

CpG Island Hypermethylation in Human Astrocytomas

Xiwei Wu², Tibor A. Rauch³, Xueyan Zhong¹, William P. Bennett¹, Farida Latif⁴, Dietmar Krex⁵, and Gerd P. Pfeifer¹

Abstract

Astrocytomas are common and lethal human brain tumors. We have analyzed the methylation status of over 28,000 CpG islands and 18,000 promoters in normal human brain and in astrocytomas of various grades using the methylated CpG island recovery assay. We identified 6,000 to 7,000 methylated CpG islands in normal human brain. Approximately 5% of the promoter-associated CpG islands in the normal brain are methylated. Promoter CpG island methylation is inversely correlated whereas intragenic methylation is directly correlated with gene expression levels in brain tissue. In astrocytomas, several hundred CpG islands undergo specific hypermethylation relative to normal brain with 428 methylation peaks common to more than 25% of the tumors. Genes involved in brain development and neuronal differentiation, such as *BMP4*, *POU4F3*, *GDNF*, *OTX2*, *NEFM*, *CNTN4*, *OTP*, *SIMI*, *FYN*, *EN1*, *CHAT*, *GSX2*, *NKX6-1*, *PAX6*, *RAX*, and *DLX2*, were strongly enriched among genes frequently methylated in tumors. There was an overrepresentation of homeobox genes and 31% of the most commonly methylated genes represent targets of the Polycomb complex. We identified several chromosomal loci in which many (sometimes more than 20) consecutive CpG islands were hypermethylated in tumors. Seven such loci were near homeobox genes, including the *HOXC* and *HOXD* clusters, and the *BARHL2*, *DLX1*, and *PITX2* genes. Two other clusters of hypermethylated islands were at sequences of recent gene duplication events. Our analysis offers mechanistic insights into brain neoplasia suggesting that methylation of the genes involved in neuronal differentiation, in cooperation with other oncogenic events, may shift the balance from regulated differentiation towards gliomagenesis. *Cancer Res*; 70(7); 2718–27. ©2010 AACR.

Introduction

Human solid tumors are characterized by a broad spectrum of genetic and epigenetic alterations. Astrocytomas are the most common type of human gliomas, and are primary malignancies in the central nervous system. These tumors are classified by varying degrees of malignancy with grade 4 astrocytomas (glioblastomas) being the most aggressive ones. Animal models and recent progress in stem cell research have now established that dividing cells in the subventricular zone give rise to astrocytomas (1). Human cancer genome sequencing studies have catalogued characteristic mutations in glioblastomas. These studies have uncovered

that individual tumors generally have ~30 to 40 mutations in gene coding sequences that alter the amino acid composition of a protein (2). The most commonly mutated or homozygously deleted genes include *CDKN2*, *TP53*, *EGFR*, *PTEN*, *NFI*, *PIK3CA*, and *IDHI*.

At the epigenetic level, tumor specimens are most often characterized in terms of altered DNA methylation (3). Aberrations in DNA methylation patterns may have critical effects on tumor initiation and progression (4). Both DNA hypermethylation of specific genes and DNA hypomethylation commonly affecting repetitive DNA are observed in brain cancers (5–11). CpG island hypermethylation can affect a variety of different genes, ranging from regulators of cell cycle progression and apoptosis to DNA repair genes and developmental transcription factors. With regard to the latter class of genes, it has recently been recognized that a large fraction of the genes commonly methylated in human tumors represent targets of the Polycomb complex in human embryonic stem cells (12–17). For example, in lung cancers, ~80% of the most frequently methylated genes are marked by the Polycomb histone modification, histone H3 lysine 27 trimethylation (H3K27me₃; ref. 18). These findings have raised important questions as to the mechanisms that might be involved in such specific methylation targeting, as well as to the generality of the phenomenon in different histologic types of human tumors and in terms of causality for tumorigenesis.

In the present study, we have used a comprehensive methylation profiling technique, developed in our laboratory, and termed “methylated-CpG island recovery assay” or MIRA

Authors' Affiliations: Departments of ¹Cancer Biology and ²Molecular Medicine, Beckman Research Institute of the City of Hope, Duarte, California; ³Section of Molecular Medicine, Department of Orthopedic Surgery, Rush University Medical Center, Chicago, Illinois; ⁴Section of Medical and Molecular Genetics, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham, United Kingdom; and ⁵Klinik und Poliklinik für Neurochirurgie, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

X. Wu and T.A. Rauch contributed equally to this work.

Corresponding Author: Gerd P. Pfeifer, Department of Cancer Biology, Beckman Research Institute, City of Hope, 1500 East Duarte Road, Duarte, CA 91010. Phone: 626-301-8853; Fax: 626-358-7703; E-mail: gpfeifer@coh.org.

doi: 10.1158/0008-5472.CAN-09-3631

©2010 American Association for Cancer Research.

(12). This assay was used in conjunction with CpG island and promoter microarrays (MIRA-chip) to characterize the CpG island methylome in normal human brain tissue and brain tumors (astrocytomas) representing different stages of tumor progression and malignancy.

Materials and Methods

Tissue and DNA samples. Six normal human brain tissues from accident victims were obtained from Capital Biosciences and BioChain. Twenty-four brain tumor tissues from astrocytomas (WHO grades 1–4) were obtained on Institutional Review Board–approved protocols from randomly selected patients who had brain tumor surgery at the Department of Neurosurgery at the University Hospital in Dresden during 1997 to 2007. Histologic diagnosis was performed by a local neuropathologist and confirmed by reference neuropathology. These tumors were labeled T1 (1–6), T2 (1–6), T3 (1–6), and T4 (1–6), in which the number immediately following the T reflects tumor grade. Six tumors labeled T2–7, T3 (7–9), and T4 (7–8) were astrocytomas obtained from Asterand. No patient had chemotherapy before tissue sampling.

MIRA and microarray hybridization. Tumor and normal brain DNA was fragmented by sonication to ~500 bp average size on agarose gels. Enrichment of the methylated double-stranded DNA fraction by MIRA was performed as described previously (19, 20). The labeling of amplicons, microarray hybridization, and scanning were performed according to the NimbleGen protocol. NimbleGen tiling arrays were used for hybridization (385K Human CpG Island plus Promoter arrays). The single array design covers all 28,226 University of California, Santa Cruz (UCSC) Genome Browser–annotated CpG islands and the promoter regions for all RefSeq genes. The promoter region covered is 1 kb (–800 to +200 relative to the transcription start sites). For all samples, the MIRA-enriched DNA was compared with the input DNA.

Identification and annotation of methylated regions. Log 2 ratio data were converted into *P* value scores using the Kolmogorov-Smirnov test with a 750 bp window by using the NimbleScan software [$SP = -\log_{10}(p)$]. Probes were selected as positive if their *P* value scores were above 2 ($P < 0.01$). For our analysis, we defined a methylated region of interest (methylation peak) as a region with at least four consecutive positive probes (i.e., no gaps) covering a minimum length of 350 bp. This stringent definition will give few false-positive results. Identified methylation peaks were mapped relative to known transcripts defined in the UCSC Genome Browser HG18 RefSeq database. Methylation peaks falling into 1,000 bp relative to transcription start sites were defined as 5'-end peaks; methylation peaks falling within 1,000 bp of RefSeq transcript end sites were defined as 3'-end peaks, and those falling within gene bodies (from 1,000 bp downstream of transcription start to 1,000 bp upstream of transcript end) were defined as "intragenic" peaks. Methylation peaks that were not close to any known transcripts were defined as "intergenic."

Identification of tumor-specific methylated regions. Methylation peaks in tumor samples were identified as above, and methylation peaks in normal samples were identified

using a $P \geq 1.7$ score to exclude regions with weak methylation. The methylated regions in tumor and normal samples were matched by outer join and their methylation status was determined by the presence of methylation peaks in each sample in these matched regions. Only the regions methylated in more than 25% of the 30 tumors and unmethylated in all of the six normal samples were considered as candidates. The average *P* value scores of probes within these candidates in tumor samples were compared with the average *P* value scores of probes in normal samples, and the final tumor-specific methylation peaks were selected if the difference was more than 1. The tumor-specific methylation peaks were also mapped to the 28,226 CpG islands defined as such at the UCSC Genome Browser. The CpG islands are defined as having a G/C content of >50%, CpG observed to expected ratio of >0.6 and a length of at least 200 bp. These CpG islands were also mapped relative to known transcripts in the HG18 RefSeq database using the same approach described above. The data reported in this article has been deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession no. GSE19391).

DNA methylation analysis using sodium bisulfite–based methods. DNA was treated and purified with the EpiTect Bisulfite kit (Qiagen). PCR primer sequences for amplification of specific gene targets in bisulfite-treated DNA are available upon request. The PCR products were analyzed by combined bisulfite restriction analysis (COBRA) as described previously (21). Additionally, PCR products from bisulfite-converted DNA were cloned into the pDrive PCR cloning vector (Qiagen), and 8 to 12 individual clones were sequenced.

Results

Analysis of DNA methylation in normal brain. MIRA-chip has proven to be a sensitive, robust, and reproducible technique for mapping DNA methylation patterns in mammalian genomes including cancer tissues (12, 14, 20). Six normal brain tissue DNA samples from accident victims were used to determine patterns of methylation at CpG islands and promoters using NimbleGen tiling arrays. After enrichment of the methylated DNA fractions by MIRA, MIRA and input fractions were mixed and hybridized to the arrays, and data were normalized and analyzed as described in Materials and Methods. We found a surprisingly large number of methylated CpG islands in normal human brain ranging from 6,000 to 7,000 in each sample. These numbers correspond to ~21% to 25% of all CpG islands as defined above.

We then analyzed the localization of these methylation peaks relative to known genes. We found that 9.3% of the methylation peaks were at the 5'-ends of genes (defined as $\pm 1,000$ bp relative to transcription start sites), 53.8% were within gene bodies (defined as within 1,000 bp downstream of the transcription start site and 1,000 bp upstream of the gene end of RefSeq genes), and 9.1% were near the 3'-end of genes (defined as $\pm 1,000$ bp of the gene end); 27.8% of the methylation peaks were not associated with genes (Fig. 1A).

Next, we focused on those genes with methylation peaks at the 5'-end and explored the relationship of this 5' gene end

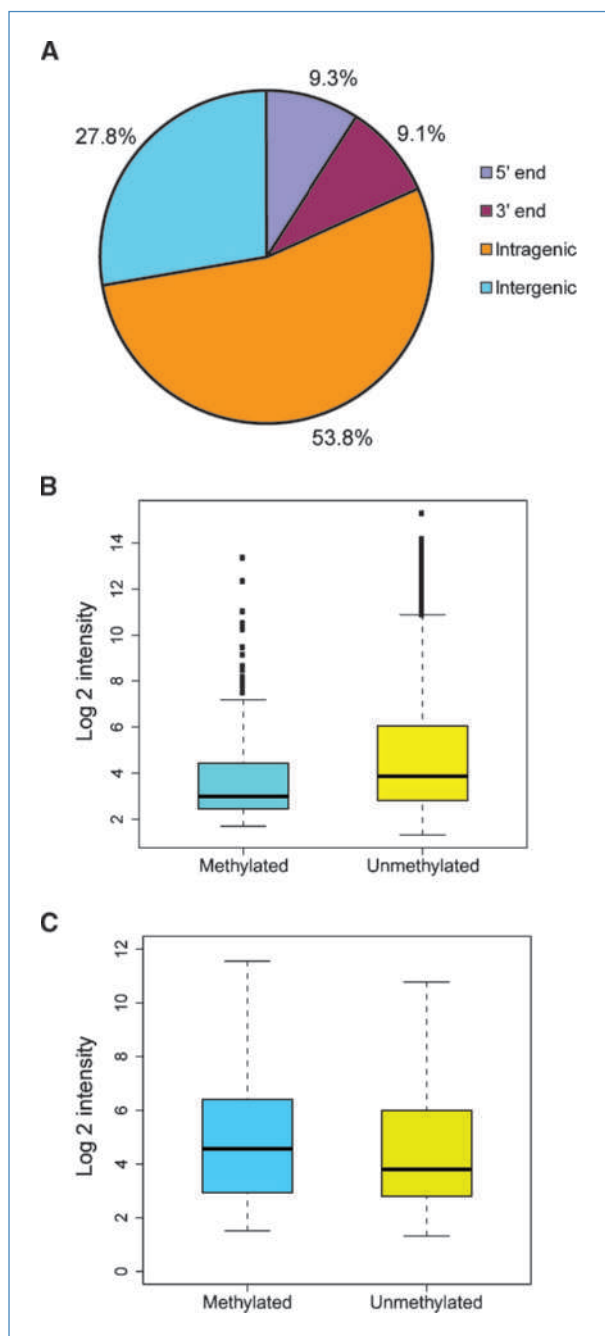


Figure 1. Localization of methylated CpG islands in normal human brain relative to known genes and correlation between gene methylation and gene expression levels. A, localization of the methylation peaks in normal brain relative to gene position. Each normal brain sample had between 6,000 and 7,000 methylation peaks. B, correlation between 5' gene end methylation in normal brain and gene expression levels. Public microarray gene expression data of whole brain tissue were obtained from the GNF SymAtlas database. Genes with methylation peaks at their 5'-end are generally expressed at lower level than genes without methylation peaks at the 5'-end (Kolmogorov-Smirnov test, $P = 1.8e-10$). C, correlation between gene body methylation in normal brain and gene expression levels. Genes with intragenic methylation peaks are generally expressed at a higher level than genes without intragenic methylation (Kolmogorov-Smirnov test, $P < 2.2e-16$).

methylation and gene expression patterns using public microarray gene expression data of whole brain tissue from the GNF SymAtlas database. There were 456 genes with 5'-end methylation in all normal samples (Supplementary Table S1). We found that genes with methylation peaks at their 5'-end are generally expressed at lower levels than genes without methylation peaks at the 5'-end, and the difference is highly significant (Kolmogorov-Smirnov test, $P = 1.8e-10$; Fig. 1B). Using the Database for Annotation, Visualization, and Integrated Discovery (version 6), we obtained annotation information for 377 of these genes. Functional annotation analysis showed that genes involved in spermatogenesis are highly enriched (Benjamini-corrected, $P = 7.28e-04$). No other gene ontology categories had corrected P values of <0.05 .

We then looked at the genes with methylation peaks within the gene body. There were 1,768 genes with gene body methylation and without 5'-end methylation. Consistent with previous findings in B cells (20, 22), we observed that genes with gene body methylation are generally expressed at higher levels than genes without gene body methylation (Fig. 1C).

Analysis of DNA methylation in brain tumors shows frequent methylation of genes involved in neuronal differentiation. To characterize methylation patterns in human brain tumors, we performed MIRA-chip analysis on 30 brain tumors including six grade 1, seven grade 2, nine grade 3, and eight grade 4 astrocytomas (glioblastomas). To confirm that the microarray data correctly reflect true methylation differences, we analyzed several genes including *BCL2L1*, *RAX*, *FOXP4*, and *ZNF274* by sodium bisulfite-based DNA methylation analysis techniques. Consistent with previous data demonstrating the reliability of the MIRA assay (12, 14, 20, 23), we confirmed methylation differences between normal and tumor tissues by COBRA analysis and by bisulfite sequencing, as shown for the proapoptotic gene *BCL2L1* also known as *BIM* (Fig. 2), for the *RAX* and *ZNF274* genes (Supplementary Figs. S1 and S2), and for *FOXP4* (data not shown). The *BCL2L1* gene was methylated in 21 of the 30 brain tumors and mostly in tumors of grade 2 and higher (Fig. 2).

We catalogued those genes that are methylated in the brain tumors and defined a tumor-specifically methylated gene as one that was not methylated in any of the six normal brain tissue DNAs. Using these criteria, we determined the number of tumor-methylated regions in each tumor as shown in Fig. 3A. The number of tumor-specifically methylated regions increased in astrocytomas going from grade 1 to grade 3. However, this number, on average, decreased in grade 4 tumors relative to grade 2 and 3 tumors. Four of the normal brain tissues available to us were from the frontal lobe, one was from the occipital lobe, and one from an unspecified region of the brain. Among these six normal brain tissues, there were very few differences in terms of number and location of methylated CpG islands. Astrocytomas came from several different regions of the brain including frontal, parietal, occipital, fronto-temporal, and temporal lobes, and three of the stage I tumors were from the cerebellum. When looking at location versus number of methylation targets adjusting for tumor stage, there was no significant correlation,

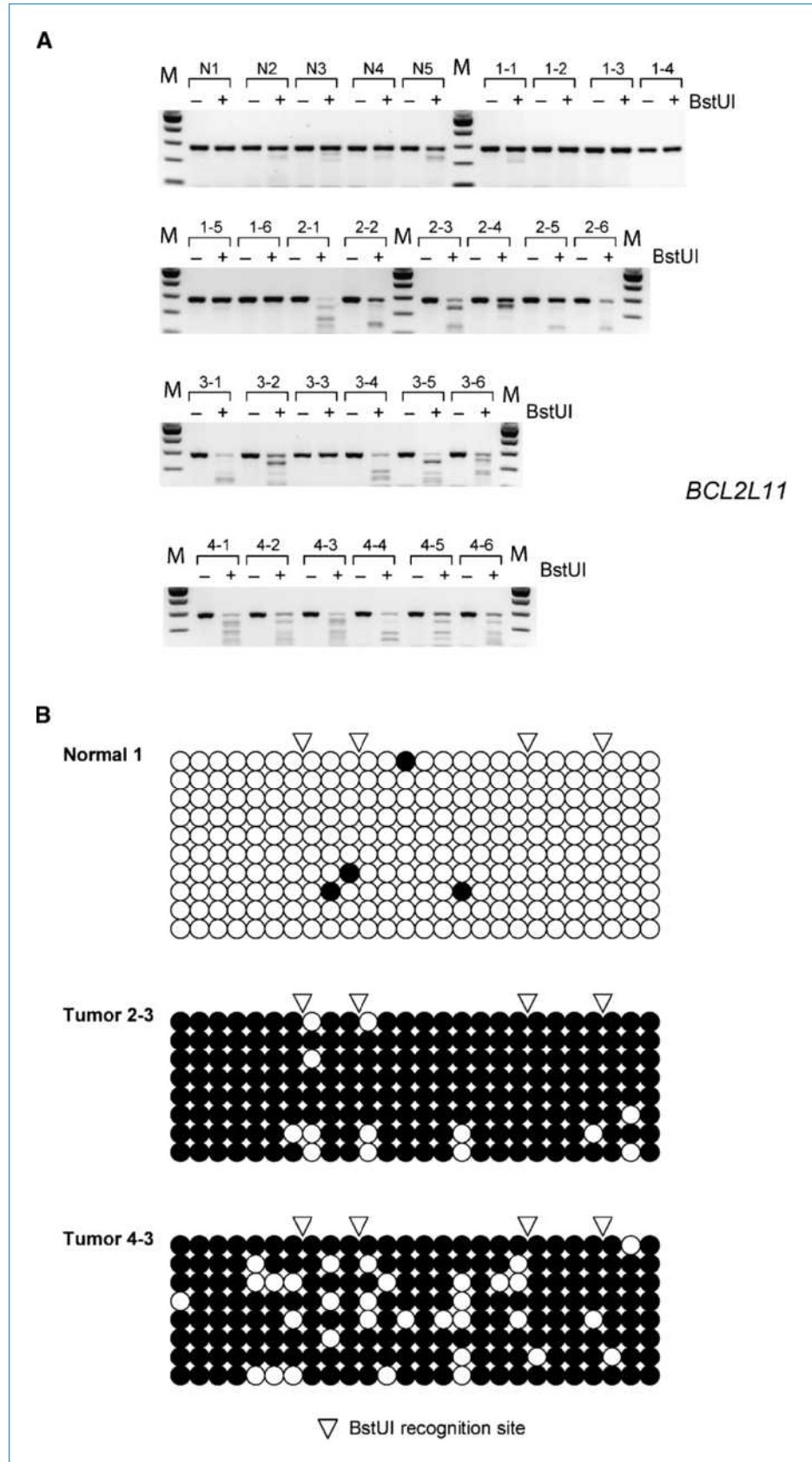


Figure 2. Confirmation of tumor-associated DNA methylation by COBRA and sodium bisulfite sequencing. The CpG island associated with the *BCL2L11* gene was analyzed in normal brain DNA (N) and in brain tumors (numbered). A, COBRA assays. In the COBRA assay, cleavage by *Bst*UI indicates methylation at 5'CGCG sequences. B, bisulfite sequencing; open circles, unmethylated CpG sites; closed circles, methylated CpG sites; triangles, *Bst*UI sites.

Downloaded from <http://aacrjournals.org/cancerres/article-pdf/70/7/2718/2646493/2718.pdf> by guest on 24 June 2024

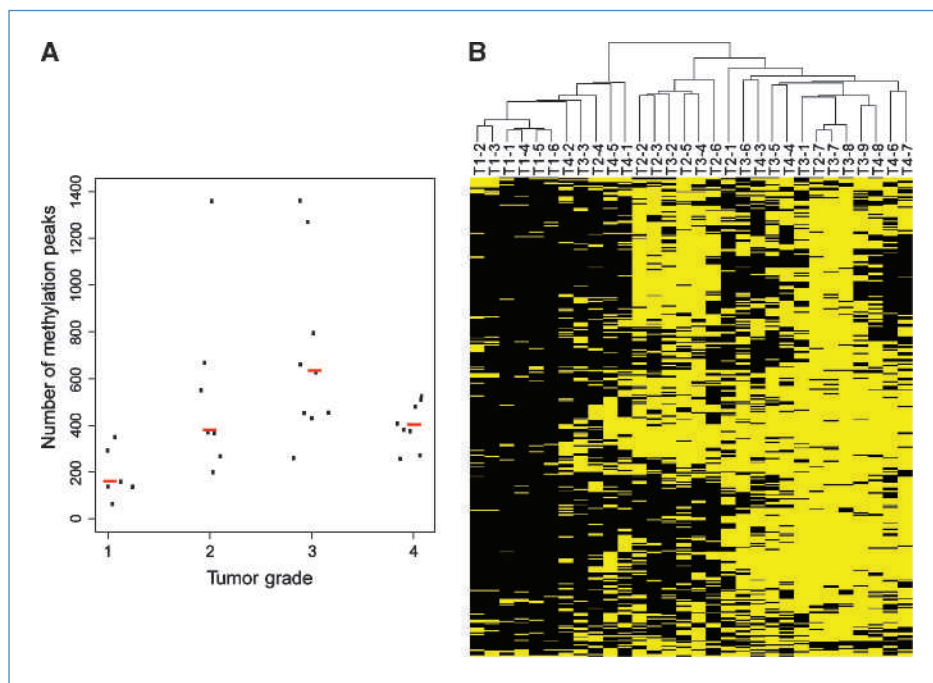


Figure 3. Tumor-specifically methylated genes and tumor grade. A, the number of methylated regions is plotted according to tumor grade. Tumor-specific methylation peaks were identified. These peaks were not found in any of the six normal samples. The red line represents the median number in each grade. B, cluster analysis. Methylation peaks commonly found in tumor samples but not in normal samples were identified (see text). Then the data were subjected to hierarchical clustering with Euclidean distance and average linkage method using Cluster v3.0 and visualized in Java TreeView. Yellow, methylated state; black, unmethylated state.

but the numbers of tumors are too small to exclude the possibility of brain region-specific differences in tumor-associated methylation.

We then established a list of genes that are most commonly methylated across different tumor specimens. These genes are unmethylated, i.e., lack a peak on the arrays in all six normal brain samples, and are methylated in at least 25% of the tumors (i.e., in at least 8 out of 30 tumors). We compiled a list of 428 methylation peaks that fit these criteria (Supplementary Table S2). A hierarchical clustering of these tumor-specific methylated regions is shown in Fig. 3B. Grade 1 tumors form a tight cluster, adjacent to five higher grade tumors, T4-2, T3-3, T2-4, T4-5, and T4-1. The other higher-grade tumors form two distinct clusters. One cluster includes T2-2, T2-3, T3-2, T2-5, T3-4, and T2-6. The rest of the tumors form another cluster and they have the most methylation peaks. Of the 428 frequently methylated regions, 403 overlap with CpG islands, and 288 fall into genes, i.e., that are within 1,000 bp upstream of the transcription start site and 1,000 bp downstream of the transcription end of a gene. There were 174 methylation peaks that could be annotated to 5'-ends of RefSeq transcripts (Supplementary Table S3). Of these, 164 genes had annotation information in the Database for Annotation, Visualization, and Integrated Discovery. Functional analysis revealed that genes involved in organ, nervous system and brain development, and neurogenesis are highly enriched (Table 1). Many of these genes are transcription factor and homeobox genes. An analysis of InterPro domains revealed 35 homeobox genes in the list of 288 genes with methylation peaks (enrichment fold = 12.2, $P = 3.7 \times 10^{-26}$; InterPro domain IPR001356). Because homeobox genes and other developmental regulatory genes are known targets of the repressive Polycomb complex in embryonic stem cells, we

compared the list of commonly methylated genes with the list of genes marked as Polycomb targets in human embryonic stem cells (24). Supplementary Fig. S3 shows that indeed 30.7% of the methylated genes are Polycomb targets. The observed number of overlapping targets between MIRA and Polycomb is more than 7-fold higher than expected by chance, indicating a strong enrichment ($P < 2.2 \times 10^{-16}$; Fisher's exact test).

Although we focused mostly on DNA hypermethylation in this study, we also looked for hypomethylation of CpG islands in the tumor samples relative to normal brain. We identified a list of 46 genes in which the hypomethylation event is frequently present at the 5' gene end (Supplementary Table S4). This list was enriched for cancer-testis antigen genes including the genes *SOHLH2*, *SSX2*, *SSX4B*, *SSX8*, *SSX9*, and *PAGE5*. This class of genes is normally expressed in germ cells, is silenced by DNA methylation in somatic tissues but can be reactivated and demethylated in cancer.

Clusters of CpG island hypermethylation in brain tumors. In addition to the many singular hypermethylated CpG islands present along all chromosomes, we identified a number of prominent clusters of hypermethylated CpG islands. These clusters were defined as a minimum of five adjacent CpG islands that underwent simultaneous hypermethylation in at least five of the tumors but are methylation-free in all normal brains. These clusters could be divided into two categories: (a) CpG islands associated with homeobox genes and (b) hypermethylation clusters not associated with homeobox genes. The first category includes the following loci: the *HOX* gene clusters *HOXD* (Fig. 4) and *HOXC* (Supplementary Fig. S4), and the homeobox genes *DLX1*, *BARHL2*, and *PITX2* (Supplementary Figs. S5 and S6 and data not shown), as well as the developmental transcription factor gene simple-minded one

(SIM1; Supplementary Fig. S7). The *HOX* clusters *HOXA* and *HOXB* were also hypermethylated in tumors but several of the genes in these two clusters had considerable methylation in normal brain. The hypermethylated CpG islands at these homeobox gene loci were found not only at the 5' gene ends but also within the gene bodies and at the 3'-ends. Chromosome 14 shows a particularly illustrative example of homeobox gene-associated hypermethylation of CpG island clusters (Supplementary Fig. S8). In this case, several homeobox and transcription factor genes, including *NKX2-8*, *PAX9*, and *FOXA1* were located within a 1.8 mb segment, and no less than 24 contiguous CpG islands became hypermethylated in astrocytomas.

A 360 kb stretch at chromosome 1q21.2 (145,940,000–146,300,000) contained 12 hypermethylated CpG islands. In this case, four consecutive tumor-specifically hypermethylated islands and eight adjacent tumor-specifically methylated islands were separated by two CpG islands that were methylated in tumors and in normal brain. Interestingly, this chromosomal region is rich in segmental duplications

(Supplementary Fig. S9). This locus contains the neuroblastoma breakpoint family (*NBPF*) of genes named after a chromosomal translocation found in a neuroblastoma patient (25). The *NBPF* genes encode proteins of unknown function and have a repetitive structure with high intragenic and intergenic sequence similarity in both coding and noncoding regions. No discernable *NBPF* orthologues have been identified in the genomes of mouse or rat. This gene family shows primate-specific duplications that result in species-specific arrays of *NBPF* homologous sequences (25).

The hypermethylation cluster on chromosome 6p21 (25,700,000–30,500,000) spanning 4.8 mb is particularly striking. This hypermethylation cluster was present in 16 out of 30 tumors analyzed (Fig. 5). The region contains the major histone cluster 1 and several *HLA* genes of the *MHC* locus. This large cluster of histone genes contains 55 histone genes (26). Thus, this chromosomal region represents a segment in which closely related genes, the histone gene families, and *HLA* genes are present and are subject to hypermethylation

Table 1. Functional annotation of 5'-end methylated genes in tumor samples

GO ID	GO term	Count	P	Benjamini	Genes
GO:0048513	Organ development	34	7.50E-09	1.33E-05	<i>IL12B, PDGFA, LECT1, BMP4, OTX2, MYF6, CNTN4, PAX9, FYN, SHOX2, EN1, CHAT, GSX2, KRT19, SIX6, POU4F3, RAX, FOXD3, GDNF, HOXD12, NPR3, HOXD3, HOXD10, MEST, ONECUT2, ALX1, NKX2-6, ONECUT1, OTP, GHRSR, LAMC2, DLX2, PAX6, NKX6-1, CFTR</i>
GO:0007420	Brain development	11	1.31E-06	1.16E-03	<i>CNTN4, ALX1, RAX, OTP, FYN, EN1, DLX2, BMP4, GSX2, OTX2, CFTR</i>
GO:0007399	Nervous system development	22	3.38E-06	1.99E-03	<i>POU4F3, RAX, NCAM2, GDNF, BMP4, OTX2, NEFM, PCDHAC1, CNTN4, ALX1, OTP, FYN, GHRSR, EN1, SHOX2, DLX2, CHAT, NKX6-1, GSX2, CRMP1, PAX6, NKX2-8, CFTR</i>
GO:0031323	Regulation of cellular metabolic process	47	8.01E-06	3.54E-03	<i>ZNF560, ZNF233, EVX2, IL12B, PDGFA, BMP4, VSX2, OTX2, KLF11, MYF6, IRX5, PAX9, GSC, KLF16, EN1, SHOX2, GSX2, GRHL2, SOX7, SIX6, NKX2-8, CPEB1, TOX2, POU4F3, RAX, FOXD3, BCL6B, HOXD12, HOXD3, HOXD10, ZNF236, ONECUT2, ALX1, C14orf39, EVX1, HOXA1, ONECUT1, NKX2-6, OLIG3, OTP, PRDM14, GHRSR, DLX2, NKX6-1, HOXA10, WT1, HOXA7</i>
GO:0022008	Neurogenesis	13	1.49E-05	1.28E-02	<i>POU4F3, GDNF, OTX2, NEFM, CNTN4, OTP, FYN, EN1, CHAT, GSX2, PAX6, NKX6-1, NKX2-8, CFTR</i>
GO:0048699	Generation of neurons	12	4.40E-05	1.29E-02	<i>CNTN4, POU4F3, OTP, GDNF, FYN, EN1, CHAT, GSX2, OTX2, NEFM, NKX2-8, PAX6, CFTR</i>
GO:0045449	Regulation of transcription	42	4.73E-05	1.19E-02	<i>ZNF560, ZNF233, EVX2, VSX2, OTX2, KLF11, MYF6, IRX5, PAX9, GSC, KLF16, EN1, SHOX2, GSX2, GRHL2, SOX7, NKX2-8, SIX6, TOX2, POU4F3, RAX, FOXD3, BCL6B, HOXD12, HOXD3, HOXD10, ZNF236, ONECUT2, ALX1, C14orf39, EVX1, HOXA1, ONECUT1, NKX2-6, OLIG3, OTP, PRDM14, DLX2, PAX6, NKX6-1, HOXA10, WT1, HOXA7</i>

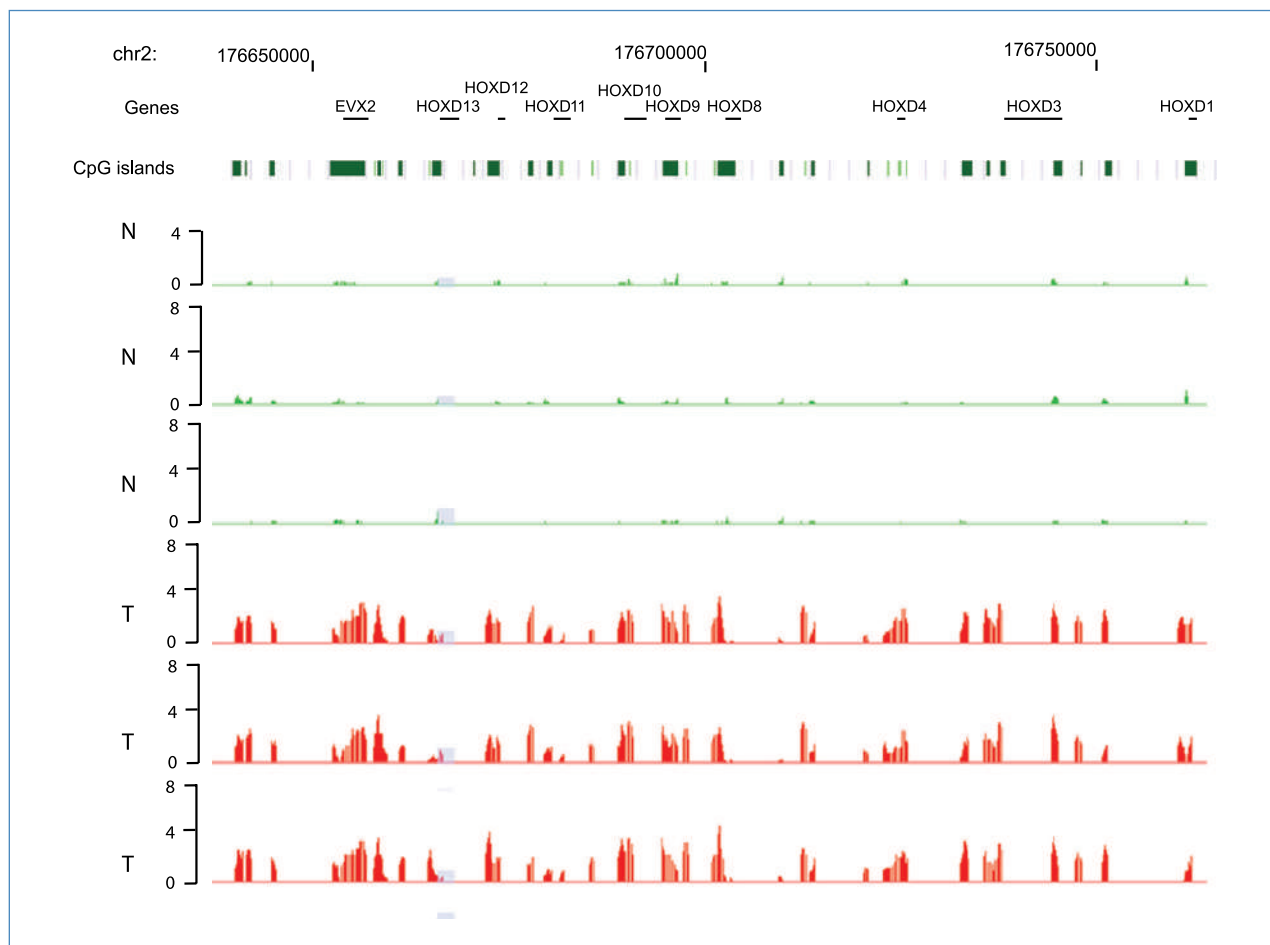


Figure 4. Clusters of hypermethylated CpG islands at the *HOXD* locus. Top, the locations of individual *HOXD* genes. The associated CpG islands are shown as green boxes. Bottom, the NimbleGen array data (methylated fraction versus input) for three normal brain tissues (N, green) and three grade 3 astrocytomas (T, red). The methylation signal is shown plotted along the chromosome as a *P* value score. Therefore, the minimum number on the Y-axis is 0 (when $P = 1$). The *P* value score was obtained by NimbleScan software and is derived from the Kolmogorov-Smirnov test comparing the log 2 ratios (MIRA versus input) within a 750 bp window centered at each probe and the rest of the data on the array.

in brain tumors. Although some CpG islands in this region are methylated in normal brain as well, there are 42 CpG islands that are subject to tumor-specific hypermethylation.

Discussion

Our work provides a global view of DNA hypermethylation events occurring in human astrocytomas. We determined that 6,000 to 7,000 CpG islands are methylated in normal brain. This number represents well over 20% of all CpG islands in total and ~5% of all CpG islands at 5' gene ends were methylated. It has become recognized recently that autosomal CpG island methylation is a rather common phenomenon in differentiated somatic tissues (20, 27–30). We found that genes with methylation peaks at their 5'-end are expressed at a significantly lower level than genes without methylation peaks at the 5'-end.

Diagnosis of brain cancer by detecting hypermethylated genes in blood serum or plasma is a promising approach,

for which initial data have already shown feasibility (31). One important attribute of such diagnostic DNA methylation markers is that they should be methylated in nearly all tumors, including early stage and low-grade tumors. To accelerate biomarker development, we created a list of 428 methylated regions that were methylated in >25% of tumors (i.e., in at least 8 of the 30 tumors analyzed; Supplementary Table S2). The most frequently methylated region was an uncharacterized CpG island on chromosome 7 that is methylated in 87% of the tumors. The most frequently methylated gene-associated region was within the gene body of the *EBF1* gene located on chromosome 5. The second most frequently methylated gene was *DHH* (desert hedgehog), methylated in 83% of the tumors. We identified six methylated regions that were methylation-positive in at least five of the six grade 1 tumors, including *EBF1* (Supplementary Table S5). One somewhat unexpected result of our study was the finding that grade 4 tumors have fewer hypermethylated regions than several of the grade 2 and 3 tumors. A larger series of

samples would need to be analyzed to determine if this is a general trend or if there is a large heterogeneity within that group as well.

Several known tumor suppressor genes are methylated in astrocytomas. These include the genes *CDKN2A*, *PTEN*, *p14ARF*, and *RB*. The reason that these genes are not listed in Supplementary Tables S2 and S3 are for the most part that these genes were not methylated in more than 25% of the tumors (our criterion to be included on the list). This finding is consistent with the relatively infrequent methylation of these genes reported in the literature. We found methylation of the *PTEN* promoter in 20% of the tumors. *CDKN2A* was methylated in 23% and two tumors displayed no signal in the entire gene area suggesting homozygous deletion. For the *MGMT* gene, for which reported methylation frequencies in gliomas generally range from 40% to >70%, we found weak signals at the promoter in normal brain and enhanced signals in ~40% of the tumors. However, the statistical cutoff used to score differential peaks (see Materials and Methods) did not allow this gene to be scored on our list.

On the list of the 288 most frequently methylated genes, homeobox genes and other developmental regulatory transcription factors were greatly overrepresented. Thirty-one percent of these genes were targets of the Polycomb complex (H3K27me3-marked) in embryonic stem cells (Supplementary Fig. S3). This is similar to what has been found in other tumor types, for example lung (14, 18), breast (32, 33), and colorectal cancer (15, 34), lymphomas (35), as well as being similar to a recent report on glioblastoma multiforme (41%, 807 genes were analyzed; ref. 36). This finding reinforces the notion that Polycomb marking has a broad effect on tumor-associated DNA methylation in many—if not all—types of human cancer. Future mechanistic studies will be necessary to determine the functional consequences of this particularly widespread epigenetic alteration in the initiation or progression of human carcinogenesis.

One important finding of our studies was the frequent occurrence of clusters of adjacent hypermethylated CpG islands. Several examples of these included homeobox genes. Mechanistically, this phenomenon is likely related to the

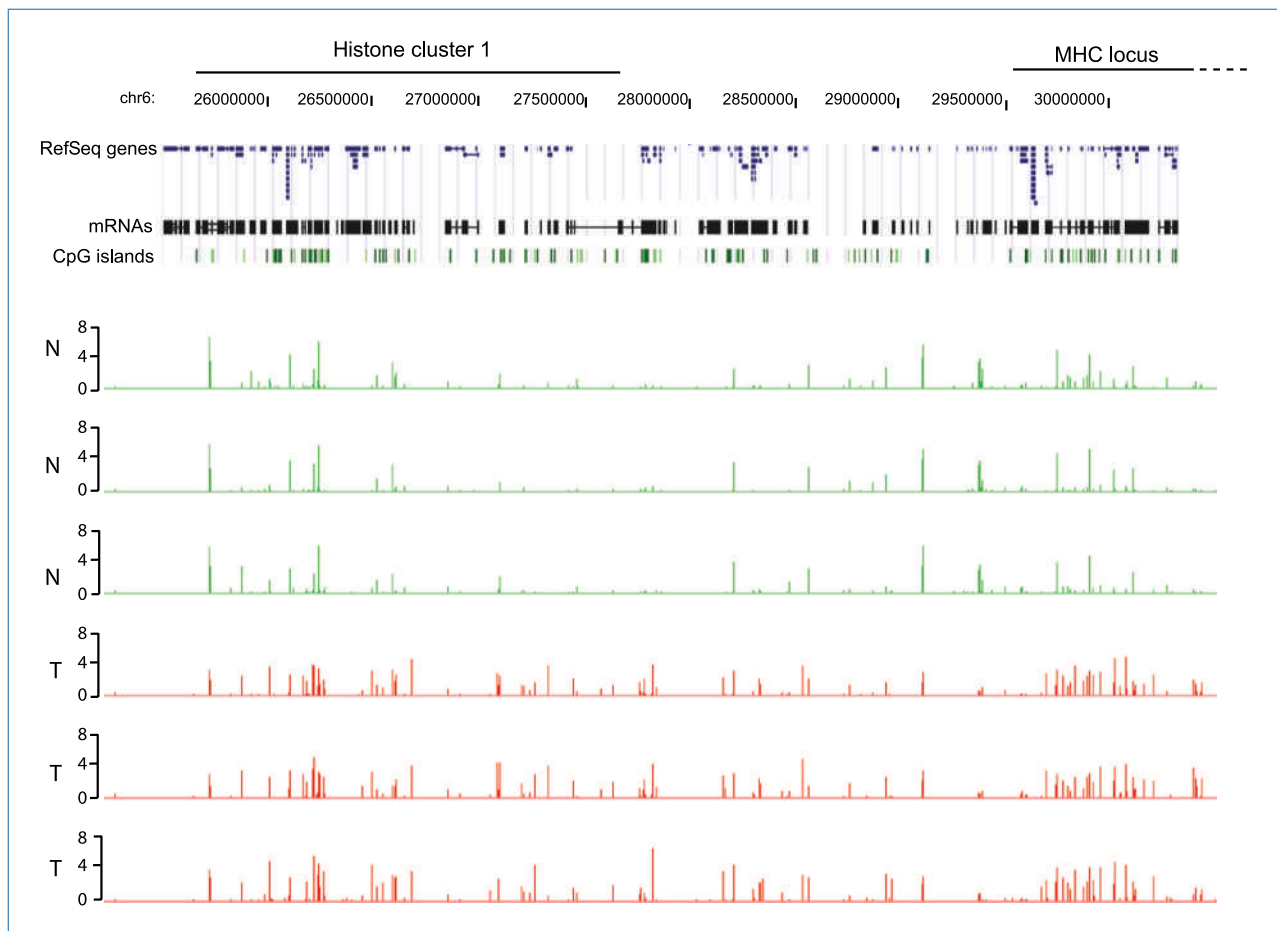


Figure 5. Clusters of hypermethylated CpG islands at the histone cluster 1 and *HLA* loci. Top, location of the histone cluster 1 and *MHC* genes. The location of RefSeq genes and mRNAs according to the UCSC Genome Browser. The associated CpG islands are shown as green boxes. Bottom, the NimbleGen array data (methylated fraction versus input) for three normal brain tissues (N, green) and three grade 3 astrocytomas (T, red). At this level of resolution, each bar represents a methylated CpG island. The methylation signal is shown plotted along the chromosome as a *P* value score.

occupancy of larger DNA sequence stretches by the Polycomb complex surrounding homeobox genes, and would, again, reflect Polycomb-targeted methylation but in a more large-scale fashion. We also identified another situation in which multiple adjacent CpG islands became hypermethylated in tumors, as a result of being present as part of multi-copy gene families. The two examples observed, reflecting extensive hypermethylation of the *NBPF* gene family and of CpG islands in histone cluster 1 and *HLA* genes in the *MHC* cluster suggest that it is gene duplication events per se that may trigger DNA hypermethylation. Paradoxically, this situation is the opposite to what we have observed in lung squamous cell carcinomas in which segmental duplications are generally hypomethylated in tumors (23). These data suggest that different types of malignancies have different aberrations concerning methylation of duplicated gene sequences.

Gene ontology analysis indicated that a significantly enriched gene category with common 5' gene end methylation included genes involved in brain development and neurogenesis, once again with frequent involvement of Polycomb target and homeobox genes (Table 1). Recent mouse models have shown that gliomas arise from neural stem cells in the subventricular zone and/or transit-amplifying progenitors rather than from normal glial progenitors or oligodendrocytes (37, 38). Neural stem cells in the adult subventricular zone give rise to highly proliferating transit-amplifying progenitor cells, which then differentiate into the different lineages, including neuroblast, astrocyte, and oligodendrocyte precursors. Aberrant or failed neurogenesis, due to methylation-induced permanent silencing of neuronal differentiation genes such as *POU4F3*, *GDNF*, *OTX2*, *NEFM*, *CNTN4*, *OTP*, *SIM1*, *FYN*, *EN1*, *CHAT*, *GSX2*, *NKX6-1*, *PAX6*, *RAX*, *DLX2*, and *BMP4* and several others (Table 1) may be a fundamental component of glioma

initiation from neural stem cells (see model in Supplementary Fig. S10). For example, *BMP4* has been shown to reduce the glioma-initiating cell pool in glioblastomas (39). Given that subventricular zone neural stem cells are the cells of origin of gliomas (37, 38), these methylation changes do not simply reflect lineage committed oligodendrocyte progenitors as glioma precursors, in which such genes are likely methylation-silenced. Rather, aberrant methylation of a group of genes involved in neuronal differentiation, in cooperation with other oncogenic events such as *TP53* mutation, may shift the balance from regulated differentiation towards cell proliferation and gliomagenesis in neural stem cells. The permanent silencing of prodifferentiation genes might lead to the accumulation of a cell population that is unable to differentiate but which could persist long enough, perhaps over the lifetime of an individual, to acquire the necessary transforming genetic or epigenetic aberrations. Our data suggest that epigenetic differentiation therapy might provide a useful approach for the management of this lethal disease.

Disclosure of Potential Conflicts of Interest

Under a licensing agreement between City of Hope and Active Motif (Carlsbad, CA), the MIRA technique was licensed to Active Motif, and the authors T.A. Rauch and G.P. Pfeifer are entitled to a share of the royalties received by City of Hope from sales of the licensed technology.

Grant Support

NIH grant CA084469 (G.P. Pfeifer).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 10/01/2009; revised 12/15/2009; accepted 01/25/2010; published OnlineFirst 03/16/2010.

References

- Alcantara Llaguno SR, Chen J, Parada LF. Signaling in malignant astrocytomas: role of neural stem cells and its therapeutic implications. *Clin Cancer Res* 2009;15:7124–9.
- Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–12.
- Jones PA, Baylin SB. The epigenomics of cancer. *Cell* 2007;128:683–92.
- Nagarajan RP, Costello JF. Epigenetic mechanisms in glioblastoma multiforme. *Semin Cancer Biol* 2009;19:188–97.
- Costello JF, Fruhwald MC, Smiraglia DJ, et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 2000;24:132–8.
- Uhlmann K, Rohde K, Zeller C, et al. Distinct methylation profiles of glioma subtypes. *Int J Cancer* 2003;106:52–9.
- Yu J, Zhang H, Gu J, et al. Methylation profiles of thirty four promoter-CpG islands and concordant methylation behaviours of sixteen genes that may contribute to carcinogenesis of astrocytoma. *BMC Cancer* 2004;4:65.
- Cadioux B, Ching TT, VandenBerg SR, Costello JF. Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. *Cancer Res* 2006;66:8469–76.
- Foltz G, Ryu GY, Yoon JG, et al. Genome-wide analysis of epigenetic silencing identifies BEX1 and BEX2 as candidate tumor suppressor genes in malignant glioma. *Cancer Res* 2006;66:6665–74.
- Kim TY, Zhong S, Fields CR, Kim JH, Robertson KD. Epigenomic profiling reveals novel and frequent targets of aberrant DNA methylation-mediated silencing in malignant glioma. *Cancer Res* 2006;66:7490–501.
- Fanelli M, Caprodossi S, Ricci-Vitiani L, et al. Loss of pericentromeric DNA methylation pattern in human glioblastoma is associated with altered DNA methyltransferases expression and involves the stem cell compartment. *Oncogene* 2008;27:358–65.
- Rauch T, Li H, Wu X, Pfeifer GP. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res* 2006;66:7939–47.
- Ohm JE, McGarvey KM, Yu X, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* 2007;39:237–42.
- Rauch T, Wang Z, Zhang X, et al. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc Natl Acad Sci U S A* 2007;104:5527–32.
- Schlesinger Y, Straussman R, Keshet I, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nat Genet* 2007;39:232–6.
- Widschwendter M, Fiegl H, Egle D, et al. Epigenetic stem cell signature in cancer. *Nat Genet* 2007;39:157–8.
- Gal-Yam EN, Egger G, Iniguez L, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3

- prostate cancer cell line. *Proc Natl Acad Sci U S A* 2008;105:12979–84.
18. Pfeifer GP, Rauch TA. DNA methylation patterns in lung carcinomas. *Semin Cancer Biol* 2009;19:181–7.
 19. Rauch TA, Pfeifer GP. The MIRA method for DNA methylation analysis. *Methods Mol Biol* 2009;507:65–75.
 20. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP. A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* 2009;106:671–8.
 21. Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* 1997;25:2532–4.
 22. Ball MP, Li JB, Gao Y, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009;27:361–8.
 23. Rauch TA, Zhong X, Wu X, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc Natl Acad Sci U S A* 2008;105:252–7.
 24. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 2006;125:301–13.
 25. Vandepoele K, Van Roy N, Staes K, Speleman F, van Roy F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. *Mol Biol Evol* 2005;22:2265–74.
 26. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. The human and mouse replication-dependent histone genes. *Genomics* 2002;80:487–98.
 27. Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;38:1378–85.
 28. Shen L, Kondo Y, Guo Y, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet* 2007;3:2023–36.
 29. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 2007;39:457–66.
 30. Illingworth R, Kerr A, Desousa D, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* 2008;6:e22.
 31. Weaver KD, Grossman SA, Herman JG. Methylated tumor-specific DNA as a plasma biomarker in patients with glioma. *Cancer Invest* 2006;24:35–40.
 32. Fourkala EO, Hauser-Kronberger C, Apostolidou S, et al. DNA methylation of polycomb group target genes in cores taken from breast cancer centre and periphery. *Breast Cancer Res Treat*. In press 2010.
 33. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Res* 2009;11:R14.
 34. Rada-Iglesias A, Enroth S, Andersson R, et al. Histone H3 lysine 27 trimethylation in adult differentiated colon associated to cancer DNA hypermethylation. *Epigenetics* 2009;4:107–13.
 35. Bennett LB, Schnabel JL, Kelchen JM, et al. DNA hypermethylation accompanied by transcriptional repression in follicular lymphoma. *Genes Chromosomes Cancer* 2009;48:828–41.
 36. Martinez R, Martin-Subero JL, Rohde V, et al. A microarray-based DNA methylation study of glioblastoma multiforme. *Epigenetics* 2009;4:255–64.
 37. Alcantara Llaguno S, Chen J, Kwon CH, et al. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer Cell* 2009;15:45–56.
 38. Wang Y, Yang J, Zheng H, et al. Expression of mutant p53 proteins implicates a lineage relationship between neural stem cells and malignant astrocytic glioma in a murine model. *Cancer Cell* 2009;15:514–26.
 39. Piccirillo SG, Reynolds BA, Zanetti N, et al. Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature* 2006;444:761–5.