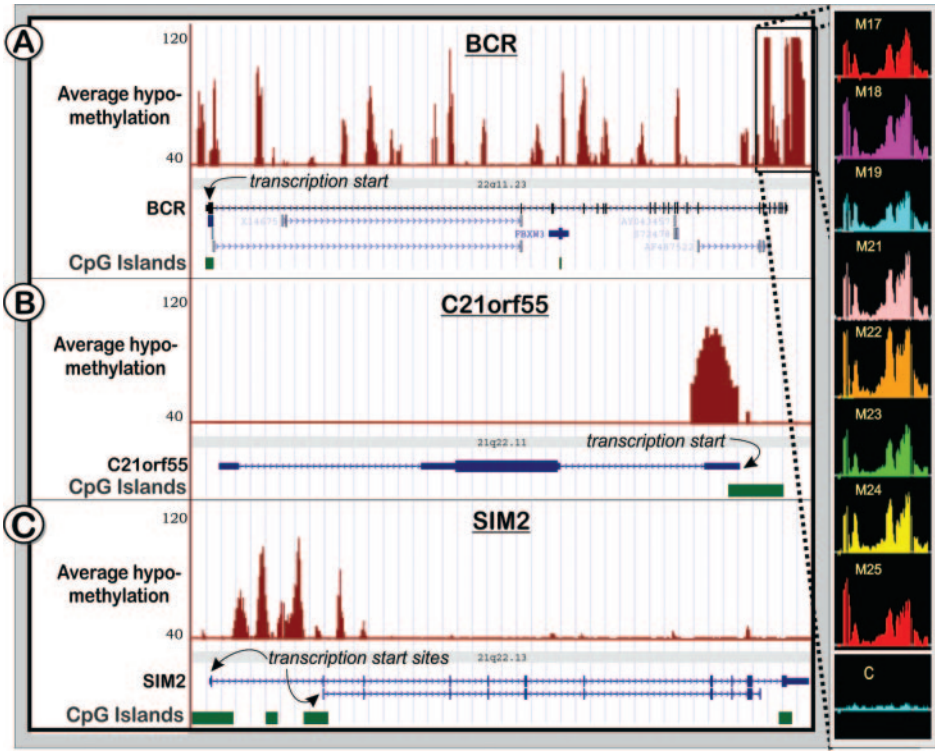


Figure 8. Profiles of unmethylated sites in three loci on human chromosomes 21 and 22 (501 bp window, Materials and Methods): *BCR* (**A**), *C21ORF55* (**B**) and *SIM2* (**C**) for human brain DNA (average of eight individuals, M17-M25). The graphs are based on *P*-values for each individual interrogation that show the significance of the enrichment in the unmethylated fraction versus total gDNA. The *P*-values were converted to the $(-10\ \log_{10})$ scale, such that, for example, *P*-value of 10^{-4} becomes 40. The vertical axes are adjusted to represent probes in the 40–120 range (*P*-values of 10^{-4}–10^{-12}), thus only probes that pass $P < 10^{-4}$ threshold are shown. Enlarged is a part of the chr 22q11.21 region (181 bp window), spanning breakpoints found in the generation of the two alternative forms of the Philadelphia chromosome translocation. C = gDNA control.

A comparison of the hypomethylation tracks with data from the Affymetrix transcriptome project (26,36) indicates that many of the unmethylated chromosomal regions overlap with mapped transcriptional active regions (Figure 9A–C, bottom tracks). These DNA methylation data complement existing studies on transcriptional activity and histone modifications on human chromosomes 21 and 22 (37). We found that in the majority of cases, specific histone modification patterns reported by Bernstein *et al*. (37) for the human hepatoma cell line HepG2 overlapped notably with the observed DNA methylation patterns. An example is shown in Figure 9C for the PEX26 gene that is ubiquitously transcribed in most tissues. The gene harbours an extensively unmethylated CpG rich region in its promoter. The comparison of the different epigenetic profiles of both studies shows that the same genomic region was also highly acetylated at Lysine 9 and 14 of histone 3 (H3), accompanied with H3 di- and trimethylation of Lysine 4. A comparison of histone modification tracks and our hypomethylation patterns for the q-arms of chromosome 21 and 22 revealed that H3 acetylation and Lys4 methylation usually correlated with unmethylated CpGs.

DISCUSSION

Microarray based technology for DNA modification analysis enables the highly parallel screening of numerous restriction fragments representing DNA methylation profiles over large segments of gDNA. Building on the principles described in earlier publications (11–23) our method addresses a series of critical issues and exhibits several advantages. An earlier method (18) used a sucrose gradient to enrich the unmethylated DNA fraction. This method, however, requires a large amount of DNA template and is rather imprecise in terms of the upper limit of the fragments that are subjected to hybridization. Other microarray methods for DNA methylation analysis can be categorized into three main classes which are based on: (i) identification of bisulfite induced C→T transitions (11–13,38,39), (ii) cleavage of gDNA by methylation-sensitive restriction enzymes and (iii) immunocapturing with antibodies against methylated cytosines. In the bisulfite arrays, each tested CpG is represented by a pair of either C(G) or T(A) nucleotides. The arrays contain oligonucleotides that measure the C(G)/T(A) ratio in the bisulfite treated DNA (corresponding to ^{met}C/C in the native DNA). Although informative and precise, these microarrays can contain only a limited number of oligonucleotides because treatment with bisulfite degenerates the 4 nt code, resulting in a loss of specificity for a large portion of the genome. For example, after bisulfite treatment all of the possible 16 permutations of a four base sequence containing unmethylated C and T (CCCC, CTCT, CCCT, CCTT, TCTC, TTTC, TTTT and so on) will become identical TTTT. The bisulfite method is also laborious and cannot be easily applied to profile a large set of samples. Furthermore, it is difficult to design suitable oligonucleotides that would

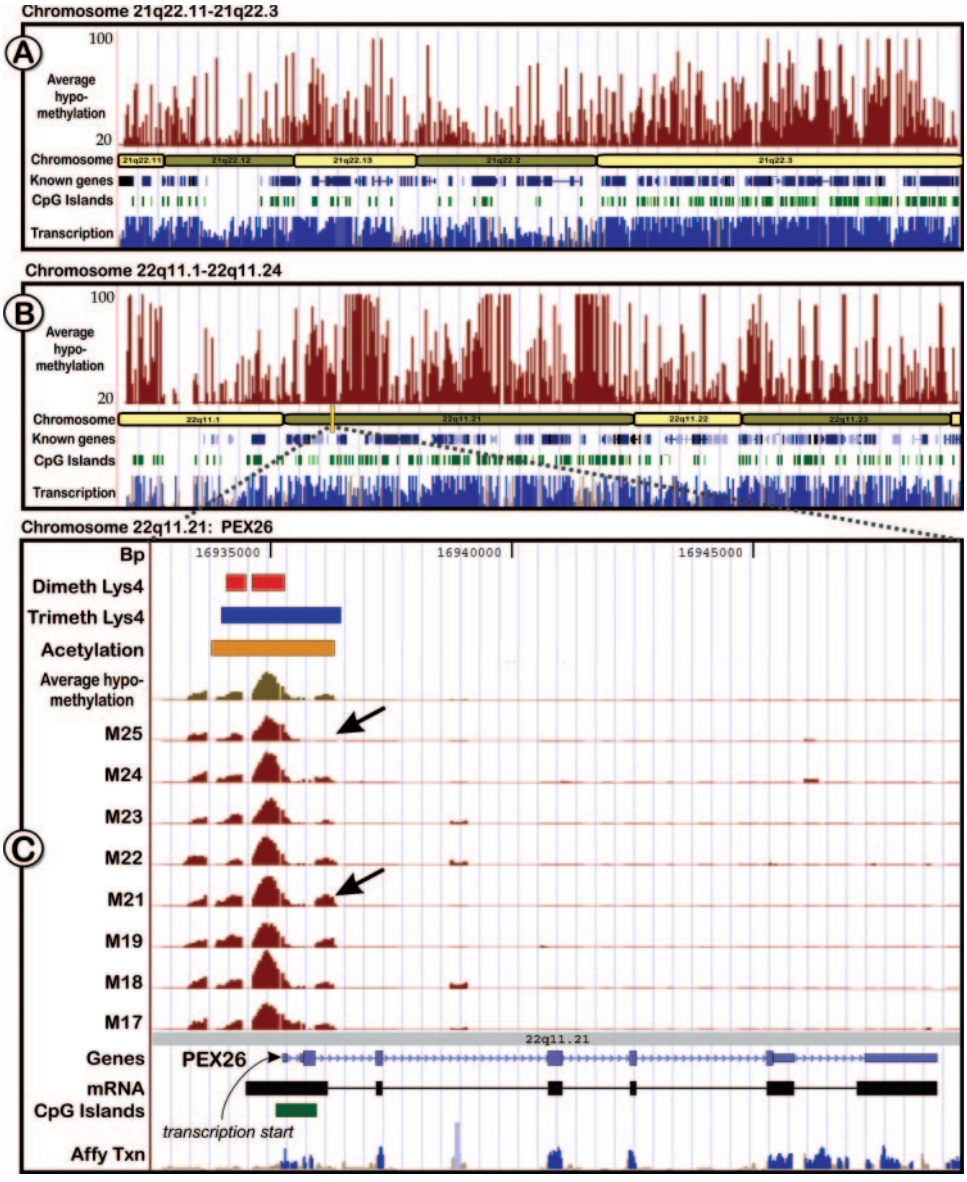


Figure 9. Genomic views showing unmethylated regions on chromosomes 21 and 22. (**A** and **B**): The top tracks (dark red) in the two chromosomal graphs shows the average amount of hypomethylation in the brain cortex of eight adult individuals. Also displayed are known genes (dark blue) and CpG islands (green). The bottom tracks display transcriptome data derived from 11 different tissues from the *Affymetrix* transcriptome phase 2 study (36). The track is coloured blue in areas that are thought to be transcribed at a statistically significant level. Regions that have a significant homology to other chromosomal regions or that overlap putative pseudogenes are coloured in lighter shades of blue. All other regions of the track are colored brown. (**C**) Enlarged is a part of chromosome 22q11.21, containing the peroxisome biogenesis factor 26 (*PEX26*, MIM 608666) that shows correlation between histone modifications and unmethylated DNA in its promoter region. The top three tracks represent histone modification data for H3 Lys4 dimethylation (orange bar), H3 Lys4 trimethylation (blue bar) and H3 Lys9/14 acetylation (yellow bar) (37). Underneath are the tracks for the average methylation patterns (unmethylated sites) observed in brain and the individual methylation patterns of all tested individuals (dark red). It is noteworthy that methylation patterns exhibit some interindividual differences (indicated by arrows).

exhibit similar melting temperatures since the specificity of base discrimination varies considerably (12). Using our approach, arrays can contain an almost unlimited number of oligonucleotides: coverage can range from individual genes to entire chromosomes represented by millions of oligonucleotides on glass chips. Whole genome tiling arrays are already available for *Arabidopsis thaliana* and *Escherichia coli*, and will soon be available for the entire human genome.

Restriction enzyme based methods are used to enrich either the hypermethylated or unmethylated fraction of gDNA. Methods relying on the enrichment and detection of hypermethylated DNA have predominantly been used to identify abnormally methylated CpG islands in malignant cells (15–17,31). Although this strategy seems to be useful for detecting major epigenetic changes in some regions of the genome, the overall proportion of interrogated CpG sites is substantially lower compared with that achieved using approaches based on the analysis of the unmethylated fraction. As shown in Results, we have estimated that interrogation of the unmethylated fraction of gDNA could be up to several hundred

folds more efficient than analysing the hypermethylated fraction. Furthermore, since unmethylated cytosines are less abundant in the genome than methylated cytosines (depending on the tissue, 70–90% of cytosines are methylated), analysis of the smaller unmethylated fraction of gDNA is more sensitive to detect subtle changes. For example, an increase of 10% from the normal density of ^{met}C would result in a 100% (from 20 to 10%) difference in the unmethylated fraction, but only a 12% (from 80 to 90%) difference in the hypermethylated fraction of gDNA. The unmethylated fraction has been used in some approaches employing class II microarray methods, for instance by using the methylation-specific McrBC enzyme (23) to deplete the hypermethylated fraction. However, the remaining unmethylated DNA fragments (>1 kb) have to be gel-purified, requiring large amounts of starting material. Additionally, the McrBC method may not be able to differentiate between dense and sparse methylation within relatively short DNA fragments. For example, the 2 kb human *COMT* promoter region, which contains 27 McrBC target sites, can be cut to shorter than 1 kb fragments in cases where there are 2 (7%) or 27 (100%) methylated McrBC sites. Furthermore, the McrBC method cannot differentiate between unmethylated and polymorphic cytosines. Another method to enrich the unmethylated fraction uses the rare cutter NotI (5′-GCGGCCGC-3′) (19–21). However, NotI sites are not well represented in the genome and will only provide a very superficial overview of genomic methylation patterns. An alternative to these methods is the use of antibodies specific for methylated cytosines [MeDIP (22)]. In this method, antibodies are used to immunocapture methylated genomic fragments. However, this approach requires large amounts of gDNA (>8 μg) and also relies on the enrichment of the less informative hypermethylated fraction of the genome.

In our analyses, we have addressed another important issue: the interference of DNA polymorphisms that may simulate DNA modification differences across individuals. Data from the SNP consortium indicate that roughly every 360th nucleotide in the human genome represents an SNP. In humans, ∼2.16 million SNPs are detected in CpG dinucleotides, and such CpG SNPs are 6.7-fold more abundant than expected (40). Depending on the restriction enzyme combination, our CpG island array-based studies demonstrated that 10–30% of all outliers initially detected as methylation differences contained SNPs (Figure 4). Information on the SNPs and other polymorphisms such as deletions, inversions or duplications within the restriction sites of the enzymes used for the enrichment of the unmethylated or hypermethylated fractions is helpful in differentiating the epigenetic variations from the DNA sequence ones. To minimize the effects of DNA polymorphisms, it may be also beneficial to compare affected tissue and healthy cells from the same individual.

Another advantage of PCR-based methylation profiling methods is the ability to work with limited DNA resources. Although our basic protocol requires about 500 ng of gDNA, the amount of template DNA can be significantly lower. In our recent experiments, methylation patterns at the *COMT* region generated from a relatively small number of Jurkat tissue culture cells (up to 500 cells, or 3 ng of DNA) did not reveal any significant differences when compared with the methylation patterns generated from a substantially larger number of cells from the same tissue.

There are, however, also some of limitations to the technology described in this article. The methylation sensitive restriction enzymes do not interrogate every cytosine, and with our current design, more than half of CpG sites remain uninterrogated. This may be critical when the phenotypic outcomes are determined by a methylation change at an isolated cytosine that is not within the restriction site of a methylation sensitive restriction enzyme. This problem may be partially overcome by the application of the same arrays to the CpG specific immunoprecipitation technique (MeDIP) (22) in addition to histone modification analysis through ChIP technology, which identifies DNA sequences associated with modified histones (10). DNA and histone modifications seem to be inter-dependent, and consequently the possibility of a combined approach that interrogates both DNA methylation and chromatin modification in parallel might be a productive approach to the fine mapping of epigenetic changes. Also, asymmetrical ^{m}C sites (CpNpN) that are found in plants and some fungi such as *Neurospora crassa* are difficult to detect, although some methylation-sensitive type IIs restriction enzymes are available (e.g. Esp3I or BveI). However, methylation of asymmetrical sites in animal organisms is not common. Additionally, this array method can also be modified for analysis of methylated adenines in plants and bacteria.

In summary, the use of microarrays targeted at unmethylated cytosines is a high-throughput approach to profile DNA methylation patterns across the genome. The ability to analyse minute amounts of DNA may enable the epigenetic screening of DNA in plasma, serum or other body fluids, as well as in prenatal diagnostics. Although all the examples provided in this work investigated human DNA, the same strategies can be used for the epigenetic analyses of numerous other species. It is evident that epigenetic profiling should be performed in a systematic, unbiased fashion and not limited to the traditionally preferable regions such as CpG islands. Outside of CpG islands, numerous other genomic loci exist that may be sites for important epigenetic modification, including enhancers, imprinting control elements (41) or the regions that encode regulatory RNA elements.

The above described technology, in combination with existing epigenetic profiling methods, may help to identify inter-individual variation in genome-wide methylation patterns as well as epigenetic changes that arise during tissue differentiation and the understanding of the epigenetic effects of various environmental factors. Of particular interest is the application of high-throughput DNA methylation analyses to address the molecular basis of various non-Mendelian irregularities of complex diseases, such as discordance of monozygotic twins, remissions and relapses of a disease, parent of origin- and sex-effects, and tissue- and site-specificity (42). Further technological developments may include building high-resolution oligonucleotide-based microarrays spanning the entire human genome, improving the enrichment strategies through the application of more specialized methylation sensitive restriction enzymes, and substantial reduction in the amount of initial template DNA down to the amount of a haploid or diploid genome. All these developments will provide the basis for identifying the methylation profile

of the entire genome in a single cell, one of the 'quantum leaps' in post-genomic biology (43).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This research has been supported by the Special Initiative grant from the Ontario Mental Health Foundation, and also by NARSAD, CIHR, NIH, the Stanley Foundation, and the Crohn's and Colitis Foundation of Canada. A.S. holds a CIHR Michael Smith Fellowship. Funding to pay the Open Access publication charges for this article was provided by the Ontario Mental Health Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Bird,A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
2. Henikoff,S. and Matzke,M.A. (1997) Exploring and explaining epigenetic effects. *Trends Genet*., **13**, 293–295.
3. Wolffe,A.P. and Matzke,M.A. (1999) Epigenetics: regulation through repression. *Science*, **286**, 481–486.
4. Reik,W., Dean,W. and Walter,J. (2001) Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–1093.
5. Grewal,S.I. and Moazed,D. (2003) Heterochromatin and epigenetic control of gene expression. *Science*, **301**, 798–802.
6. Suter,C.M., Martin,D.I. and Ward,R.L. (2004) Germline epimutation of MLH1 in individuals with multiple cancers. *Nature Genet*., **36**, 497–501.
7. Walter,J. and Paulsen,M. (2003) Imprinting and disease. *Semin. Cell Dev. Biol*., **14**, 101–110.
8. Laird,P.W. (2003) The power and the promise of DNA methylation markers. *Nature Rev. Cancer*, **3**, 253–266.
9. Frommer,M., McDonald,L.E., Millar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
10. van Steensel,B. and Henikoff,S. (2003) Epigenomic profiling using microarrays. *Biotechniques*, **35**, 346–350, 352–344, 356–347.
11. Adorjan,P., Distler,J., Lipscher,E., Model,F., Muller,J., Pelet,C., Braun,A., Florl,A.R., Gutig,D., Grabs,G. *et al*. (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res*., **30**, e21.
12. Balog,R.P., de Souza,Y.E., Tang,H.M., DeMasellis,G.M., Gao,B., Avila,A., Gaban,D.J., Mittelman,D., Minna,J.D., Luebke,K.J. *et al*. (2002) Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal Biochem*., **309**, 301–310.
13. Gitan,R.S., Shi,H., Chen,C.M., Yan,P.S. and Huang,T.H. (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res*., **12**, 158–164.
14. Hatada,I., Kato,A., Morita,S., Obata,Y., Nagaoka,K., Sakurada,A., Sato,M., Horii,A., Tsujimoto,A. and Matsubara,K. (2002) A microarray-based method for detecting methylated loci. *J. Hum. Genet*., **47**, 448–451.
15. Shi,H., Wei,S.H., Leu,Y.W., Rahmatpanah,F., Liu,J.C., Yan,P.S., Nephew,K.P. and Huang,T.H. (2003) Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. *Cancer Res*., **63**, 2164–2171.
16. Yan,P.S., Efferth,T., Chen,H.L., Lin,J., Rodel,F., Fuzesi,L. and Huang,T.H. (2002) Use of CpG island microarrays to identify colorectal tumors with a high degree of concurrent methylation. *Methods*, **27**, 162–169.
17. 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.
18. Tompa,R., McCallum,C.M., Delrow,J., Henikoff,J.G., van Steensel,B. and Henikoff,S. (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol*., **12**, 65–68.
19. Li,J., Protopopov,A., Wang,F., Senchenko,V., Petushkov,V., Vorontsova,O., Petrenko,L., Zabarovska,V., Muravenko,O., Braga,E. *et al*. (2002) NotI subtraction and NotI-specific microarrays to detect copy number and methylation changes in whole genomes. *Proc. Natl Acad. Sci. USA*, **99**, 10724–10729.
20. Yamamoto,F. and Yamamoto,M. (2004) A DNA microarray-based methylation-sensitive (MS)-AFLP hybridization method for genetic and epigenetic analyses. *Mol. Genet. Genomics*, **271**, 678–686.
21. Ching,T.T., Maunakea,A.K., Jun,P., Hong,C., Zardo,G., Pinkel,D., Albertson,D.G., Fridlyand,J., Mao,J.H., Shchors,K. *et al*. (2005) Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. *Nature Genet*., **37**, 645–651.
22. 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. *Nature Genet*., **37**, 853–862.
23. Lippman,Z., Gendrel,A.-V., Colot,V. and Martienssen,R. (2005) Profiling DNA methylation patterns using genomic tiling microarrays. *Nature Methods*, **2**, 219–224.
24. Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nature Genet*., **6**, 236–244.
25. Heisler,L.E., Torti,D., Boutros,P.C., Watson,J., Chan,C., Winegarden,N., Takahashi,M., Yau,P., Huang,T.H., Farnham,P.J. *et al*. (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.
26. Kapranov,P., Cawley,S.E., Drenkow,J., Bekiranov,S., Strausberg,R.L., Fodor,S.P. and Gingeras,T.R. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, **296**, 916–919.
27. Cawley,S., Bekiranov,S., Ng,H.H., Kapranov,P., Sekinger,E.A., Kampa,D., Piccolboni,A., Sementchenko,V., Cheng,J., Williams,A.J. *et al*. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.
28. Hajkova,P., el-Maarri,O., Engemann,S., Oswald,J., Olek,A. and Walter,J. (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol. Biol*., **200**, 143–154.
29. Sutherland,E., Coe,L. and Raleigh,E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol*., **225**, 327–348.
30. Kruger,T., Wild,C. and Noyer-Weidner,M. (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. *EMBO J*., **14**, 2661–2669.
31. Yan,P.S., Chen,C.M., Shi,H., Rahmatpanah,F., Wei,S.H. and Huang,T.H. (2002) Applications of CpG island microarrays for high-throughput analysis of DNA methylation. *J. Nutr*., **132**, 2430S–2434S.
32. Chen,C.M., Chen,H.L., Hsiau,T.H., Hsiau,A.H., Shi,H., Brock,G.J., Wei,S.H., Caldwell,C.W., Yan,P.S. and Huang,T.H. (2003) Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes. *Am. J. Pathol*., **163**, 37–45.
33. Mathieu-Daude,F., Welsh,J., Vogt,T. and McClelland,M. (1996) DNA rehybridization during PCR: the 'Cot effect' and its consequences. *Nucleic Acids Res*., **24**, 2080–2086.
34. Kampa,D., Cheng,J., Kapranov,P., Yamanaka,M., Brubaker,S., Cawley,S., Drenkow,J., Piccolboni,A., Bekiranov,S., Helt,G. *et al*. (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res*., **14**, 331–342.
35. Bird,A.P. (1995) Gene number, noise reduction and biological complexity. *Trends Genet*., **11**, 94–100.
36. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al*. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
37. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J.,III, Gingeras,T.R. *et al*. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.

38. Shi,H., Maier,S., Nimmrich,I., Yan,P.S., Caldwell,C.W., Olek,A. and Huang,T.H. (2003) Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J. Cell Biochem.*, **88**, 138–143.
39. Hou,P., Ji,M., Liu,Z., Shen,J., Cheng,L., He,N. and Lu,Z. (2003) A microarray to analyze methylation patterns of p16(Ink4a) gene 5′-CpG islands. *Clin. Biochem.*, **36**, 197–202.
40. Tomso,D.J. and Bell,D.A. (2003) Sequence context at human single nucleotide polymorphisms: overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. *J. Mol. Biol.*, **327**, 303–308.
41. Schumacher,A., Buiting,K., Zeschnigk,M., Doerfler,W. and Horsthemke,B. (1998) Methylation analysis of the PWS/AS region does not support an enhancer-competition model. *Nature Genet.*, **19**, 324–325.
42. Petronis,A. (2001) Human morbid genetics revisited: relevance of epigenetics. *Trends Genet.*, **17**, 142–146.
43. Collins,F.S., Green,E.D., Guttmacher,A.E. and Guyer,M.S. (2003) A vision for the future of genomics research. *Nature*, **422**, 835–847.