Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets

J.M.Ordway, J.A.Bedell, R.W.Citek, A.Nunberg, A.Garrido, R.Kendall[2,3], J.R.Stevens[3], D.Cao[2,3], R.W.Doerge[2,3], Y.Korshunova, H.Holemon, J.D.McPherson[4], N.Lakey, J.Leon, R.A.Martienssen[5] and J.A.Jeddeloh*

Orion Genomics, St Louis, MO, [2]Department of Agronomy, Lilly Hall of Lifesciences 915 West State Street, Purdue University, West Lafayette, IN, [3]Department of Statistics, Purdue University, 150 North University Street, West Lafayette, IN, [4]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX and [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

*To whom corresondence should be addressed at Science and Technology, Orion Genomics, LLC, 4041 Forest Park Avenue St. Louis, MO 63108, USA. Tel: +1 314.615.6977; Fax: +1 314.615.6975; Email: jjeddeloh@oriongenomics.com

Using a unique microarray platform for cytosine methylation profiling, the DNA methylation landscape of the human genome was monitored at more than 21 000 sites, including 79% of the annotated transcriptional start sites (TSS). Analysis of an oligodendroglioma derived cell line LN-18 revealed more than 4000 methylated TSS. The gene-centric analysis indicated a complex pattern of DNA methylation exists along each autosome, with a trend of increasing density approaching the telomeres. Remarkably, 2% of CpG islands (CGI) were densely methylated, and 17% had significant levels of 5 mC, whether or not they corresponded to a TSS. Substantial independent verification, obtained from 95 loci, suggested that this approach is capable of large scale detection of cytosine methylation with an accuracy approaching 90%. In addition, we detected large genomic domains that are also susceptible to DNA methylation reinforced inactivation, such as the HOX cluster on chromosome 7 (CH7). Extrapolation from the data suggests that more than 2000 genomic loci may be susceptible to methylation and associated inactivation, and most have yet to be identified. Finally, we report six new targets of epigenetic inactivation (IRX3, WNT10A, WNT6, RARα, BMP7 and ZGPAT). These targets displayed cell line and tumor specific differential methylation when compared with normal brain samples, suggesting they may have utility as biomarkers. Uniquely, hypermethylation of the CGI within an IRX3 exon was correlated with over-expression of IRX3 in tumor tissues and cell lines relative to normal brain samples.

Abbreviations: CGI, CpG islands; CH7, chromosome 7; Ct=cycle threshold; DD, double digest; DM=densely or uniformly methylated; FDR, false discovery rate; IM, intermediately methylated; M=mock digested; MD, methylation-depleted; MDRE=methylation dependent restriction enzyme explanations; MSRE, methylation sensitive restriction enzyme; R=fraction refractory to analysis, or Purine nucleotide (IUPAC); T=treated with McrBC; TSS, transcriptional start sites; U=unmethylated or sparsely methylated; UT=untreated; W=analytical window.

Introduction

Identifying molecular differences that distinguish tumor tissue from normal tissue is a current topic area of intense interest. Although, tumor genomes display only a limited number of primary sequence differences from the nearly isogenic normal tissues in proximity to them (1), a large number of molecular differences exist. In particular, the spectrum of sequences that normal and tumor genomes specify and mark as silent chromatin have been used as 'epigenetic signatures' to molecularly discriminate these cells (2).

Disruption of normal gene regulation is important for carcinogenesis (3,4) resulting in loss, or gain of genetic function. The molecular events that underlie this altered regulation include point-mutations and macro-mutations, such as deletion, amplification or genomic rearrangement (e.g. translocation), that can result in more complex interactions when regulatory genes are affected. Recently, the importance of epigenetic perturbation of gene regulation in the form of changes in chromatin structure has begun to be more fully appreciated (5). In the context of cancer, inappropriate chromatin packaging of genes can lead to gene silencing, or in some cases, ectopic gene expression (6).

Cytosine methylation is a chemically stable mark that may establish, or follow as a consequence of, the packaging of a particular region into silent chromatin. Therefore identification of aberrant genomic DNA methylation that is associated with carcinogenesis should identify loci that are important for disease progression (2).

Mammalian DNA methylation patterns that transmit cellular silencing signals are mitotically maintained with 96–99% fidelity (7). This lies in stark contrast with the primary sequence which is maintained with fidelity over 99.9999% (8). The scale of the difference suggests that epigenetic deregulation may be under appreciated, and recent studies have highlighted the role of the environment, e.g. through dietary folate metabolism, in maintaining DNA methylation and gene silencing, hinting at a mechanism underlying predisposition for cancer (9,10).

Several cancer therapies targeting the maintenance of cytosine methylation and silencing states are in human clinical trials (11–13). Currently, the therapies target either the DNA methylation machinery, or the histone modification machinery. This machinery works synergistically to maintain gene silencing. While such therapies are very promising, the clinical success rates achieved thus far have been similar to more conventional chemotherapies. Therefore, selecting patients for

such epigenetic therapies or measuring their success may require an understanding and characterization of the sequences affected by epigenetic perturbation in particular diseases.

Finding and characterizing the genomic loci capable of driving carcinogenesis from the epigenetic perspective, as well as those capable of serving as clinically meaningful disease or therapy-specific markers is a pressing need. We chose to address the need through the adaptation of a microarray-based DNA methylation profiling approach for use in the human genome. We report here the results obtained in the course of characterizing the oligodendroglioma derived cell line LN-18 (CRL-2610). Future applications of this technology will serve not only to directly identify those sequences which undergo epigenetic alteration in disease, but also as an epigenetic 'Ames' test for environmental history.

Materials and methods

Cell line, normal and tumor brain tissue and blood nucleic acid samples
Cell lines LN-18 (glioblastoma), A-172 (glioblastoma), LN-229 (glioblastoma), T-98G [glioblastoma multiforme(GBM)], U-138 MG (astrocytoma) and U-87 MG (astrocytoma) were obtained from American Type Culture Collection (Manassas, VA) and cultured under supplier's recommendations (http://www.atcc.org/). Histologically normal/non-malignant brain tissue (~0.4 mg ea) from the cerebrum (right temporal lobe) of three different patients (15084A1, 16920A1 and 8387B1) were acquired from Asterand plc. (Detroit, MI). Primary tumor samples [two astrocytoma (1FMXAFPB, JA4CIFNL) and one GBM specimen (CAURPFJD)] were acquired from Genomics Collaborative Inc (Caimbridge, MA). All the tumor samples exhibited >90% neoplastic cellularity. The age range of the normal samples (45,56,87) was controlled relative to the tumors (48,59,59). Genomic DNA was extracted by the MasterPure DNA Purification kit under manufacturer's recommended conditions (Epicentre Biotechnologies, Madison, WI). Male whole blood genomic DNA, representing a pool of peripheral blood samples obtained from approximately six individuals, was obtained from Novagen (Madison, WI).

RNA was purified from the tumor tissues and cell lines using the RNeasy Kit from Qiagen (Valencia, CA). cDNA from normal brain and astrocytoma tissues was purchased from Invitrogen (Carlsbad, CA). cDNA from the cell lines and tumor samples was produced using the oligo dT primers provided in the Superscript III first strand system from Invitrogen (Carlsbad, CA) under the manufacturer's recommended conditions.

Microarray experimental and analytical approach
Methylation profiling was performed as described previously (14,15) but with adaptations to apply the method to the human genome. The adaptations included the use of target mass normalization before labeling, the employment of long oligonucleotide probes (60mer) on the NimbleGen systems CGH platform, and the use of sequences within the human genome that are incapable of digestion by McrBC as controls.

Methylation-dependent DNA fractionation
LN-18 DNA was mechanically sheared into a uniform molecular weight distribution using a GeneMachines HydroShear apparatus. Sheared DNA was split into four equal portions of 100 μg each. Two portions (treated technical replicates) were digested with McrBC (NEB) in 300 μl total volume including 1× NEB2 buffer, 0.1 mg/ml BSA, 2 mM GTP and 200 U McrBC. The remaining two portions were mock-treated under identical conditions except that 20 μl water was added instead of McrBC. Treated and mock-treated reactions were incubated at 37°C overnight. All reactions were treated with 5 μl proteinase K (50 mg/ml) for 1 h at 50°C, and precipitated with EtOH under standard conditions. Pellets were washed twice, dried and resuspended in 50 μl H$_2$O. A total of 50 μg of each fraction was resolved on a 1% low melting point SeaPlaque GTG Agarose gel (Cambridge Bio Sciences, Rockland, ME). Untreated and treated replicates were resolved side-by-side. DNA (1 kb) sizing ladder was run adjacent to each untreated/treated pair to guide accurate gel slice excision. Gels were visualized with long-wave ultraviolet (UV), and gel slices including DNA within the modal size range of the untreated fraction (~1–4 kb) were excised with a clean

razor blade. DNA was extracted from gel slices using the GenElute agarose spin columns (Sigma–Aldrich).

Human array (OGHA v1.0) design
To identify epigenetic changes in gene regulation, our efforts were focused largely upon the transcription start sites of genes. Since promoter annotation and TSS identification are generally less reliable than gene annotation, we tested the ENSEMBL gene predictions using a highly characterized set of human promoters from the Eukaryotic Promoter Database (EPD, http://www.epd.isb-sib.ch/). Substantial similarity was observed between ENSEMBL gene predictions and EPD annotations. Therefore, we decided that the ENSEMBL annotated genes would be an ideal probe design source. The TSS information for the 24 847 ENSEMBL annotated human genes (NCBIv34) was extracted. 60mer probes were designed and uniqueness tested using custom primer selection scripts (J. D. McPherson, unpublished data). The average position of the TSS associated 60mer probes relative to +1 is −500 bp.

The array design consists of 85 176 probes. Each feature is represented with four on-board replicates yielding a total of 21 294 unique feature probes. The features represent 19 595 transcriptional start sites (TSS) (79% of the identified human genes), 1395 GenBank BAC annotated CG islands (CGi), 161 probes spanning ~165 kb along the MTAPase/CDKN2A/B locus on CH9, 66 additional probes dedicated to cancer gene promoters, and 77 probes designed as copy number (HERV, LINEs and SINE) and other controls. We analyzed CGI in two ways. The first was using GenBank BAC accessions with annotated CGI upon them as a source for probe design. The second was mapping University of California Santa Cruz (UCSC) annotated CGI relative to our probe set. Together, the TSS probes and CGi probes scan more than 9000 UCSC annotated human CG islands. Feature coverage of the OGHA v1.0 microarray in the human genome is depicted in Figure 2. The vertical lines extending up from the chromosome represent features designed to the Watson strand, and the vertical lines down from the chromosome represent features designed to the Crick strand.

Dye labeling and microarray hybridization
Each fraction was labeled with Cy3 and separately with Cy5 by *in vitro* synthesis reactions. A total of 200 ng of fractionated template DNA and 20 μl 2.5× random hexamer cocktail (Invitrogen) were combined in a total volume of 25.5 μl and incubated in a boiling water bath for 5 min. Reactions were placed on ice for 3 min. Reaction volumes were adjusted to 50 μl including 140 μM each dATP, dGTP and dTTP, 60 μM dCTP (Invitrogen), 1.5 μl 25 nmol Cyanine 3-dCTP or Cyanine 5-dCTP (NEN), and 100 U 3′–5′ exo$^-$ Klenow (NEB). Reactions were incubated at 37°C for 45 min then inactivated by addition of 50 μl TE. Labeled DNA was purified using MinElute spin columns (Qiagen) and eluted in 35 μl EB buffer. Synthesis yields and Cy-dye specific activities were determined by Nanodrop spectrophotometry. One-half of each labeled reaction was used per microarray hybridization. A duplicated dye-swap experimental design was employed and four microarray hybridizations were performed. Microarray hybridization, washing and scanning were performed by Nimblegen Systems of Iceland under their standard CGH operating protocols.

Statistical analysis of microarray data
A two-stage ANOVA model was employed (16). In the first stage the model fit was: $y_{gijkl} = \mu + T_i + A_j + D_k + \varepsilon_{gijkl}$, where y = log(intensity), μ = global mean, T = fixed treatment effect, $i = 1, 2$; A = fixed array effect, $j = 1, 2, 3, 4$; D = fixed dye effect, $k = 1, 2$; $g = 1, 2, \ldots, 21294$ (gene), $l = 1, \ldots, g_l$, (some genes have replicates for combinations of T, A and D), ε_{gijkl} = error, iid ~ $N(0, \sigma^2)$.

Let T_1 and T_2 correspond to untreated samples (UT) and treated methylation-depleted (MD) samples, respectively. In the second stage of the ANOVA the residual for each gene g from the first stage analysis, r_{gijkl}, was used as the response variable in the model: $r_{gijkl} = G_g + (GT)_{gi} + (GA)_{gj} + (GD)_{gk} + \gamma_{gijkl}$, where the error terms, γ_{gijkl}, are assumed to be independent $N(0, \sigma_g^2)$.

Normalization to RCG zero features
Subsequent analysis of the LN-18 data utilized those features having no McrBC half-sites within 1000 bp of genomic sequence spanning the probe [i.e. RCG/kb frequency values of zero (RCG 0)] as controls. One would not expect the RCG0 features to be affected by McrBC due to their lack of McrBC half-sites. As such, any variability observed from these features' intensities should be solely attributed to the overall treatment, array and dye effects, plus chance variability. Using the RCG0 features to estimate these effects allows for a clearer picture of the methylation status of the other features. For this reason, the first-stage ANOVA model was fit using only the RCG0 features. The estimates of μ, T_i, A_j and D_k were obtained

and then used in the normalization of the remaining non-control genes. The normalized methylation levels $\tilde{y}_{gijkl} = y_{gijkl} - (\hat{\mu} + \hat{T}_i + \hat{A}_j + \hat{D}_k)$ were obtained for all non-control features and the two-stage model fit.

Testing for methylation

Once the two-stage analysis was performed, the following hypotheses were tested to detect methylated features (i.e. log scale): H_0: MD=UT versus H_1: MD $\neq$ UT. Equivalently, for each gene g the hypotheses H_0: $GT_{g2} = GT_{g1}$ versus H_1: $GT_{g2} \neq GT_{g1}$ was tested. The test statistic for testing expression of untreated versus methylation-depleted samples, $V^*_g = GT_{g1} - GT_{g2}$, is easily interpreted as UT-MD (or U-T), the untreated treatment effect minus the methylation-depletion treatment effect for gene g.

Because of the number of statistical tests made (21 294) two methods were used to control for multiple testing errors. Holm's sequential adjustment provides strong control of the family-wise error rate (FWER) below significance level α, with greater power than the standard Bonferroni method (17). The false discovery rate (FDR) controlling method of Benjamini and Hochberg (18) provides weak control of the FWER, and controls the FDR below level α. The FDR is defined as the expected proportion of falsely rejected null hypotheses relative to the total number of rejections. The FDR was controlled at 0.05. Matlab (Mathworks) and SAS statistical packages (SAS Institute) were used for the analyses and generation of figures.

Bisulphite sequencing

A total of 2 µg LN-18 genomic DNA was bisulphite converted with an EZ DNA Methylation Kit (Zymo Research) using the manufacturer's recommended protocol. Primers and loci selected for validation, along with PCR conditions are indicated in the Supplementary Data. Sequencing and analysis was performed as described previously (19).

Quantitative PCR-based validation of DNA methylation predictions (MethylScreen)

For each analyzed genomic DNA sample, 4 µg of genomic DNA was digested with McrBC (NEB) in 100 µl total volume including 1× NEB2 buffer, 0.1 mg/ml BSA, 2 mM GTP and 50 U McrBC overnight at 37°C. In parallel, 4 µg of each genomic DNA sample was mock-treated under identical conditions with the exception that water was substituted for McrBC. Following digestion, enzyme activity was heat inactivated at 65°C for 20 min. A total of 40 ng McrBC-digested and 40 ng mock-treated DNA were analyzed by quantitative real-time PCR. In all cases, digested and mock-treated templates were analyzed in adjacent wells. Reactions included 1× FailSafe Buffer G (EpiCentre), 5 µM each primer (see Supplementary Data), and 1 U *Taq* polymerase (Invitrogen) in 25 µl total volume. Standard quantitative PCR cycling conditions were used with a 'hot' plate read of 72°C for 2 min. The melt curve of each amplicon was calculated within a temperature gradient from 60 to 95°C at 1°C increments with a 10 s hold time for each read. The cycle number at which the McrBC-digested sample crossed the threshold was subtracted from the cycle number at which the mock-treated sample crossed the threshold to determine the ΔCt of each locus. Since McrBC digests only DNA including purine-5mC, thereby decreasing the amplifiable copies of loci containing DNA methylation and increasing the Ct relative to the mock-treated sample, increasing ΔCt values reflect increasing levels of local DNA methylation. All average ΔCt values (UT-McrBC) less than −1 were set to −1. Four condition MethylScreen assays employed the two treatments described above [mock and methylation dependant restriction enzyme (MDRE)], as well as a methylation sensitive restriction enzyme treatment [(MSRE) e.g. HhaI (NEB (Beverly, MA))] and a double digest (DD) treatment (e.g. HhaI + McrBC).

DNA methylation occupancy calculations

The amount of DNA remaining after each treatment can be calculated using a comparative threshold method, wherein the relative amount of a population of interest is determined by comparing it to a reference value. In this case, the reference value is internal and can be either the mock treated reaction or the DD reaction. Because it is based on the same locus, the same DNA sample, and the same primers there is no amplification efficiency bias difference to consider. Using this approach, if the efficiency of amplification with a particular primer pair is 100%, the quantity of DNA molecules in the variously treated populations of interest is proportional to the equation:

$$P(T1|T2) = 2^{(Ct2-Ct1)}, \tag{1}$$

where P is the proportional amount of template DNA remaining in Treatment 1 (T1) compared to Treatment 2 (T2), Ct1 is the cycle-threshold for T1 and Ct2 is the cycle-threshold for T2.

For example, given an MSRE treatment Ct of 27 and a mock treatment Ct of 25, the proportion of template DNA in the MSRE reaction is 2 (25–27) = $\frac{1}{4}$ = 25% of the mock treatment. The proportion of template remaining after

the MSRE treatment is an indication of the amount of densely methylated DNA (all MSRE sites must be blocked; i.e. 25% in the example above). The proportion of template remaining after MDRE treatment is an indication of the amount of unmethylated or sparsely methylated DNA (methylated cytosines are not prevalent enough to allow for digestion between the primers). DNA that is not accounted for by either of these treatments must be in an intermediate state of methylation.

The following equations describe some features of the quantitative PCR:

$$Tc \propto 2^{-Ct}, \tag{2}$$

where Tc is the total copies of DNA (absolute number can be calculated by comparisons to known dilutions of standards) and Ct is cycle-threshold.

$$W = CtDD - CtM, \tag{3}$$

where W is the analytical window, CtDD is the cycle-threshold for the DD and CtM is the cycle-threshold for the Mock (untreated) sample

$$R = 2 - W, \tag{4}$$

where R is the proportion of fragments that are refractory to treatment.

The window (W, in Equation 3) establishes the number of molecules that participate in the restriction reactions. Windows of 2 cycles or less indicate that >25% of the molecules are refractory (R, in Equation 4) to enzyme digestion and amplification. As such, samples with analytical windows less then 2 cycles are typically considered to be analytical failures.

$$DM = 2^{-(CtMSRE-CtM)}, \tag{5}$$

$$U = 2^{-(CtMDRE-CtM)}, \tag{6}$$

where DM is the proportion of densely methylated template DNA, U is the proportion of Unmethylated DNA, CtMSRE is the Cycle-threshold of MSRE sample and CtMDRE is the Cycle-threshold of the MDRE sample.

If CtMSRE − CtM is less than 1.0, then it is more accurate to calculate DM using the CtMDRE and calculate DM as the remainder after subtracting off the proportion of U and R (i.e. DM = 1 − U − R). Similarly, if CtMDRE − CtM is less than 1.0, then it is more accurate to calculate U using the CtMSRE and calculate U as the remainder after subtracting off the proportion of DM and R (i.e. U = 1 − DM − R). One can verify the accuracy of these calculations by considering the difference between CtDD and CtMSRE/MDRE. If the methylated gene copies are in fact present following MSRE restriction one would expect them to be destroyed in a DD. Conversely, if the unmethylated gene copies are in fact present following MDRE restriction, one would expect them to be destroyed by the DD. Confidence in the DM and U calls made in this way is assessed by the size of Ct difference between DD and the MSRE/MDRE. Large Ct differences elicit higher confidence. Measurements with a ΔCt less than 0.5 cannot be resolved apart from machine error.

$$IM = 1 - DM - U, \tag{7}$$

where IM is the proportion of Intermediately methylated (IM) DNA.

Some molecules can be methylated such that both enzymes are capable of cutting them. That is, they contain 5 mC but are not completely methylated. They exist as part of each of the above calculations. Typically the fraction of IM molecules are identified and calculated only if the ΔCt of the UT-MSRE and UT-MDRE are each above or near 1.0 (e.g. >0.7). An important note about this method is that the IM copy calculation contains both the error functions associated with the MD and U calculations; typically adjusting this calculation for the amount refractory will not affect the calculation.

All ΔCt calculations yield a unit-less parameter reflecting relative proportion of template DNA between various treatments. The fold treatment effects can be applied to the total calculation so that each of the molecular proportions (DM, U and IM) can be converted back into a unit of copies.

RT-PCR expression analysis

A five point (neat, 1:5, 1:25, 1:125: 1:625) 5-fold dilution curve of cDNA into water was used as template for limited cycle end-point PCR. No-template (water) and genomic DNA (positive) controls were run in parallel. Primers designed to exonic sequences which span the 3′ most intron in either the IRX3 gene or the GAPDH gene were used to monitor expression. The IRX3 primers (Supplementary Table) would amplify a 463 bp product if genomic DNA were the template, while cDNA should yield a product of 293 bp. Similarly, the GAPDH control primers (Supplementary Table)
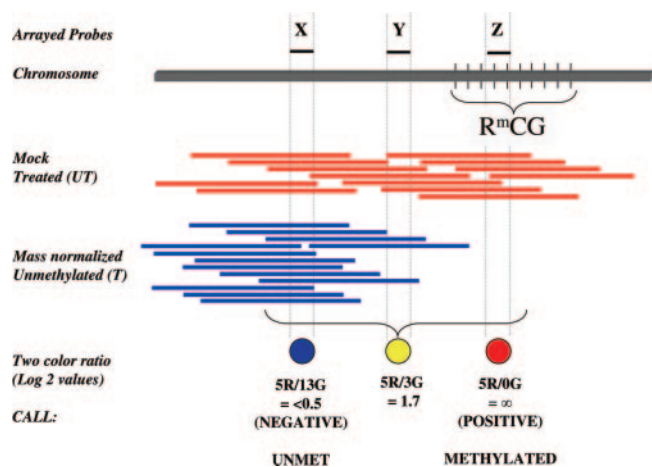
Fig. 1. MethylScope array analysis procedure and theoretical results A schematic of three array probes (X, Y and Z) arranged along a chromosome is shown. If the DNA near feature Z is heavily methylated (vertical bars) approximately half of those methylated CG dinucleotides will be half-sites for McrBC (R^mCG). Sheared and size-selected genomic DNA was labeled with Cy5 (red), while McrBC treated DNA was labeled with Cy3 (blue). There are 13 red fragments and 13 blue fragments as a consequence of the mass normalization prior to target synthesis. Fragments which contain two R^mCG sites have been depleted from the Cy3-labeled population due to the action of McrBC, while unmethylated blue fragments are enriched by mass normalization. The two color array hybridization results are depicted as blue circles for unmethylated, yellow for intermediate or adjacent methylation, or red for densely methylated. The relative signal from the two colors is indicated as a $\log_2$ ratio.

should yield a product of 380 bp if genomic DNA were the template, while cDNA should yield a product of 273 bp. The PCR was performed using the Epicenter (Madison, WI) FailSafe 2× buffer G in a 12 μl reaction. The following cycling parameters were used for both loci: 1 repetition of 95°C for 5 min, 30 repetitions of 95°C for 30 s, 65°C for 15 s, 72°C for 15 s, 1 repetition of 72°C for 10 min. Amplification products were visualized using ethidium-stained agarose gels, and were quantified using an Agilent Bioanalyzer 3100 (Palo Alto, CA) and the DNA 12000 kit.

RESULTS

Methylation profiling was performed using the methylation-dependent restriction enzyme McrBC to separate methylated and unmethylated DNA fragments for labeling and hybridization to microarrays (14). The profiling scheme is depicted in Figure 1. First, purified genomic DNA from a sample of interest is randomly sheared into a uniform size range. The sheared DNA is then split into four portions: two mock treated and two McrBC-digested. In order to restrict DNA, McrBC requires that two methylated half sites (R^mC, where R = A or G) lie within 40–3000 bp of each other (optimal spacing 55–103 bp); cutting has been demonstrated to occur between the two half sites in the proximity of one half-site (20–27). Each matched treatment pair was then size fractionated by electrophoresis and the high-molecular weight DNA that was resistant to treatment was purified from excised agarose gel slices. We have employed independent digests so that any variance during digestion can be assessed across technical replicates. The purified DNA represents the whole genome in the mock-treated sample and the unmethylated fraction of the genome in the McrBC-treated sample.

Next, the same mass of the whole genome sample (untreated, UT) and the methylation-depleted genome sample (treated, T) was independently labeled with different fluorophores. The ratio of mock-treated or 'untreated' genome (UT) to methylation-depleted or 'treated' genome (T) dye-signal is obtained for each feature (UT/T). Digestion with McrBC removed more than half of the DNA from the T sample, but because the same amount of T and UT DNA was used for each labeling, unmethylated sequences will be enriched in the T sample. This mass normalization step increases the signal from unmethylated sequences, and makes these measurements more reproducible, but requires careful subsequent normalization to internal control probes. The ideal set of control probes are sequences which are devoid of McrBC half-sites for a large stretch of sequence (i.e. >1000 or >2000 bp). The resultant two color sample datasets yield hybridization ratios for each feature, and these are assigned a color: blue (low UT/T ratio = unmethylated), yellow (moderate UT/T ratio) or red (high UT/T ratio = methylated) in Figure 1. Sample analysis utilized a duplicated dye swap on pairs of independently treated and UTs for each genome, requiring four arrays and allowing a balanced Latin-square analysis design (See Materials and methods).

Methylated sequences will have a ratio UT/T that approaches infinity (because of the depletion of methylated DNA in the T sample) while unmethylated sequences will have a low UT/T ratio that should be less than 1 because of the enrichment caused by the mass normalization of the UT and T targets. In practice, because the genomic DNA is sheared to a finite size, probes that are adjacent to methylated sequences are also capable of detecting that methylation (feature Y, Figure 1; yellow and orange probes, Figure 2). The distance at which probes can detect adjacent methylation is related to the degree of shearing, and was termed 'wingspan'.

DNA methylation profile of LN-18

The stage IV oligodendroglioma derived cell line LN-18 was chosen as the subject genome for this experiment because it was previously demonstrated to be homozygous for a deletion of the p16(CDKN2A) locus on CH9, which acted as a control region for ploidy assessment (see below). Features representing the CDKN2A region were included on the microarray in addition to the CpG islands (CGI) and TSS throughout the annotated genome (above and Materials and methods). The features in Figure 2 are colored according to the representative methylation profile data [Robust Multi-array Average (28) (performed by NG Systems, WI), linear interpretation] obtained from a duplicated dye swap analysis of LN-18. The color scheme is depicted as a gradient from blue (no methylation) to red (highest-methylation). Yellow features detect intermediate levels of methylation or methylation in regions adjacent to the probe due to wingspan. A substantial number of methylated regions (red) are evident, as well as a large number of unmethylated regions (blue). Globally, each autosome displayed a trend of increasing methylation from pericentromeric regions to telomeres (Figures 2 and 5).

The results from the statistical analysis of the LN-18 genome's cytosine methylation profile are depicted in Figure 3. The signal at each feature was first corrected for technical and biological variation using ANOVA (see Materials and methods), and then normalized to features that represent 1 kb segments of the human genome that contain no McrBC half sites (RCG = 0/kb, denoted hereafter as RCG0). The normalized data were then tested using a

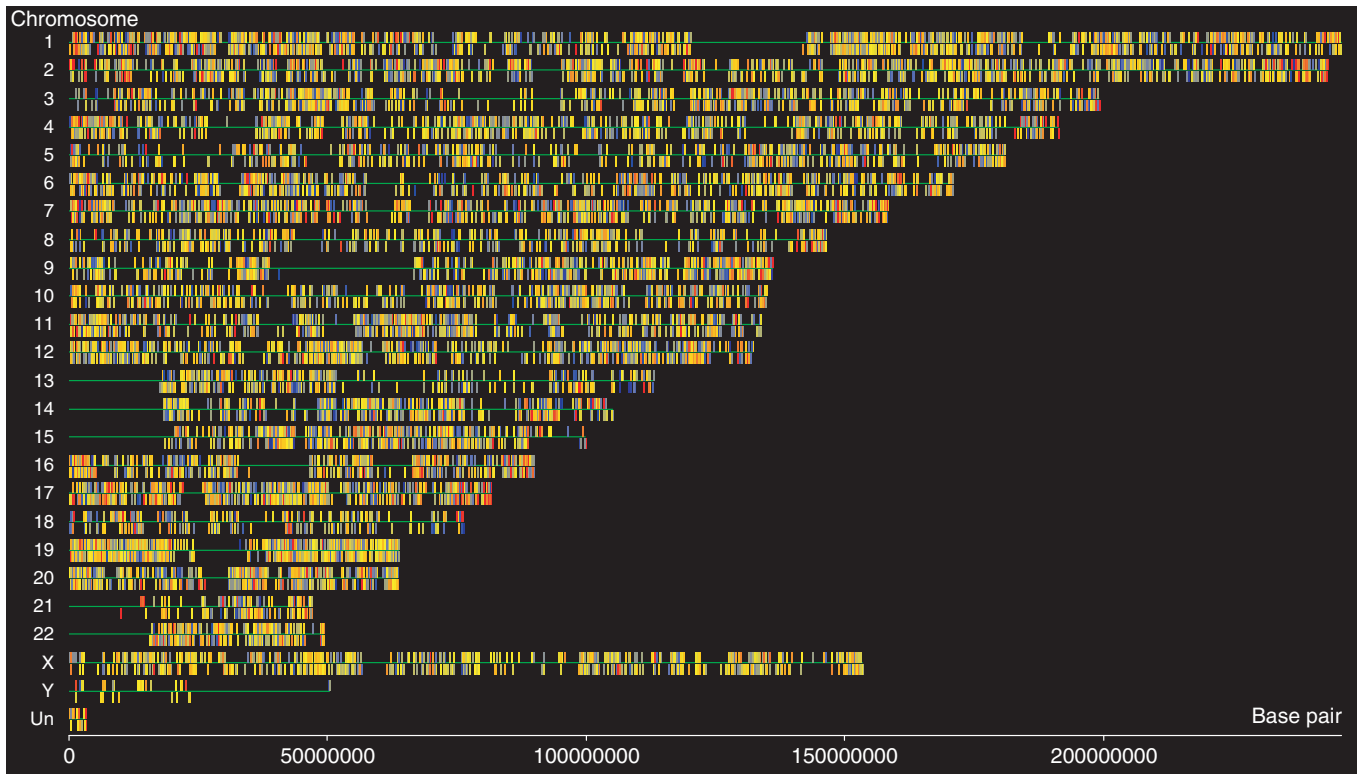**Fig. 2.** Methylation profile of LN-18. (**A**) Each of 21 294 60mer probes were mapped onto the human genome (NCBIv35) by BLAST and is depicted as a vertical line from the Watson (above the line) or Crick (below the line) DNA strand. Areas devoid of probes represent the centromeres, NOR or 5S gene clusters. DNA methylation from the LN-18 genome is measured by the UT/T ratio (see text for details) depicted as blue (unmethylated), yellow (IM) to red (densely methylated).
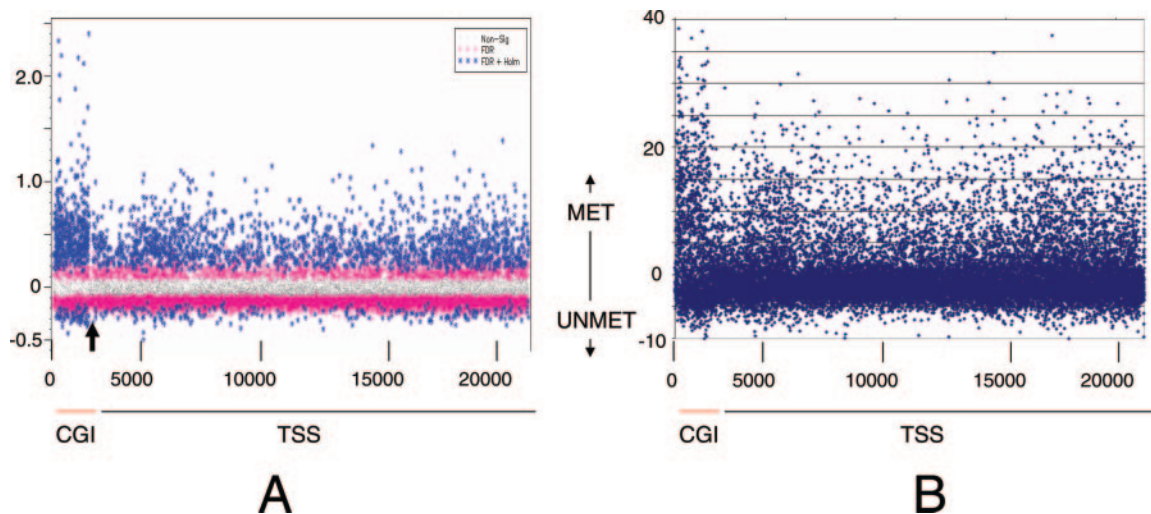

Fig. 3. ANOVA results of LN-18 DNA methylation profile DNA methylation values are plotted for each feature, grouped by category [CGI (black bar), CDKN2A/B (black arrow) and TSS (red bar)]. Methylated sequences have a large positive log ratio while unmethylated sequences have a zero or negative log ratio. Each probe is represented by a color which reflects the statistical significance of the methylation measurement. Blue probes pass the Holm multiple testing correction and have a P-value <0.001. Pink probes pass the FDR multiple testing correction and have a P-value <0.05. Features indicated by gray probes are not significantly different from the RCG0 controls (and are likely unmethylated). (**B**) depicts the ANOVA corrected signal data for each probe divided by its standard error (i.e. t-test values) on the y-axis. The x-axis is the same as (A).

null-hypothesis in which whole genome (UT) signal should be the same as the methyl-depleted (T) signal if the sequences surrounding a given feature were unmethylated. Panel A depicts a graph of the LN-18 methylation ratio (y-axis) by feature name (x-axis). The methylation data are depicted as the ANOVA corrected signal ratio [$\log_2$ (UT/T)].

Each gene was allowed to have its own variance. In this analysis, methylated sequences are expected to have a positive log ratio, while unmethylated sequences are expected to have a log ratio of zero (corresponding to a UT/T ratio of 1.0). The data points are colored according to the multiple comparison corrected P-values based upon an

2413

Table I. ANOVA results and CpG Island methylation analysis from LN-18

	Total	P < 0.001 Holm	P < 0.05 FDR	Significant 5 mC				Not Sig
				% Sig	V > 0.75 Dense	0.05 < V < 0.75 Sparse	V < 0 Unmethylated	
CGI	9786	1387	6461	66%				3326
Within probe	2964	662	2232	75%	71	797	1361	732
Adjacent to probe	6822	725	4229	62%	32	1321	2976	2594
					2%	33%	67%	
CGI TSS	8403	971	5510	66%				2893
Within probe	1761	288	1312	75%	19	353	940	449
Adjacent to probe	6642	683	4198	63%	30	1261	2907	2444
					1%	29%	70%	
Total features	21294							
Features considered	21047	2117	11146	53%				9901
TSS	19595	1648	9946	51%	95	3425	6426	9649
BAC annotated CGI	1395	449	1069	77%	75	573	421	326
Other	304	20	131	43%	1	32	98	173
					1%	19%	33%	47%
Not considered	247							

FDR= false discovery rate, Holm = Holm error correction. The significance level for both tests was set at 0.05. CGI = UCSC annotated CpG Island. Adjacent to probe means annotated CGI was located by position ± 500 bp of probe. Dense, Sparse and Unmethylated refer to the array ratio range and the implied 5 mC content based upon the MethylScreen calibration (Figure 5C). TSS = Transcription start site. The methylation values are represented by V and correspond to the corrected signal (RCG normalized $\log_2$(UT/T)), cut-offs set based upon results from Figure 5C. Dense = methylated probe and region, Sparse = either IM probe and region or unmethylated probe adjacent to methylated DNA, Unmethylated= substantially devoid of 5 mC. Features not considered include the 62 used for normalization, and 185 that mapped to multiple loci in the genome.

alpha level of 0.05. Blue data points passed the Holm multiple testing correction and have a *P*-value <0.001. Pink data points passed the FDR correction and have a *P*-value <0.05. Ratios that are not significantly different from the control RCG0 features are depicted in gray. Panel B depicts the same data using the respective *t*-test statistic values (*y*-axis) with the same *x*-axis. The *t*-test data representation allows a visual assessment of the variation on a per feature basis since the t-test values are the $\log_2$ (UT/T) signal divided by the standard error for that probe; it corresponds directly with the significance results of panel A.

Table I summarizes the ANOVA results. Approximately 60 loci were used as RCG0 controls for normalization (Table I). Over half of the 21 047 measurements detected statistically significant DNA methylation levels relative to the controls, with 4152 significant methylated regions (log ratio > 0) and 6994 significantly unmethylated regions (log ratio < 0) (Table I). The existence of unmethylated regions suggests that the controls could in fact detect some methylation, probably because of 'wingspan' from surrounding, RCG containing features.

CpG Islands (CGI) have been characterized as CG rich genomic sequences that do not show CG depletion; they are often in association with genes [(29), for a review see (30)]. CGI are generally thought to be devoid of 5mC, which accounts for the absence of CG depletion which is driven by the mutagenesis of 5mC over evolutionary time, (31–33). Aberrant cytosine methylation of CGI has been associated with inactivation of tumor suppressor genes in human cancers, consequently the DNA methylation status of CGI is of immense interest (5). There are 9787 CGI probes in our design, which reflects roughly 1/3 of the total CGI content in the human genome. The BAC based CGI designed features performed similarly to the UCSC identified CGI; both were present in the same frequency in both the significant methylated and unmethylated loci (Table I).

As expected, very few (1–2%) of the 9786 CGI features reported dense methylation, whether or not they corresponded to known TSS. However, almost 1 in 3 CG islands had low levels of methylation. In these cases, methylation may be confined to sequences near, but not within the CGI. Features in the region would still detect signal because of 'wingspan'. Bisulphite sequencing and other validation methods were used to test this idea, as described below.

Independent verification suggests high accuracy with a quantitative capacity

Independent verification of DNA methylation levels was performed using two methods. The first was clone-based bisulphite sequencing. 13 loci were selected from a range of the microarray data, and 1 locus devoid of RCG motifs was selected. More than 10 clones were analyzed per locus (19). The locus devoid of RCG motifs was substantially methylated, and because no McrBC half-sties were in proximity of the probe it yielded a negative log ratio (Supplementary Table). Of the 13 remaining loci, 11 yielded statistically significant measurements: 6 were methylated and 5 were unmethylated according to the microarray experiments. In 9 of the 11 cases, the methylation occupancy information determined by bisulphite sequencing agreed with the array prediction (Figure 5A) (Supplementary Table). Interestingly, the data from the 9 concordant suggested that there may be a relationship between the array's output ratio value and the average regional 5 mC occupancy (Figure 4A–C). In the two discordant cases, the data were consistent with a wingspan effect upon the probes (Figure 4A, two points off line). Panel B contains bar-charts depicting the average methylation density at each locus. For each locale the *y*-axis depicts the average methylation occupancy at the CGs within the analyzed interval. The CGs are spaced along the *x*-axis by their relative position.

The second verification technique employed a more high-throughput quantitative PCR approach recently developed by
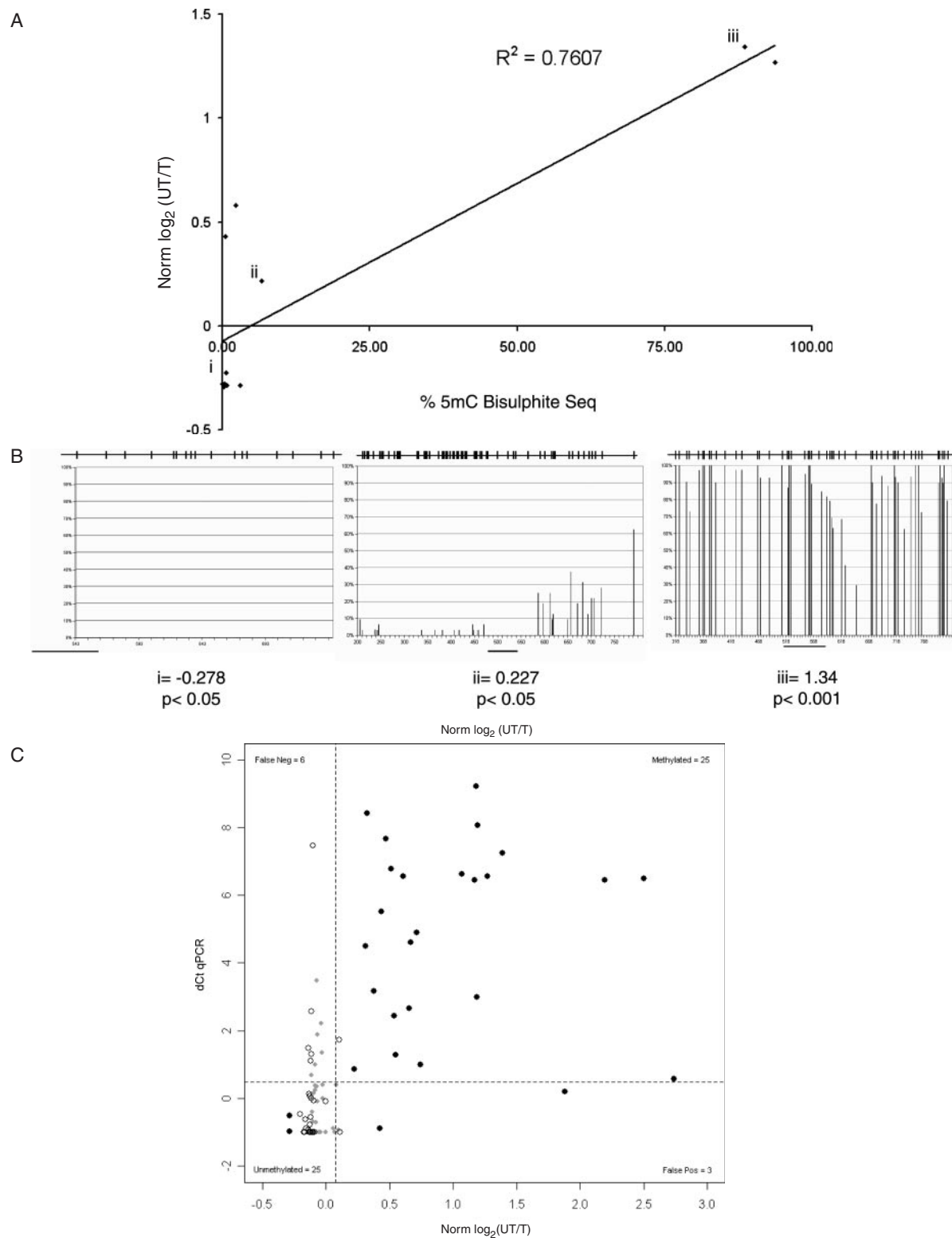
Fig. 4. Independent accuracy analysis suggests quantitative capacity within array analysis. (**A**) A scatter plot of the average 5 mC density determined by bisulphite sequencing obtained from the 11 statistically significant loci are plotted versus the normalized $\log_2$ (UT/T) for each feature. The R^2 value obtained reflects the result of a linear correlation analysis. Two of the 11 were considered to be discordant with the regression line. (**B**) Bar charts depicting high resolution cytosine methylation analysis from loci including the points i, ii and iii. Each of the charts indicate the position of each CG analyzed within each window on the *x*-axis, and the average relative 5 mC occupancy for greater than 10 clones at each site within the locus on the *y*-axis. The ANOVA corrected signal log ratio for each locus is indicated, along with the multiple testing correction *P*-value intervals the measurements fell within, are the inset. The position of the array probe within the interval is denoted by the black bar in each window. (**C**) A scatter-plot of methylation values for 84 genomic loci is shown. Signal ratios from MethylScope are plotted against the change in cycle threshold (ΔCt) measured by MethylScreen quantitative PCR (34). The data are categorized into four quadrants: In clockwise order, false negatives, methylated predictions, false positives and unmethylated predictions. Methylated and unmethylated features are those for which both assays agree. A cut-off of 0.5 cycles is used for the MethylScreen data, as explained in the text (34). The data points are colored according to their significance level: FDR=open circle, Holm=filled circle, grey points are not significant.

our group [MethylScreen, see Materials and methods and (34)]. We randomly selected 50 loci without regard to the magnitude of the microarray measurement, and selected an additional 34 loci representing the range of measurements. We performed MethylScreen upon these 84 loci, 4 of which were included in the bisulphite sequencing study. Figure 4C depicts the log ratio of UT/T (x-axis) plotted against the ΔCt obtained by MethylScreen (y-axis). Like Figure 3A, the data points are colored by their statistical significance level, gray data were not significant. The limit of detection of methylation by MethylScreen was set to a ΔCt of 0.5 cycles. This was because under ideal circumstances, quantitative PCR has a platform error of ±0.5 Ct. Using these criteria, the upper left quadrant represents false negatives in which significantly hypermethylated regions were not detected by the array (6 total). The upper right quadrant represents methylated features (25 total) as determined by both assays. The lower right quadrant represents false positives (3 total). Finally, the lower left quadrant represents unmethylated features (25 total). A total of 25 of the randomly selected loci were no more or less methylated than the controls and so met the null hypothesis (gray). Of these, only 5 were substantially methylated according to MethylScreen, indicating that only a small number of features in this class failed to be detected by the array hybridization, with the remainder being unmethylated. Altogether, the qPCR (MethylScreen) and microarray (MethylScope) approaches agreed 85% of the time. These validation experiments, along with the bisulphite sequencing data, indicate that the MethylScreen and MethylScope technologies are highly reproducible, and support the view that the MethylScope array hybridization approach generates an accurate characterization of genome-wide cytosine methylation. In the few cases where they disagree, the level of methylation was almost always very low, and in only three cases (one false negative and two false positives) did the MethylScope and MethylScreen techniques differ markedly. We cannot exclude the possibility that both 'false positives' were detecting methylation adjacent to the probe.

Simultaneous methylation profile and copy number assessment

LN-18 is homozygous for a deletion of the cell cycle regulatory gene CDKN2A (35), and so we included 161 probes from this region on the array as controls. Close inspection of Figure 3A appeared to indicate there was a lack of hybridization of either the UT or T fractions to the probe elements (black arrow) alphabetically between the CGI features (red line) and the promoter features (black line). The 161 probes dedicated to CDKN2A/B fall into this region. In an attempt to confirm that the array could detect the absence of the CDKN2A locus, we sorted the t-values (signal/error) such that the features were depicted in chromosome order. CH7, CH8 and CH9 are illustrated in Figure 5. Another region of the genome with a similar feature density to that present within the CDKN2A/B region is the HOX gene cluster on CH7, represented by 41 features. The boxes highlight each region in Figure 5. The HOX cluster is hypermethylated while the CDKN2A region on CH9 has no significant signal. The mean of the t-values corresponding to the 161 features from the CDKN2A region is −1.8 with a standard deviation of 1.3, indicating that for the vast majority of the probes the signal (UT/T) was of the same order as that of its standard error, and therefore not above background
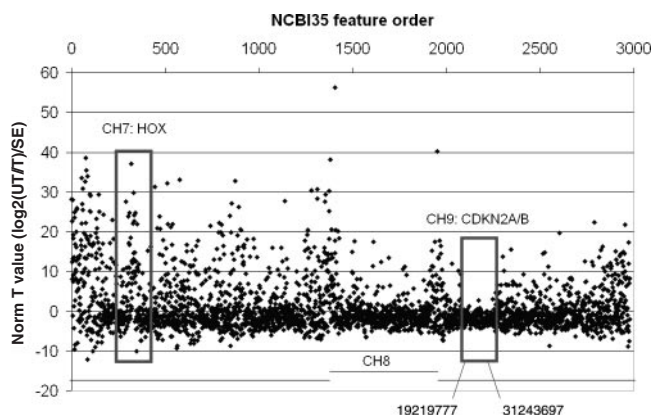

Fig. 5. LN-18 methylation profile reveals ploidy monitoring capacity Methylation data from chromosomes 7, 8 and 9 are represented by the normalized test statistic value t for each feature, aligned by genome co-ordinates on the *x*-axis. Large positive *t*-test values reflect high-density methylation with small variance, while ratios near 1.0 represent low signals with high variance (background signal). The boxes highlight the HOX cluster on CH7, and the CDKN2A/B locus on CH9. LN-18 is homozygous for a deletion of CDKN2A, while the HOX cluster is heavily methylated. The estimated chromosomal positions of the deletion are indicated on CH9.

(Supplementary Data). We estimated the size of the deletion to be <12 Mb; it likely lies between base pair positions CH9: 19219777 and 31243697.

LN-18 displays HOX cluster hypermethylation
The HOX interval is ~165 kb, and our array contains approximately 1 feature per 4 kb. The values from our analysis are indicated as a blue line in Figure 6. Methylation profiling data from the same region was previously obtained from colon cancer and fibroblast cell lines, by using immuno-precipitation with antibodies raised against 5 mC and overlapping BAC clones on the array (36), and these data are indicated as a green line in Figure 6. In this particular region, the BACs were larger than average, with more than 80 kb between the midpoints of the BACS. The TSS for the genes in the interval are depicted as arrows pointed in the direction of their transcription (not to scale). Thus, while both analyses indicate methylation, locally unmethylated regions could also be detected using our higher resolution approach, afforded by a higher density and smaller size of probes used. Unmethylated regions included the TSS from only 3 of the HoxA genes surveyed (see Discussion).

Indications of biological relevance
In an effort to discern whether or not the methylation events detected were biologically meaningful, we ran the 25 quantitative PCR assays that detected regional methylation against a panel of genomic DNAs from six brain cancer derived cell lines, as well as pooled male and female peripheral blood which were collected independently from the tumor origin. We employed the CGI in the TH2B gene as a positive control for single copy gene associated 5 mC detection (37). Among these loci, 8 were differentially methylated in the majority of cell lines relative to peripheral blood (Figure 7). Two of these 8 loci were previously demonstrated to be subject to DNA methylation mediated gene silencing; RASSF1A (38) and E-Cadherin (CDH1) (39). The remaining six genes that displayed
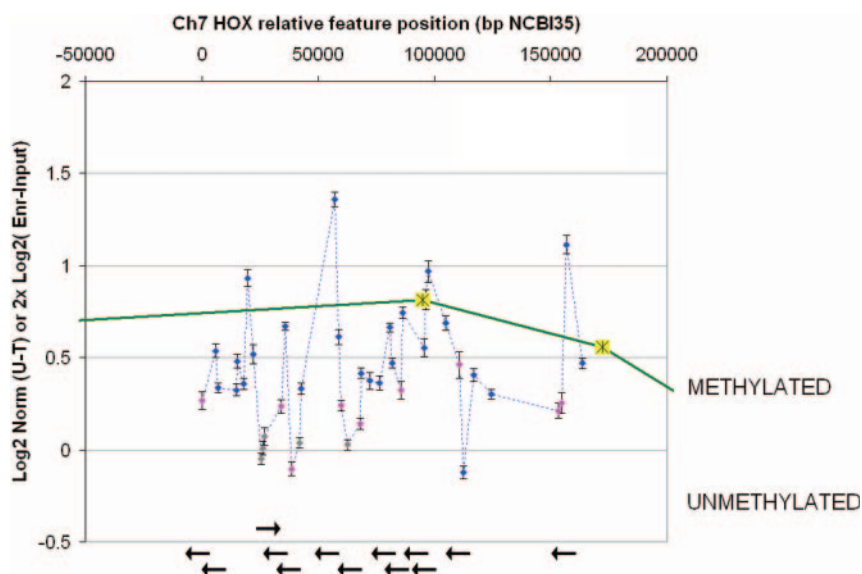
Fig. 6. High resolution methylation profile of the CH7:HOX cluster in LN-18 cells. Methylation in the CH7 HoxA gene cluster. Genes are represented as black arrows (not to scale). The arrows reflect the direction of transcription for the TSS features, their tail is aligned to the end of the feature. Methylation data is from LN-18 cells reported here (blue dashed line) and from immuno-analysis of SW48 cells (green line) reported previously (36) plotted against CH7 (NCBIv35). The average signal ratio from SW48 was normalized by the average normal fibroblast control, scaled for comparison and the midpoint of the BAC insert was assigned the normalized array signal ratio (yellow x-box). The standard error of the LN-18 measurements is indicated by error bars. Features passing the Holm correction are indicated by blue ($P < 0.001$), those passing the FDR correction ($P < 0.05$) are pink. Those whose methylation level was not significantly different from unmethylatable (RCG0) controls are indicated in gray.
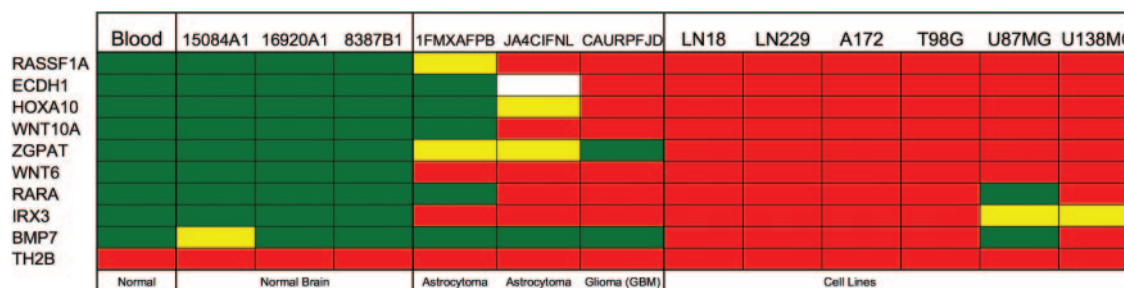

Fig. 7. Discovered differentially methylated loci molecularly classify brain tissues and cell lines. Results from two independent qPCR methylation assays at 10 genomic loci are depicted. Red cells indicate substantial DNA methylation at the genomic locus ($\Delta Ct \gg 2$). Green cells indicate little detectable 5 mC at the locus ($\Delta Ct < 1$). Yellow cells indicate an intermediate amount of methylation ($1 < \Delta Ct < 2$).

hypermethylation relative to blood, and normal brain tissue have not been reported to be subject to epigenetic silencing. These newly discovered methylated loci were associated with the IRX3, BMP7, WNT10A, WNT6, RARα and ZGPAT genes. The same panel of loci was differentially methylated in a comparison of primary brain tissue from three apparently normal patients in comparison to primary tumor tissues derived from two astrocytomas and one glioma (GBM).

IRX3 CGI methylation is correlated with IRX3 overexpression
Interestingly, the methylated CGI detected at the IRX3 locus corresponds to a an exon rather than the gene's promoter (Figure 8). A high resolution analysis of the annotation and DNA methylation data obtained from the MethylScope analysis of the LN-18 genome is depicted in Figure 8A. The promoter of IRX3 was significantly unmethylated, while the CGI in exon 2 was methylated. Bisulphite genomic sequence was obtained from the IRX3 exonic CGI

(Figure 8B). The graph displays the average methylation occupancy per CG in the interval when considering the 38 clones sequenced, the X and Y axes are the same as those in Figure 4B. MethylScreen and bisulphite sequencing analyses demonstrated that this region is unmethylated in normal blood and normal brain DNA (Figure 7 and data not shown). The position of the array feature is denoted by the gray bar in Figure 8B. There are six HhaI restriction sites in the analyzed region, five of which appear to be nearly always occupied. The MethylScreen amplicon designed for the qPCR analysis presented in Figure 7 surveys all six of these HhaI sites. This amplicon was then employed in a MethylScreen assay (Figure 9). The kinetic profiles obtained from each of the four conditions of the assay obtained from each genomic sample are displayed. The data depict results from cell lines (upper row), normal brain tissue (middle row) and three brain tumors (lower row); each genome's results are color coded by their treatment: mock restriction (red), HhaI restriction (green), McrBC restriction (blue) and a HhaI +
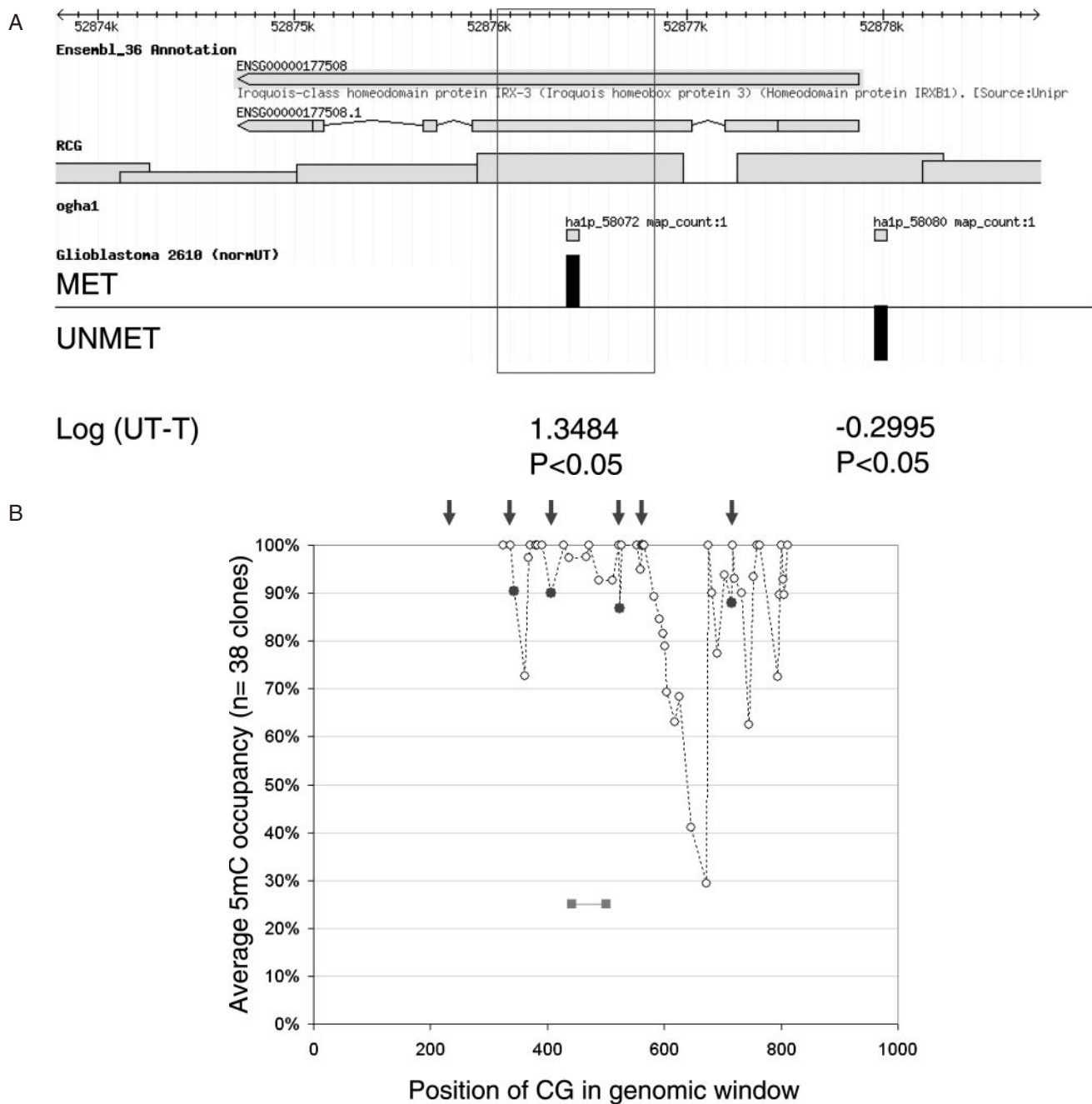
2417

Fig. 8. High resolution analysis of IRX3 hypermethyaltion reveals an exonic rather than promoter location. (**A**) A gbrowse view of the NCBIv35 genome build with the ENSEMBL_36 annotation is depicted along with the LN-18 microarray measurements of local DNA methylation. The arrow scale at the top denotes the base pair position along the chromosome. The first data track contains a bar representing the IRX3 locus, the arrow depicts its direction of transcription. The second track is the splicing model of the transcript as determined by ENSEMBL_36. The third data track (RCG) depicts the relative abundance of methylateable McrBC recognition sites within a defined 1 kb genomic window. The fourth track (OGHA) denotes the positions of the two array features (60mer probes). The fifth track depicts a scaled black bar chart of the ANOVA corrected array measurements for the two features Log_2 (UT/T); the actual measurements along with their *P*-values are indicated below each feature. Positive log ratios indicate 5 mC and negative log ratios indicate its relative absence. (**B**) Bisulphite sequencing results from analysis of the LN-18 methylation pattern surrounding the array feature are depicted as a scatter plot. The *x*-axis depicts the relative position of each base pair within the interval. The average methylation occupancy at each CG is represented as an open circle. The area under the dashed line between points represents the methylation density. Black arrows denote the position of HhaI restriction sites in the interval (one was outside the sequenced region). The black circles represent methylation occupancies of CGs at HhaI restriction sites within the interval. The gray bar is the position of the array probe.

McrBC DD (orange). The pie chart insets display the result of each assay where red = fraction of molecules with all HhaI sites occupied, green = fraction of molecules devoid of RmC methylation between the primers, and the yellow = fraction of molecules susceptible to both treatments (see Materials and methods for calculations). The exon appeared to be hypermethylated in all the tumors and cell lines relative to the three independent normal brain samples, reminiscent of the differentially methylated regions (DMR) in imprinted genes like IGF2 and H19. We therefore
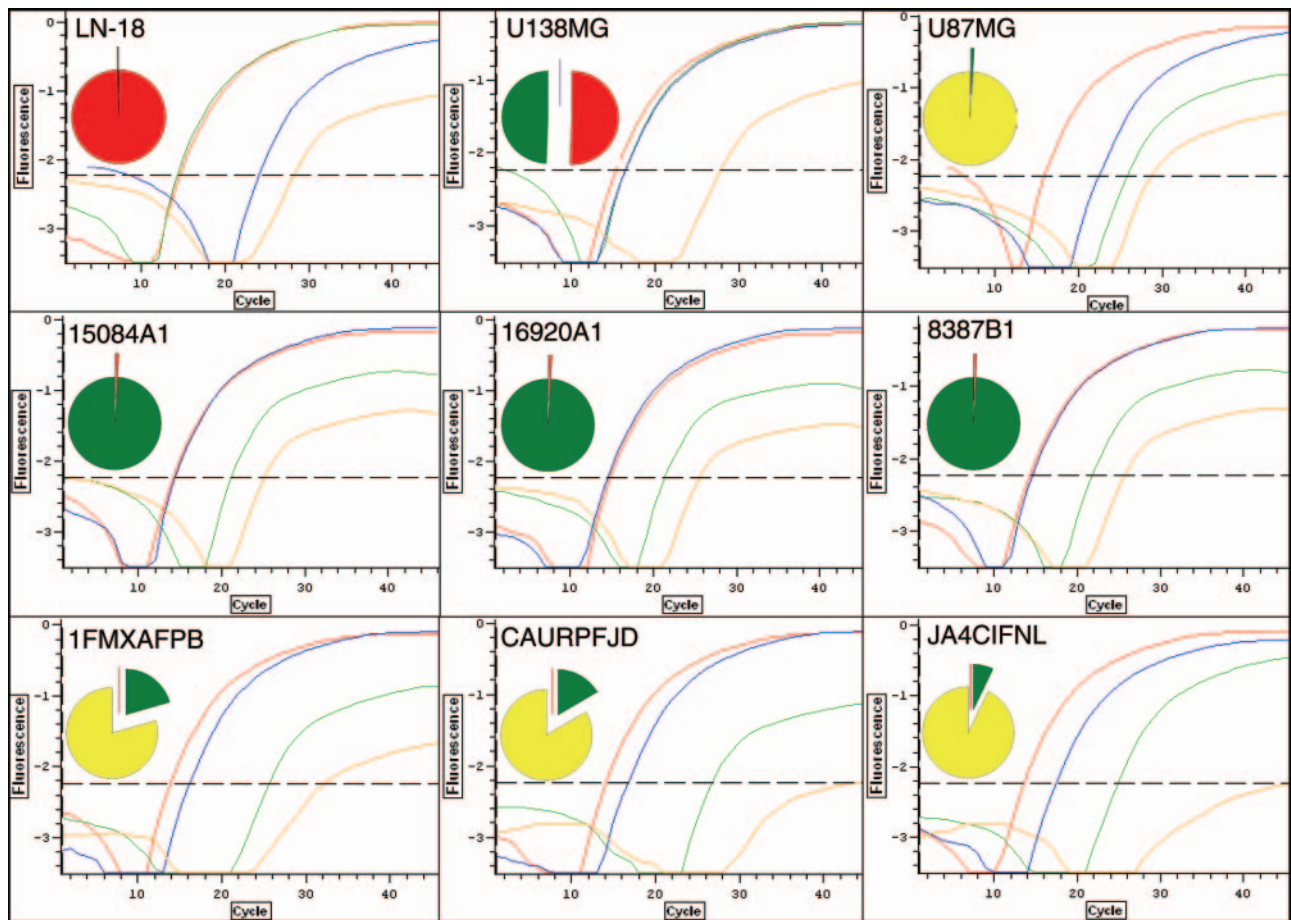
Fig. 9. The IRX3 exonic CGI is hypermethylated in brain cancer cell lines and tumors but not normal brain samples Representative SYBR green quantitative PCR kinetic reaction profiles are depicted for nine templates following four MethylScreen treatments. The top row is brain cancer cell lines. The middle row is normal brain samples. The bottom row consists of two primary astrocytoma and one GBM sample. Within each profile the four treatments of MethylScreen are indicated by colors: Red = mock treatment, Green= HhaI treatment, Blue= McrBC treatment and Orange= DD (HhaI + McrBC). The inset pie-charts reflect the interpretation of the methylation status of the molecular population (See Materials and methods) based upon the average profile data. Red = proportion with uniform HhaI methylation, Green = proportion with low to no methylation, Yellow= proportion IM. All the tumor samples had a McrBC conditioned change in Ct > 2 cycles; indicating the majority of molecules present contain aberrant methylation relative to the normal samples.

hypothesized that IRX3 may be over-expressed in tumors and cell lines (due to its intragenic position) rather than silenced. Since the MethylScreen results from normal brain tissues are not consistent with hemizygous methylation, we do not suspect IRX3 as being imprinted (Figure 9, middle row).

The expression of IRX3 in tumor and normal brain samples was assessed using semi-quantitative PCR amplification of dilutions of oligo-dT primed cDNA libraries. IRX3 expression was measured relative to GAPDH expression as a positive control (Figure 10A). The MethylScreen IRX3 DNA methylation result from each sample is depicted in Figure 10B. The expression amplicon selected for each target spanned an intron of ~100 bp or more to eliminate genomic DNA contamination of the cDNA libraries. In all cases, the level of DNA contamination was negligible (data not shown). Comparison of amplification from each dilution indicated that all three of the tumors expressed between 5- and 8-fold more IRX3 than the normal tissue library (with a SD of 3-fold). Commercially acquired astrocytoma cDNA libraries behaved similarly (data not shown), although matched DNA samples were not available. LN-18 produced more than 20-fold more IRX3 than normal brain, and had the

most methylated DNA (Figures 10A and B, 9). The U87MG cell line expressed ~4-fold less IRX3 than LN-18, but ~5-fold more than the normal tissue, making its expression level most similar to the tumor libraries (data not shown). Interestingly, it was also the sample with a methylation pattern most similar to the tumors (Figure 9).

DISCUSSION

Precise, accurate and quantitative whole genome DNA methylation profiles

DNA methylation is a stable epigenetic signal associated with gene silencing (2), and quantitative measurement of DNA methylation during disease has immense appeal as a diagnostic tool, especially in cancer. Several DNA methylation profiling approaches have been developed (14,36,40–50). Few have demonstrated the ability to identify both methylated and unmethylated sequences with statistical confidence, and fewer yet, have the capacity to monitor the entire genome.

Biologically meaningful changes in DNA methylation typically occur over regions with multiple CG dinucleotides,
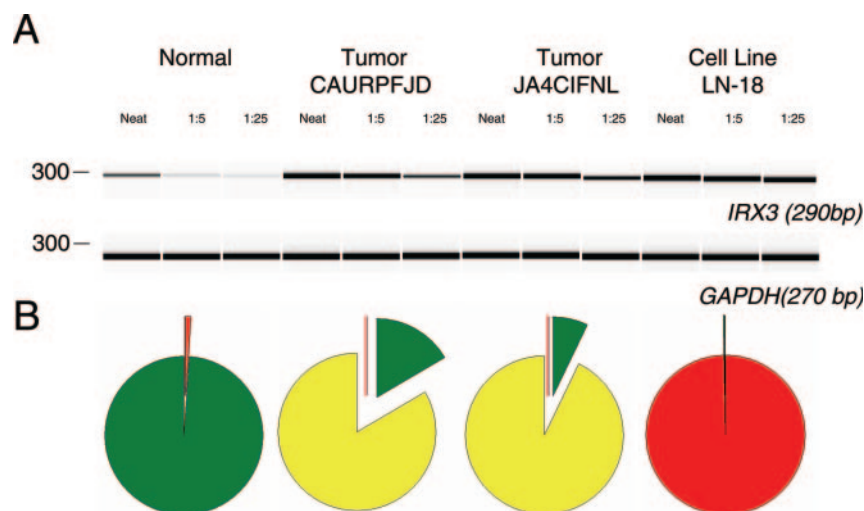
Fig. 10. IRX3 CGI hypermethylation correlates with over-expression (**A**) depicts Bioanalyzer results obtained following quantification of RT-PCR products in reactions from cDNA libraries prepared from normal brain, tumor and cell line controls. The expression of IRX3 was monitored relative to GAPDH as a benchmark. The normal library was purchased from Invitrogen. Neat, 1:5 and 1:25 reflect the cDNA template dilution factor (into water). (**B**) depicts the IRX3 exonic CGI MethylScreen results from Figure 9.

which are methylated in concert. Thus, regions of methylated sequence rather than single bases may be the most important unit of information in the epigenome. The unique enzymatic activity of McrBC is well suited for identifying these regions, as it recognizes at least half of the potentially methylated CG-dinucleotides in the human genome (RCG half sites); no other enzymatic system has such a broad capacity. Further, the methods described here monitor methylation of repeated sequence elements as well as unique sequences. Because each array feature has the capacity to detect methylation adjacent to it, unique features may be employed to monitor adjacent repeat methylation (Figure 1). The first reported epigenetic alteration in cancer genomes was a loss of methylation from repeated elements (51), and such changes in methylation may have profound consequences for the genome as a whole (52). Our approach is also unique in its ability to report the relative density of 5 mC around each feature (Figure 4) as well as to identify changes in regional genomic copy number (Figure 5); although a relatively high probe density was required to observe this effect.

The purpose of this study was to demonstrate that methylated and unmethylated sequences could be detected within the human genome, and to understand the precision and accuracy of the cytosine methylation profiling procedure. Because a duplicated dye-swap experiment was performed within each DNA sample, and each feature was replicated on the array, a stringent measure of precision can be gauged though a correlation analysis between the independent technical replicates. The average correlation (R^2) was 0.77 for LN-18 and ranged between 0.72 and 0.83. Recent experiments involving nearly 50 arrays have displayed R^2 averages of 0.87 and a range of 0.81 to 0.93 (J. M. Ordway, R. M. Kendall and J. A. Jeddeloh, unpublished data). Using two independent validation approaches, we have demonstrated that the MethylScope array hybridization method is robust and accurate. In total, between the two verification approaches, 95 loci were tested, 70 of which were identified as either significantly methylated or unmethylated according to the microarray. The array measurements agreed with the verification measurements

59 times so that the accuracy of the array approach is at least 84% (59/70), assuming that bisulphite sequencing and MethylScreen are almost always accurate. Importantly, the verification of data obtained from measurements which were not statistically significant yielded appropriate results; i.e. these features were predominantly unmethylated as predicted by our analytical hypothesis (20/25). However, future analyses would not be able to identify the true status of these measurements without a second analytical method (i.e. not array only).

These values should be considered a conservative estimate of accuracy because of wingspan effects (Figure 1). The MethylScreen quantitative PCR approach is only sensitive to DNA methylation between the primers while the array itself is capable of reporting methylation of a larger region.

LN-18 epigenetic landscape

Our analysis of the LN-18 genome revealed more than 4000 methylated loci, by design the majority of which were associated with the transcription start sites of human genes (Table I). CpG islands (CGI) were disproportionately unmethylated in agreement with previous reports [for review a see (30)]; however far more were hypermethylated than anticipated if both sparse and dense methylation are considered (Table I). A conservative estimate of the CGI hypermethylation is revealed by the CGI features reporting dense methylation, which were 2% of the total CGI features. However, since approximately half of the features with intermediate array ratios were also methylated within the feature, upwards of 19% (2 + 17%) of the CGI in this genome are methylated (Table I). Perhaps this unexpected finding reflects the more extreme epigenomic representation displayed by cancer derived cell lines.

Importantly, the CGI associated with TSS were most often associated with unmethylated and intermediate array ratios. This observation highlights the importance of their unmethylated status, since half of these intermediate measurements will likely represent unmethylated features with adjacent methylated DNA. Further, it suggests that many more TSS may be susceptible to the spread of silencing

from these proximally methylated elements than originally anticipated (5,30). The use of ultra-high density microarrays with much higher feature densities will permit the more accurate mapping of methylation changes in the future.

Among the independently verified hypermethylated loci that the array discovered were E-cadherin and RASSF1A. These loci were important for two reasons. First, they served as an internal positive control since their ability to become hypermethylated had been shown previously (38,39). Second, the methylation pattern present at the RASSF1A promoter region and the high density feature placement suggested wingspan: As predicted, features some distance from the verified center of methylation, detected methylation proximal to them, though with lower confidence (data not shown). Wingspan was confirmed elsewhere by bisulfite sequencing (Figure 4).

Because a given feature seemed to be capable of detecting adjacent methylation, the quantitative value obtained from any probe represents a complex output. Signal from a feature provides not only data regarding the density of methylation surrounding the feature, but also information about its chromosomal context. Future studies are aimed at analytical methods to de-convolute the contribution of both signals.

Comprehensive nature of analysis supports epigenetic mapping

Recently, the need for large-scale epigenetic mapping of the human genome (The Human Epigenome Project) has been realized in the context of cancer, neurological and other diseases (53). The American Association for Cancer Research, National Human Genome Research Institute, as well as National Cancer Institute have recently called for a Human Epigenome Project to determine genome-wide patterns of DNA and histone modifications. Such a coordinated effort would run in parallel with the Cancer Genome Anatomy Project which will identify genetic changes in cancer cells. Epigenetic profiles are also important in testing stem cell lines for their suitability for therapeutic cloning (54,55). Preliminary efforts to generate such epigenomic maps have included the large scale bisulphite sequencing of several human chromosomal regions (56,57); the analysis of modified histones associated with chromosomes 21 and 22 by chromatin immunoprecipitation (ChIP) chip (58), as well as array-based methylation profiling using plasmid and BAC subclones from the human genome (36,59).

The method we describe here, in addition to detecting biomarkers differentially methylated in disease, has potential to rapidly map quantitative methylation differences throughout the genome at very high resolution, and could make a substantial contribution to epigenomic maps of this sort. The analysis presented here demonstrates that any whole genome effort will need better than 4 kb resolution in order to fully capture the profile of the epigenetic landscape and resolve the wingspan effects. Higher probe density would mitigate any negative effects of this possibility and are readily achievable.

Targets of epigenetic modification identified

Several targets of DNA hypermethylation were identified in the course of this characterization. We have independently confirmed that the testis specific histone 2B (TH2B) locus is methylated in at least 7 non-testes tissues (brain, cervix, ovary, lung, colon, breast and lymphocytes), suggesting conservation of epigenetic regulation from rodents to

humans (Figure 7 and data not shown) (37). This finding makes the locus useful for a positive control. Ten of 13 HOX cluster transcriptional starts sites were in methylated regions in the chromosome 7 HOX cluster (Figures 5 and 6). It appeared that only three TSS in the region were substantially unmethylated, Q96MZ23, HOXA3 and HOXA11 (Figure 6 and Supplementary Table). Members of this HOX cluster have previously been found to be concomitantly hypermethylated in cell lines and tumors (60). These findings suggest that domains larger than individual genes may also be targets of epigenetic alteration. HOX hypermethylation may be general to cell lines other than LN-18, since the SW48 (colon tumor cell line) profile reported by Weber *et al*. (36) has similar characteristics (Figure 6). However, the local highs and lows in the HOX region (Figure 6) are only apparent from our analysis. Interestingly, inappropriate CDKN2A and HOX gene regulation in brain cancer, and altered DNA methyltransferase functions have been mechanistically connected through interactions with the polycomb repressive complex member BMI1 (61).

Finding that 6 of the 8 differentially methylated loci identified were novel hypermethylation targets suggests that the full number of genomic loci susceptible to epigenetic modification is large, and that the majority have yet to be revealed. Four of these genes have been implicated in normal brain development [IRX3 (62), BMP7 (63), WNT6 (64) and WNT10A (65)]. Finally, RARα is a member of a gene family that includes other genes subject to hypermethylation in cancer (66).

If this fraction of differentially methylated sequences (8 of 24) is representative of the whole dataset, we would predict that approximately 1500 loci in the human genome might behave similarly (∼35% of the 4152 methylated loci). This value is in close agreement with an estimate put forth by Baylin and colleagues (43). However, neither estimate considers the likely existence of thousands of loci with the opposite molecular phenotype (unmethylated in tumors).

This study identified a hypermethylated region in exon 2 of the IRX3 gene and coincident upregulation of IRX3 transcription in glial-derived primary tumors and tumor cell lines, relative to normal brain. Given the reported functions of IRX (*iroquois*) family proteins, this finding may have implications for the development of gliomas. IRX family genes encode homeobox transcription factors conserved from nematodes to humans (67). In vertebrates, these factors participate in regulation of proneural genes during early neurulation (68) as well as in antero-posterior and dorso-ventral subdivision of the neural plate [reviewed in (69)]. In these contexts, regulation of IRX3 activity has been shown to be directly responsive to signaling from both the Wnt and sonic hedgehog (Shh) pathways (70,71). Traditionally, high-grade gliomas (including glioblastoma and astrocytoma) have been presumed to arise from glial cells. However, recent studies support a model in which human glioma arises from neural stem-like cells (72–75). Given that Shh functions in maintaining neural stem cell characteristics (76), and ectopic expression of Irx3 in chick embryos is sufficient to induce Shh responsiveness (77), it is tempting to speculate that overexpression of IRX3 may be functionally involved in glial tumorigenesis. Furthermore, the interesting finding of coordinate differential methylation of HOXA10, WNT10A, WNT6 and IRX3 genes in primary glial tumors and tumor cell lines raises the possibility of an underlying epigenetic

signature that decouples these developmental factors from their normal regulatory mechanisms.

The power of genomic survey to detect biomarkers is best illustrated by the exonic CGI within the IRX3 gene. Current views of the importance of epigenetic regulation in cancer are based upon the belief that the regulatory regions of tumor suppressor genes become hypermethylated and/or that the regulatory regions of oncogenes may become unmethylated, and that these alterations lead to changes in expression. While certainly true, IRX3 serves as a reminder of the danger implicit in strictly interpreting rules before the landscape of epigenetic alteration has been fully characterized.

Clearly, comparing methylation profiles of normal brain tissues and primary tumors with this data set would be the optimal means to discover and characterize the number and types of novel hypermethylation and hypomethylation targets. Such studies using our array assay are presently underway.

Supplementary data

Supplementary data are available at *Carcinogenesis* online

References

1. Jackson,A.L. and Loeb,L.A. (1998) The mutation rate and cancer. *Genetics*, **148**, 1483–1490.
2. Laird,P.W. (2003) The power and the promise of DNA methylation markers. *Nat. Rev. Cancer*, **3**, 253–266.
3. DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, **14**, 457–460.
4. Golub,T.R., Slonim,D.K., Tamayo,P., Huard,C., Gaasenbeek,M., Mesirov,J.P., Coller,H., Loh,M.L., Downing,J.R., Caligiuri,M.A. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.
5. Jones,P.A. and Baylin,S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
6. Feinberg,A.P., Ohlsson,R. and Henikoff,S. (2006) The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.*, **7**, 21–33.
7. Genereux,D.P., Miner,B.E., Bergstrom,C.T. and Laird,C.D. (2005) A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc. Natl Acad. Sci. USA*, **102**, 5802–5807.
8. Kumar,S. and Subramanian,S. (2002) Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA*, **99**, 803–808.
9. Rakyan,V.K., Blewitt,M.E., Druker,R., Preis,J.I. and Whitelaw,E. (2002) Metastable epialleles in mammals. *Trends Genet.*, **18**, 348–351.
10. Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33**, 245–254.
11. Kelly,W.K. and Marks,P.A. (2005) Drug insight: histone deacetylase inhibitors—development of the new targeted anticancer agent suberoylanilide hydroxamic acid. *Nat. Clin. Pract. Oncol.*, **2**, 150–157.
12. Lyko,F. and Brown,R. (2005) DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. *J. Natl. Cancer Inst.*, **97**, 1498–1506.
13. Baylin,S.B. (2005) DNA methylation and gene silencing in cancer. *Nat. Clin. Pract. Oncol.*, **2**, S4–S11.
14. Lippman,Z., Gendrel,A.V., Black,M., Vaughn,M.W., Dedhia,N., McCombie,W.R., Lavine,K., Mittal,V., May,B., Kasschau,K.D. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature*, **430**, 471–476.
15. Lippman,Z., Gendrel,A.V., Colot,V. and Martienssen,R. (2005) Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Meth.*, **2**, 219–224.
16. Wolfinger,R.D., Gibson,G., Wolfinger,E.D., Bennett,L., Hamadeh,H., Bushel,P., Afshari,C. and Paules,R.S. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol.*, **8**, 625–637.
17. Bickell,P. and Docksum,K. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. Prentice Hall, pp. 288.
18. Hochberg,Y. and Benjamini,Y. (1990) More powerful procedures for multiple significance testing. *Stat. Med.*, **9**, 811–818.
19. Ordway,J.M., Bedell,J.A., Citek,R.W., Nunberg,A.N. and Jeddeloh,J.A. (2005) MethylMapper: a method for high-throughput, multilocus bisulphite sequence analysis and reporting. *Biotechniques*, **39**, 464, 466, 468.
20. Dila,D., Sutherland,E., Moran,L., Slatko,B. and Raleigh,E.A. (1990) Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *J. Bacteriol.*, **172**, 4888–4900.
21. Sutherland,E., Coe,L. and Raleigh,E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.*, **225**, 327–348.
22. Gast,F.U., Brinkmann,T., Pieper,U., Kruger,T., Noyer-Weidner,M. and Pingoud,A. (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. *Biol. Chem.*, **378**, 975–982.
23. Stewart,F.J. and Raleigh,E.A. (1998) Dependence of McrBC cleavage on distance between recognition elements. *Biol. Chem.*, **379**, 611–616.
24. Panne,D., Raleigh,E.A. and Bickle,T.A. (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolysis. *J. Mol. Biol.*, **290**, 49–60.
25. Stewart,F.J., Panne,D., Bickle,T.A. and Raleigh,E.A. (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.*, **298**, 611–622.
26. Panne,D., Muller,S.A., Wirtz,S., Engel,A. and Bickle,T.A. (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. *EMBO J.*, **20**, 3210–3217.
27. Pieper,U., Groll,D.H., Wunsch,S., Gast,F.U., Speck,C., Mucke,N. and Pingoud,A. (2002) The GTP-dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. *Biochemistry*, **41**, 5245–5254.
28. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
29. Cross,S.H. and Bird,A.P. (1995) CpG islands and genes. *Curr. Opin. Genet. Dev.*, **5**, 309–314.
30. Fazzari,M.J. and Greally,J.M. (2004) Epigenomics: beyond CpG islands. *Nat. Rev. Genet.*, **5**, 446–455.
31. Shen,J.C., Rideout,W.M.,3rd and Jones,P.A. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.*, **22**, 972–976.
32. Shen,J.C., Rideout,W.M.,3rd and Jones,P.A. (1992) High frequency mutagenesis by a DNA methyltransferase. *Cell*, **71**, 1073–1080.
33. Jones,P.A., Rideout,W.M.,3rd, Shen,J.C., Spruck,C.H. and Tsai,Y.C. (1992) Methylation, mutation and cancer. *Bioessays*, **14**, 33–36.
34. Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D., Jones,J., Flick,E., Rholfing,T., Fries,J., Bradford,K. *et al.* (2005) Sorghum genome sequencing by methylation filtration. *PLoS. Biol.*, **3**, e13.
35. Ishii,N., Maier,D., Merlo,A., Tada,M., Sawamura,Y., Diserens,A.C. and Van Meir,E.G. (1999) Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines. *Brain Pathol.*, **9**, 469–479.
36. Weber,M., Davies,J.J., Wittig,D., Oakeley,E.J., Haase,M., Lam,W.L. and Schubeler,D. (2005) Chromosome-wide and promoter-specific analyses

identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*, **37**, 853–862.

37. Choi,Y.C., Gu,W., Hecht,N.B., Feinberg,A.P. and Chae,C.B. (1996) Molecular cloning of mouse somatic and testis-specific H2B histone genes containing a methylated CpG island. *DNA Cell. Biol.*, **15**, 495–504.

38. Hesson,L., Bieche,I., Krex,D., Criniere,E., Hoang-Xuan,K., Maher,E.R. and Latif,F. (2004) Frequent epigenetic inactivation of RASSF1A and BLU genes located within the critical 3p21.3 region in gliomas. *Oncogene*, **23**, 2408–2419.

39. Yoshiura,K., Kanai,Y., Ochiai,A., Shimoyama,Y., Sugimura,T. and Hirohashi,S. (1995) Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl Acad. Sci. USA*, **92**, 7416–7419.

40. Yan,P.S., Chen,C.M., Shi,H., Rahmatpanah,F., Wei,S.H., Caldwell,C.W. and Huang,T.H. (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.*, **61**, 8375–8380.

41. 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.

42. 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.

43. Suzuki,H., Gabrielson,E., Chen,W., Anbazhagan,R., van Engeland,M., Weijenberg,M.P., Herman,J.G. and Baylin,S.B. (2002) A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat. Genet.*, **31**, 141–149.

44. 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.

45. 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.

46. Hou,P., Ji,M., Li,S., He,N. and Lu,Z. (2004) High-throughput method for detecting DNA methylation. *J. Biochem. Biophys. Meth.*, **60**, 139–150.

47. Yan,P.S., Wei,S.H. and Huang,T.H. (2004) Methylation-specific oligonucleotide microarray. *Meth. Mol. Biol.*, **287**, 251–260.

48. Tran,R.K., Henikoff,J.G., Zilberman,D., Ditt,R.F., Jacobsen,S.E. and Henikoff,S. (2005) DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.*, **15**, 154–159.

49. Gao,L., Cheng,L., Zhou,J.N., Zhu,B.L. and Lu,Z.H. (2005) DNA microarray: a high throughput approach for methylation detection. *Colloids. Surf. B. Biointerfaces*, **40**, 127–131.

50. 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.

51. Feinberg,A.P. and Vogelstein,B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, **301**, 89–92.

52. Martienssen,R.A. and Colot,V. (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science*, **293**, 1070–1074.

53. Jones,P.A. and Martienssen,R. (2005) A blueprint for a human epigenome project: the AACR human epigenome workshop. *Cancer Res.*, **65**, 11241–11246.

54. Jaenisch,R., Hochedlinger,K. and Eggan,K. (2005) Nuclear cloning, epigenetic reprogramming and cellular differentiation. *Novartis Found Symp.*, **265**, 107–18.

55. Hochedlinger,K. and Jaenisch,R. (2003) Nuclear transplantation, embryonic stem cells, and the potential for cell therapy. *N. Engl. J. Med.*, **349**, 275–286.

56. Rakyan,V.K., Hildmann,T., Novik,K.L., Lewin,J., Tost,J., Cox,A.V., Andrews,T.D., Howe,K.L., Otto,T. and Olek,A. (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS. Biol.*, **2**, e405.

57. Meissner,A., Gnirke,A., Bell,G.W., Ramsahoye,B., Lander,E.S. and Jaenisch,R. (2005) Reduced representation bisulphite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.*, **33**, 5868–5877.

58. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J.,3rd, Gingeras,T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.

59. Yan,P.S., Perry,M.R., Laux,D.E., Asare,A.L., Caldwell,C.W. and Huang,T.H. (2000) CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. *Clin. Cancer Res.*, **6**, 1432–1438.

60. Shiraishi,M., Sekiguchi,A., Oates,A.J., Terry,M.J. and Miyamoto,Y. (2002) HOX gene clusters are hotspots of de novo methylation in CpG islands of human lung adenocarcinomas. *Oncogene*, **21**, 3659–3662.

61. Lund,A.H. and van Lohuizen,M. (2004) Epigenetics and cancer. *Genes Dev.*, **18**, 2315–2335.

62. Bellefroid,E.J., Kobbe,A., Gruss,P., Pieler,T., Gurdon,J.B. and Papalopulu,N. (1998) Xiro3 encodes a *Xenopus* homolog of the *Drosophila* Iroquois genes and functions in neural specification. *EMBO J.*, **17**, 191–203.

63. Furuta,Y., Piston,D.W. and Hogan,B.L. (1997) Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development*, **124**, 2203–2212.

64. Amura,C.R., Marek,L., Winn,R.A. and Heasley,L.E. (2005) Inhibited neurogenesis in JNK-1 deficient embryonal stem cells. *Mol. Cell Biol.*, **25**, 10791–10802.

65. Kelly,G.M., Lai,C.J. and Moon,R.T. (1993) Expression of wnt10a in the central nervous system of developing zebrafish. *Dev. Biol.*, **158**, 113–121.

66. Cote,S. and Momparler,R.L. (1997) Activation of the retinoic acid receptor beta gene by 5-aza-2′-deoxycytidine in human DLD-1 colon carcinoma cells. *Anticancer Drugs*, **8**, 56–61.

67. Burglin,T.R. (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.*, **25**, 4173–4180.

68. Gomez-Skarmeta,J., de La Calle-Mustienes,E. and Modolell,J. (2001) The Wnt-activated Xiro1 gene encodes a repressor that is essential for neural development and downregulates Bmp4. *Development*, **128**, 551–560.

69. Gomez-Skarmeta,J.L. and Modolell,J. (2002) Iroquois genes: genomic organization and function in vertebrate neural development. *Curr. Opin. Genet. Dev.*, **12**, 403–408.

70. Braun,M.M., Etheridge,A., Bernard,A., Robertson,C.P. and Roelink,H. (2003) Wnt signaling is required at distinct stages of development for the induction of the posterior forebrain. *Development*, **130**, 5579–5587.

71. Kobayashi,D., Kobayashi,M., Matsumoto,K., Ogura,T., Nakafuku,M. and Shimamura,K. (2002) Early subdivisions in the neural plate define distinct competence for inductive signals. *Development*, **129**, 83–93.

72. Galli,R., Binda,E., Orfanelli,U., Cipelletti,B., Gritti,A., De Vitis,S., Fiocco,R., Foroni,C., Dimeco,F. and Vescovi,A. (2004) Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.*, **64**, 7011–7021.

73. Ignatova,T.N., Kukekov,V.G., Laywell,E.D., Suslov,O.N., Vrionis,F.D. and Steindler,D.A. (2002) Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers *in vitro*. *Glia*, **39**, 193–206.

74. Singh,S.K., Hawkins,C., Clarke,I.D., Squire,J.A., Bayani,J., Hide,T., Henkelman,R.M., Cusimano,M.D. and Dirks,P.B. (2004) Identification of human brain tumour initiating cells. *Nature*, **432**, 396–401.

75. Phillips,H.S., Kharbanda,S., Chen,R., Forrest,W.F., Soriano,R.H., Wu,T.D., Misra,A., Nigro,J.M., Colman,H., Soroceanu,L. *et al.* (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*, **9**, 157–173.

76. Machold,R., Hayashi,S., Rutlin,M., Muzumdar,M.D., Nery,S., Corbin,J.G., Gritli-Linde,A., Dellovade,T., Porter,J.A., Rubin,L.L. *et al.* (2003) Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches. *Neuron*, **39**, 937–950.

77. Kiecker,C. and Lumsden,A. (2004) Hedgehog signaling from the ZLI regulates diencephalic regional identity. *Nat. Neurosci.*, **7**, 1242–1249.