



## MeDIP coupled with a promoter tiling array as a platform to investigate global DNA methylation patterns in AML cells

Arzu Yalcin<sup>a</sup>, Clemens Kreutz<sup>b,e</sup>, Dietmar Pfeifer<sup>a</sup>, Mahmoud Abdelkarim<sup>a</sup>, Gregor Klaus<sup>a</sup>, Jens Timmer<sup>b,c,d,e,f</sup>, Michael Lübbert<sup>a</sup>, Björn Hackanson<sup>a,\*</sup>

<sup>a</sup> Division of Hematology/Oncology, Department of Medicine, University Medical Center Freiburg, Germany

<sup>b</sup> Institute for Physics, Albert-Ludwigs-University of Freiburg, Germany

<sup>c</sup> Freiburg Institute for Advanced Studies (FRIAS), Albert-Ludwigs-University of Freiburg, Germany

<sup>d</sup> BIOS Center for Biological Signalling Studies, Albert-Ludwigs-University of Freiburg, Germany

<sup>e</sup> Freiburg Center for Systems Biology (ZBSA), Albert-Ludwigs-University of Freiburg, Germany

<sup>f</sup> Freiburg Initiative in Systems Biology (FRISYS), Albert-Ludwigs-University of Freiburg, Germany

### ARTICLE INFO

#### Article history:

Received 28 July 2012

Received in revised form

10 September 2012

Accepted 16 September 2012

Available online 11 October 2012

#### Keywords:

Acute myeloid leukemia

DNA methylation

Epigenetics

### ABSTRACT

Hypermethylation of CpGs in promoter regions and subsequent changes in gene expression are common features in acute myeloid leukemia (AML). Genome-wide studies of the methylome are not only useful to understand changes in DNA methylation and gene regulation but also to identify potential targets for antileukemic treatment. Here we performed methylated DNA immunoprecipitation (MeDIP) in the AML cell line HL-60 and donor-derived CD34+ cells, followed by hybridization on a human promoter tiling array. The comparative analysis of HL-60 versus CD34+ cells revealed differentially methylated promoter regions including genes that are frequently methylated in AML, such as p15/INK4B, OLIG2, RAR $\beta$ 2 and estrogen receptor. Microarray data was validated by quantitative pyrosequencing. We corroborate previous reports that MeDIP, in our study combined with a promoter tiling array (MeDIP-Chip), is a robust method to identify genes that are differentially methylated in AML cells in a genome-wide manner, and is thus useful to identify new epigenetic targets for therapeutic or prognostic research.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Epigenetic alterations comprise modifications of DNA and histones, resulting in regulation of gene expression due to changes of the chromatin structure [1,2]. DNA methylation predominantly occurs at cytosines in the CpG dinucleotide context. Methylation of CpG islands (CpG dinucleotides in the 5'-untranslated region and the first exon of approximately 60% of all genes) can affect the transcriptional activation of genes. The methyl groups serve as target sites for methylation-dependent repressor proteins that induce transcriptional repression by recruiting co-repressor complexes, histone deacetylases, or histone methyltransferases [3]. In the same way, promoter methylation might prevent transcription factors from accessing their binding sites and thus directly inhibit gene expression resulting in silencing of tumor suppressors [4]. Under physiologic conditions, the proper distribution of DNA

methylation plays an essential role in development, chromosomal integrity, maintenance of gene expression states, and X chromosome inactivation [5,6]. DNA methylation patterns are severely altered in human tumors, including acute myeloid leukemia (AML). These changes include local hypermethylation and global hypomethylation [7,8] of CpG islands, resulting in genomic instability and transcriptional silencing of genes [9–11]. Gene specific patterns of hypermethylated DNA are associated with transcriptional silencing and can correlate with cancer progression, prognosis, and treatment response, thus representing promising biological markers [10,12,13].

AML is a heterogeneous disease characterized by cytogenetic and molecular aberrations. AML also frequently shows silencing of genes (e.g. tumor suppressors) through hypermethylation of promoter CpG islands. The anti-leukemic effect of low-dose DNA demethylating agents supports the hypothesis that dysregulation of tumor suppressor genes via hypermethylation might contribute to the malignant phenotype of AML [14]. However, there is still a lack of robust epigenetic biomarkers to predict response and monitor the clinical course of AML patients treated with drugs such as 5-azacytidine or decitabine. Nonetheless, the expectation that DNA methylation signatures could be used as biomarkers for early

\* Corresponding author at: Department of Internal Medicine, Division of Hematology and Oncology, Albert-Ludwigs-University, Freiburg, Hugstetter Str. 55, 79106 Freiburg, Germany. Tel.: +49 761 270 77150; fax: +49 761 270 77150.

E-mail address: [bjorn.hackanson@uniklinik-freiburg.de](mailto:bjorn.hackanson@uniklinik-freiburg.de) (B. Hackanson).

detection of cancer as well as for prognostic and predictive purposes has led to the development of a number of technologies for DNA methylation analysis [15]. Some of these can be combined with microarrays for detection in order to screen many target sequences in one experiment. These can be classified in: (i) methods using methylation-sensitive (HpaII and NotI) or methylation-specific restriction (McrBC) (e.g. CHARM [16]; HELP [17]), (ii) techniques capturing methylated DNA by means of a recombinant, methyl-CpG-DNA binding protein MBD2-Fc (mCIP [18,19]); MBD-affinity purification [20]), which is used to fractionate genomic DNA depending on its methylation level, (iii) immunoprecipitation-based methods using either 5'-methylcytosine (MeDIP) [21,22] or methyl-binding protein (MBD) domains [23] specific to methylated CpGs to enrich methylated DNA. The application of an anti-5mC antibody for DNA profiling was introduced by Weber et al. [21,22]. The main advantage of capturing methods such as MeDIP is that they are not limited to restriction sites and appears to be an efficient, reproducible method to analyze the methylome of large collections of DNA samples [24–26]. Indeed, one caveat with affinity approaches is that methylated CpG-rich sequences are more enriched than CpG-poor sequences [27], but since CpG islands comprise high CpG density, changes in tumor cells and genes that undergo promoter CpG hypermethylation can easily be detected.

Here, we describe the application of MeDIP combined with a GeneChip tiling promoter array (Affymetrix). This combination of immunoprecipitation of methylated CpGs with a specific antibody and a promoter array (MeDIP-Chip) that consists of multiple short (20–25 bp) probes per target is a useful technique for identification of genes with hypermethylated CpG islands.

## 2. Material and methods

### 2.1. CD34+ cell purification, DNA preparation and *in vitro* methylation

Following G-CSF (granulocyte-colony stimulating factor, Amgen, Germany) stimulation and leukapheresis, bone marrow derived CD34+ cells were isolated using magnetic-activated cell separation (MACS, Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's recommendation. Genomic DNA from the human leukemia cell line HL-60 [28,29], CD34+ cells and peripheral blood mononuclear cells of healthy donors was extracted using the DNeasy blood & tissue kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). All donors had provided ethic board approved informed consent. The DNA of CD34+ human progenitor cells was pooled after extraction to reduce inter-individual epigenetic variability (for microarray data evaluation). The DNA was sonicated to obtain fragments ranging between 300 bp and 1000 bp using a Bioruptor (Diagenode, Liège, Belgium). For the dilution series experiment whole-genome amplification was used to generate an unmethylated copy (0% methylated) of genomic PBL DNA according to the manufacturer's protocol (REPLI-g Mini Kit, Qiagen, Hilden, Germany). The amplified unmethylated DNA was treated with CpG-methyltransferase M.SssI (NEB, Frankfurt, Germany) for 6 h at 37 °C [30] to add methyl-groups to all cytosine residues within CpG dinucleotides to generate 100% methylated genomic DNA, followed by DNA purification with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Methylated DNA (100%) was diluted with unmethylated (0%) DNA to create a dilution series of 100%, 50%, 30%, 20%, 10% and 0% methylation. The methylation level of the dilution series was verified by pyrosequencing for ten genomic loci.

### 2.2. Methylated DNA immunoprecipitation (MeDIP) assay

Methylated DNA immunoprecipitation was performed as described by Weber et al. [21]. MeDIP was performed on technical replicates (*n*), (HL-60 *n* = 7; CD34+ (pooled DNA of 4 donors) *n* = 7; 0% to 100% *n* = 3). Briefly, 4 µg of sonicated denatured DNA was incubated with 10 µg of mouse monoclonal antibody against 5-methylcytosine (Eurogentec, Seraing, Belgium) in 10 × IP buffer (100 mM Na-Phosphate pH 7.0, 1.4 M NaCl, 0.5% Triton X-100) for 6 h at 4 °C. Antibody-bound DNA was collected with 80 µl of Dynabeads with M-280 sheep anti-mouse IgG (Invitrogen, Karlsruhe, Germany) for 2.15 h at 4 °C on a rotating wheel and washed twice with 1 × IP buffer (10 mM Na-Phosphate pH 7.0, 0.14 M NaCl, 0.05% Triton X-100). The beads were resuspended in 250 µl Proteinase K buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS, 70 µg proteinase K) and incubated for 5 h at 50 °C. DNA was extracted by standard phenol/chloroform procedure and isopropanol precipitated.

### 2.3. Quantitative PCR

For the comparison of DNA methylation of HL-60 and CD34+ cells, quantitative real-time PCR was used to verify the amount of enrichment for promoter regions of *C/EBPA*, *RARβ2*, *OLIG2* and *GSTP1* in the DNA fraction. qPCR was performed with the LightCycler 480 Real-Time PCR System, and the reaction mix contained 1 × SYBR green master mix (Roche, Mannheim, Germany) and 0.5 mM of forward and reverse primers, respectively, in a volume of 10 µl. PCR cycling consisted of 95 °C for 10 min, 45 cycles of 95 °C for 10 s, 60 °C for 10 s and 72 °C for 20 s, followed by a melting curve analysis. Primer sequences for qPCR are shown in [Supplementary Table 2](#).

### 2.4. Whole genome amplification and promoter array hybridization

Immunoprecipitated DNA from MeDIP was amplified with the REPLI-g FFPE Kit (Qiagen, Hilden, Germany). The amplified DNA was purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). 7.5 µg of amplified DNA was fragmented and labeled according to Affymetrix Chromatin Immunoprecipitation Assay Protocol.

DNA samples were hybridized to the array using Affymetrix hybridization reagents, and a hybridization oven 640. Subsequently, the arrays were stained and washed using the Fluidics Station 450 according to the manufacturer's protocol (Affymetrix, High Wycombe, UK). Arrays were scanned with the GeneChip Scanner 3000 7G. Raw data was extracted with the GeneChip Operating System (GCOS) software from Affymetrix. The Affymetrix GeneChip Human Promoter 1.0R Array comprises over 4.6 million probes tiled through over 25,500 transcription start sites. The 25-mer probes are tiled at an average resolution of 35 bp. Each promoter region covers approximately 7.5 kb upstream through 2.45 kb downstream of 5' transcription start sites (<http://www.affymetrix.com>).

### 2.5. Data analysis

Raw data were imported into the Partek Genomic Suite Software (<http://www.partek.com/>) [31,32]; and preprocessed with the robust multichip averaging (RMA) algorithm [33] that includes background correction, quantile normalization and log<sub>2</sub>-transformation.

A principal component analysis (PCA) was performed as a quality control, to identify potential outliers and evaluate whether these significantly affect the data. For analysis of variance (ANOVA) and

single-sided *t*-test, a significance level  $\alpha = 0.05$  has been used, i.e. *p*-values  $p < 0.05$  have been considered as significant. The trimmed mean of the *t*-values of nearby probes (to reduce the effects of noisy probes) is calculated by averaging probes next to each other, excluding the highest (we defined as 10%) and lowest values (trimming). The MAT (Model-based Analysis of Tiling-arrays) algorithm [34] uses this trimmed mean in a window of fixed genomic length (here 600 bp) to generate MAT scores. If the MAT score of a window or a compound of windows that are near to each other is significant based on *p*-value cutoff and background distribution, then it is reported. The MAT algorithm removes systematic errors in the data to elucidate the real biological signals in the microarray data [34]. To detect significant hypermethylated regions, in the comparison of the two cell lines, the mean signal from each probe for the HL-60 cells was subtracted from that of the CD34+ cells for all probes on the microarray. These detected regions were annotated to their corresponding genes using the Affymetrix U133.Plus.2 Expression Array Probeset ID annotation file. To visualize differences in methylation at specific promoters, Partek Genomic Suite software was used to create custom tracks (.bed files) for visualization in the University of California at Santa Cruz (UCSC) Blat genome browser (<http://genome.ucsc.edu/>). Alternatively, for more detailed analysis, the Bioconductor rMAT package was used for preprocessing [35]. The fold-changes HL-60 minus CD34 have been calculated at the log<sub>2</sub>-scale. A regularized *t*-test was applied to test for significant differences between cell line and CD34+ cells. To stabilize the *t*-test, the population variance estimated from all probes and of the variance estimates for the individual probes were averaged. The probe sequences as well as their genomic locations were obtained from the Affymetrix bmap Hs.PromPR.v02-3.NCBIv36 file. The location of CpG islands was downloaded from the UCSC goldenpath genome database (version hg18, Build 36.1).

Because the DNA fragments obtained after sonication are larger than the probe sequences, also methylated CpGs close to the genomic positions matching to the complementary probe sequences can be detected by immunoprecipitation. This effect has been accounted for by assuming a typical distribution [27] of the DNA fragment length. The distribution has been translated to a smoothing kernel (600 bp) to calculate the effective CpG number around genomic positions complementary to the probe sequences, so that the probes on the tiling array were subdivided based on their CpG-density.

## 2.6. Pyrosequencing

Pyrosequencing analysis was performed to validate the DNA-methylation differences measured by the microarray. Bisulphite treatment of the DNA was conducted using the EZ DNA Methylation-Gold Kit according to the manufacturer's protocol (Zymo Research, Freiburg, Germany). Pyrosequencing primers were designed using Pyrosequencing Assay Design Software. Sequences of PCR primers, genomic location of the bisulfite pyrosequencing assay and the number of investigated CpG sites in each assay are shown in [Supplementary Table 3](#). A universal tag was placed on either the forward or the reverse primer (depending on the strand to be sequenced), and a universal biotinylated primer was used for all reactions [36]. Pyrosequencing was performed using PyroMark Gold Q96 Reagents and the Pyro Q-CpG software on a PyroMark Q96 MD pyrosequencer (Qiagen, Hilden, Germany) according to manufacturer's recommendation. Pyrosequencing analyses were performed in triplicates and data were analyzed using the Pyro Q-CpG software. All assays were optimized with fully methylated (100%) and unmethylated (0%) genomic DNA. Methylation levels were calculated relative to the average number of CpG sites within each assay.

## 3. Results

### 3.1. 5mC antibody immunoprecipitation enables selective enrichment of methylated DNA

To test the selective affinity of the 5'-methylcytosine (5mC) antibody we performed immunoprecipitation (IP) on HL-60 and CD34+ DNA and compared the pull-down with a mock antibody (IgG). We analyzed the 5'-methylcytosine and IgG antibody-precipitated DNA for enrichment of genes of known highly methylated promoter regions of *RARβ2* [37], *C/EBPA* [38] and *OLIG2* [39] and one low methylated gene *GSTP1* [40] in HL-60 by DNA-qPCR (Fig. 1A). We detected low enrichment (4.2 fold) for *GSTP1* with the 5mC antibody. However, high enrichment (> 17 fold) was achieved for *RARβ2* (17.6 fold), *C/EBPA* (29.1 fold) and *OLIG2* (29.6 fold). The enrichment increased with CpG content in the promoter. In contrast, very low enrichment (from 0.02 fold to 0.05 fold) was detected for DNA immunoprecipitated with the IgG antibody, implying selective affinity of the 5mC antibody.

In order to test the efficiency of enrichment of the 5mC antibody, we applied a dilution series of in vitro methylated peripheral blood mononuclear cell DNA (0, 10, 20, 30, 50 and 100% methylated). The enrichment after immunoprecipitation cannot be seen as a function of density of methylated CpGs [22] referred to the dilution series as these samples were made from different ratios of 100% artificially methylated DNA and completely unmethylated DNA. The IP DNA was examined for the genes *GAPDH* and *TSH2B* by DNA-qPCR (Fig. 1B). As the samples, starting from 10% methylated DNA sample comprise uniformly in vitro methylated DNA, we expect the CpGs of the genes *GAPDH* and *TSH2B* to be methylated. At 30% methylated DNA, the enrichment increased to 25% for *GAPDH* and 33% for *TSH2B*, respectively. At 100% methylated DNA, the yield increased to 100% for *GAPDH* and 92% for *TSH2B*, respectively. For methylation levels below 30%, the detected enrichment was below 2%, indicating a reliable immunoprecipitation.

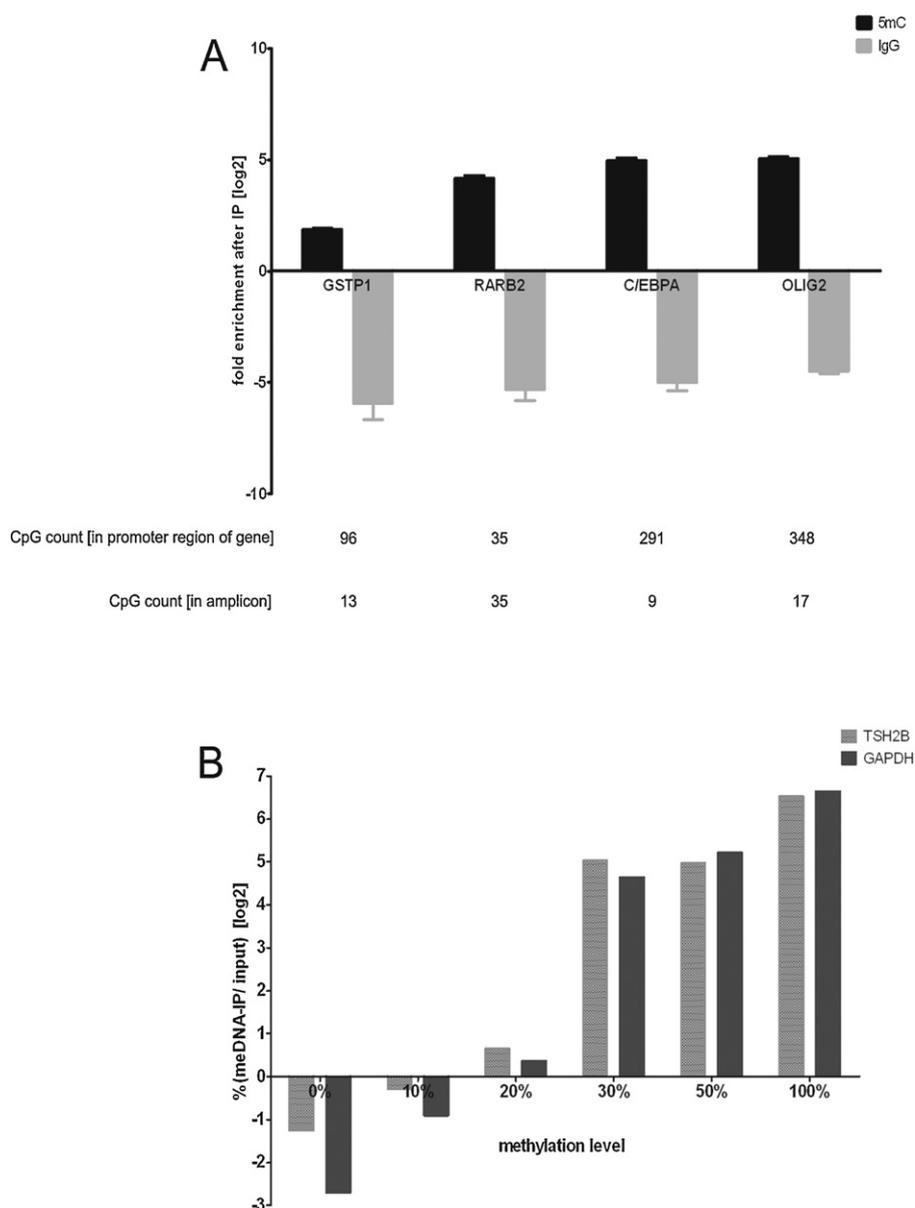
### 3.2. Principal component analysis (PCA) of microarray data separates HL-60 cells from CD34+ cells

In a first step to compare meDIP samples of HL-60 and CD34+ cells, a principal component analysis was performed on the microarray data. Principal component analysis is a multivariate procedure reducing the dimension of the data to the components retaining the bulk of the signals' variance. The difference in signal intensities of the probes binding hybridized IP DNA separated HL-60 and CD34+ cells into two groups (Fig. 2A). The grouping of spheres of HL-60 and CD34+ cells shows the similarity within the data set of technical replicates of HL-60 and CD34+ cells, respectively. Furthermore, the partition into two groups reveals that the DNA methylation profile of both cell lines is different. Two replicates of HL-60 spread considerably apart from the remaining replicates of the cluster. We performed further analysis with and without these two replicates, respectively and regarded them as no outliers, because they did not substantially affect the results.

The PCA of three technical replicates of the in vitro methylated DNA dilution series underlines the separation in different groups and thus the difference in the signal intensities of the probes derived from the precipitated DNA (Fig. 2B).

### 3.3. MeDIP preferentially enriches highly methylated and CpG-dense DNA regions

To determine whether the immunoprecipitated DNA consists of CpG-dense regions, we divided probes covered by the tiling array into three groups, based on their CpG-density, as described in the methods: the "low CpG" group consists of the 10% of probes with



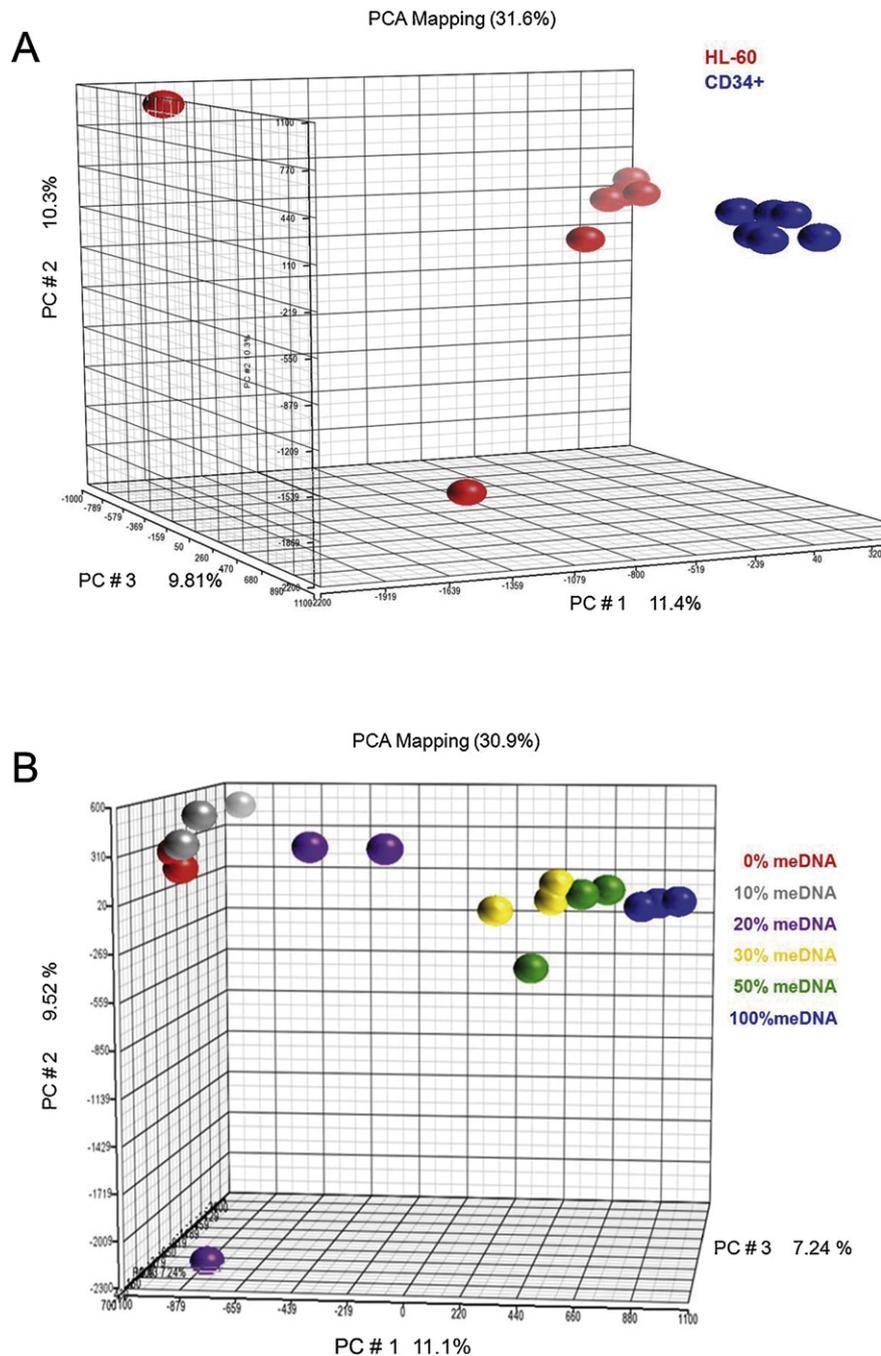
**Fig. 1.** Bar graph representing qPCR validation results for enriched DNA. (A) Fold enrichment of candidate genes enriched by MeDIP. *RARB2*, *C/EBPA*, *OLIG2* (all known to be highly methylated in HL-60) and *GSTP1* (known to be unmethylated in HL-60) of precipitated DNA with specific antibody as compared with IgG pull-down quantified by qPCR relative to input DNA. Graphs represent the average values  $\pm$ S.D. of three replicates on a log<sub>2</sub> scale. The CpG count according to the gene promoter region and the analyzed amplicon are given below. (B) Recovery rate [%] of TSH2B and GAPDH target sequences for in vitro methylated (0–100%) DNA quantified by qPCR relative to input DNA. Graphs represent the values of one qPCR on a log<sub>10</sub> scale.

the lowest CpG density, the “medium CpG” group consists of 80% of the probes with medium CpG density and the “high CpG” group consists of the 10% of the probes with the highest CpG density (Fig. 3). We then quantified the mean signal raw intensity in each CpG group for the dilution series relative to the unmethylated (0%) group. Normalization is not feasible for a dilution series because global different intensities are expected for different dilution settings which are not intended to be removed by a normalization step. As we do not expect chromosome dependency for MeDIP of in vitro methylated dilution series, we performed this analysis exemplarily on chromosome 10.

Probes with low CpG density show minor signal intensity differences (Fig. 3) for 10% (log<sub>2</sub> fold change:  $-0.03$ ) through 30% methylated dilution series. An increase was detected for 50% (log<sub>2</sub>:  $0.17$ ) and 100% (log<sub>2</sub>:  $0.21$ ) methylated samples. The mean signal intensity difference for probes with medium CpG already started

to increase at 20% (log<sub>2</sub>:  $-0.024$ ) and continued to 100% (log<sub>2</sub>:  $-0.69$ ). For probes with high CpG density we detected a constant increase between 20% (log<sub>2</sub>:  $0.327$ ) and 100% (log<sub>2</sub>:  $1.68$ ) of mean signal intensity differences for probes with high CpG density. Taken together, an increase of the CpG density of probes leads to a consistent increment of the mean intensity difference. With increasing methylation level of the in vitro methylated samples, the mean signal intensity differences are more prominent.

We additionally calculated the fold-change of signal intensities of individual probes in order to compare methylated DNA versus non-methylated DNA at the level for single probes of chromosome 10 (Fig. 4). The fold changes of probes with low CpG density are similarly distributed for 20% and 30% (45% of the probes) and also for 50% and 100% methylated dilution series (about 65 and 68% of the probes, respectively). However, samples with 50 and 100% methylation revealed more positive fold changes (Fig. 4A).

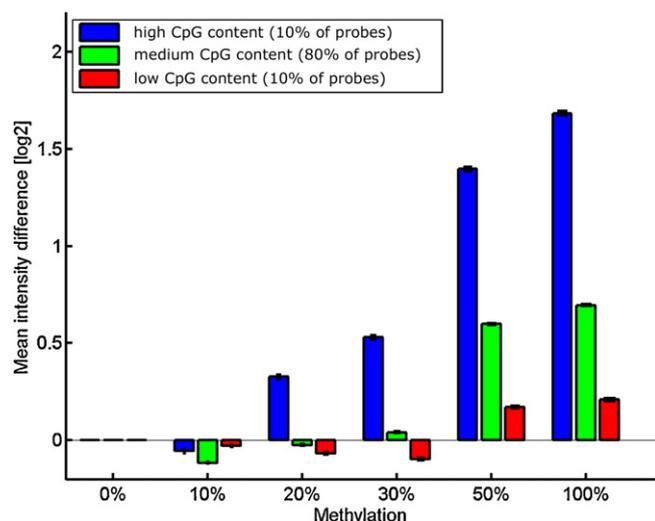


**Fig. 2. Principal component analysis (PCA) of microarray data.** PCA has been applied to microarray data of (A) HL-60 ( $n=7$ ; blue), CD34+ ( $n=7$ ; red) and (B) in vitro methylated DNA dilution series ( $n=3$  for 10%, 20%, 30%, 50% and 100%;  $n=2$  for 0%), respectively. The data for each sample type is represented by different colors and each replicate is presented by a sphere of the same color. Replicates of 0% 10%, 30%, 50% and 100% methylated DNA as well as CD34+ show clear grouping pattern. (For interpretation of the references to color in the artwork, the reader is referred to the web version of the article.)

The fold-change distribution of probes with medium CpG density are shifted to positive values in 50 and 100% methylated samples (about 85% of probes with medium CpG density) as compared to 30 and 20% (about 53% and 48% of the probes, respectively) and are more distinguished between the samples. The differences in the number of probes and fold changes between 20% and 30% as well as for 50% and 100% methylated dilution series (fold change ranges roughly from  $-2$  to  $3$ ) are visually detectable (Fig. 4B). The distances between the fold change distributions of the samples become more distinct for probes with higher CpGs density. The majority of probes for 20% methylated dilution series (79% of probes with high CpG density) and for 30% methylated dilution series (88% of probes

with high CpG density) show a shift to positive fold changes, from  $-1$  to  $2$  and  $<3$ , respectively. For 50% and 100% methylated dilution series all probes with high CpG density (100%) shifted from  $-0.5$  to positive fold changes  $>3$  and  $4$ , respectively (Fig. 4C).

Next, we calculated the  $p$ -values by a two-sample  $t$ -test for each probe (for chromosome 10) (Fig. 4D–F). For 20% and 30% methylated dilution series the  $p$ -values of the probes with low and medium CpG density are uniformly distributed, indicating no significance in detecting methylation differences (Fig. 4D and E). However, for 20% and 30% methylated dilution series 2% of the probes with high CpG density are significant ( $p < 0.01$ ) (Fig. 4F). The  $p$ -values for 50% and 100% methylated dilution series increased substantially for probes



**Fig. 3. Signal intensity histogram of probes according to their CpG density.** CpG density analysis of the probes on the array was realized by counting the number of subsequent GC or CG base pairs in the probe sequences and classification into three different groups. The y-axis of the bar plot shows the mean intensity difference of the probes with 0% methylation on a log<sub>2</sub> scale. The three colors representing high, medium and low CpG density according to the samples hybridized on the microarray. For low CpG density in the DNA sequence at the probe locations, there is only a slight increase in the signal level (red bars). The highest sensitivity is observed for high CpG abundance. The blue bars indicate the measurements for the probes with most CpGs. A difference of one at the vertical axis corresponds to a factor of two in concentration. For the illustration purpose, only the probes on chromosome 10 have been analyzed. (For interpretation of the references to color in the artwork, the reader is referred to the web version of the article.)

with low, medium and high CpG density. The number of significant probes ( $p < 0.01$ ) rises with increasing CpG density of probes and DNA methylation level of the dilution series (from 8% to 39% for 50% methylated DNA and 5% at medium CpG density to 27% at high CpG density for 100% methylated DNA).

To confirm that MeDIP is able to identify global hypermethylation of HL-60 cells compared to CD34+ cells, we plotted the fold-change distribution of the significant probes ( $p < 0.01$ ) comprising the total number of probes on the array (Fig. 5A). The outcome is compared to a randomized data set. Here, the group labels are randomly drawn for each probe to calculate a null distribution, i.e. the distribution of the fold change estimates for probes equally methylated in HL-60 and CD34+ cells. The number of significant probes with positive fold change is considerably higher than for those ones with negative fold change indicating that there are more genes methylated in HL-60. Because the null distribution is located between  $-2$  and  $2$ , significant probes with a log<sub>2</sub>-fold change  $> 2$  indicate an enrichment of methylated DNA in HL-60 compared to CD34+ cells. Depending on the HL-60 and CD34+ dataset, the  $p$ -values are smaller than for the null distribution (Fig. 5B), showing that the cutoff for testing differences is between 0.01 and 0.05 and indicating that there is a global effect in the samples. The flatness of the null distribution and of the right tail of the observed distribution confirms that the shift toward zero is not due to variance in homogeneities and that  $t$ -test is applicable.

### 3.4. Pyrosequencing analysis validates MeDIP-Chip results

The MAT algorithm is used to detect enriched regions on the IP hybridized microarrays. A  $t$ -statistics was calculated for each probe. The resulting list of data comprises chromosome, start and end of detected enriched regions and  $p$ -value. The analysis of the HL-60 vs. CD34+ experiment by the MAT algorithm revealed 6832 detected regions with a positive MATscore (ranging from 4.6 to 51.5) and

8985 regions with a negative MATscore (ranging from  $-15.5$  to  $-3.1$ ) (Supplementary dataset 1).

To evaluate the hypothesis that a positive MATscore reflects an enrichment of methylated DNA regions in HL-60 cells and a negative MATscore indicates an enrichment in CD34+ cells, we selected genes with both, positive and negative MAT scores (Supplementary Fig. 1) and performed bisulfite pyrosequencing to quantitatively measure the methylation level of the included CpGs. The promoter regions of *OLIG2*, *SOCS1*, *ESR1*, *BMPER*, *CDKN2B* (*p15INK4B*), *CYP26C1*, *RARβ2*, *CYFIP1*, *WNT5A*, *GUCY1B3* and *C/EBPα* with a positive MATscore were significantly hypermethylated in the HL-60 cells compared to healthy donor CD34+ cells (Fig. 6A). The promoter regions of *BCL2*, *IGF1-R* and *SYTL3*, with a negative MATscore showed significant hypermethylation in CD34+ cells as compared to HL-60 cells (Fig. 6B). However, the level of the MAT score did not correlate with the level of methylation. In total, 21 of 23 (91%) analyzed promoter regions with a positive MATscore verified that these CpGs were hypermethylated in HL-60 in contrast to CD34+. Regions with a negative MATscore showed a hypomethylation of CpGs in HL-60 in 4 of 8 genes (50%), while sequenced amplicons of four genes revealed equally low DNA methylation status in HL-60 and CD34 and two amplicons (*NAT2* and *ID3*) showed hypomethylation of analyzed CpGs in CD34 compared to HL-60 (Fig. 6 and Supplementary Table 1).

Thus, pyrosequencing analysis confirmed that the 5mC antibody enriched highly methylated DNA of HL-60 and CD34+, respectively and that the MAT algorithm is a robust tool to detect differentially methylated genes when comparing two distinct cell populations.

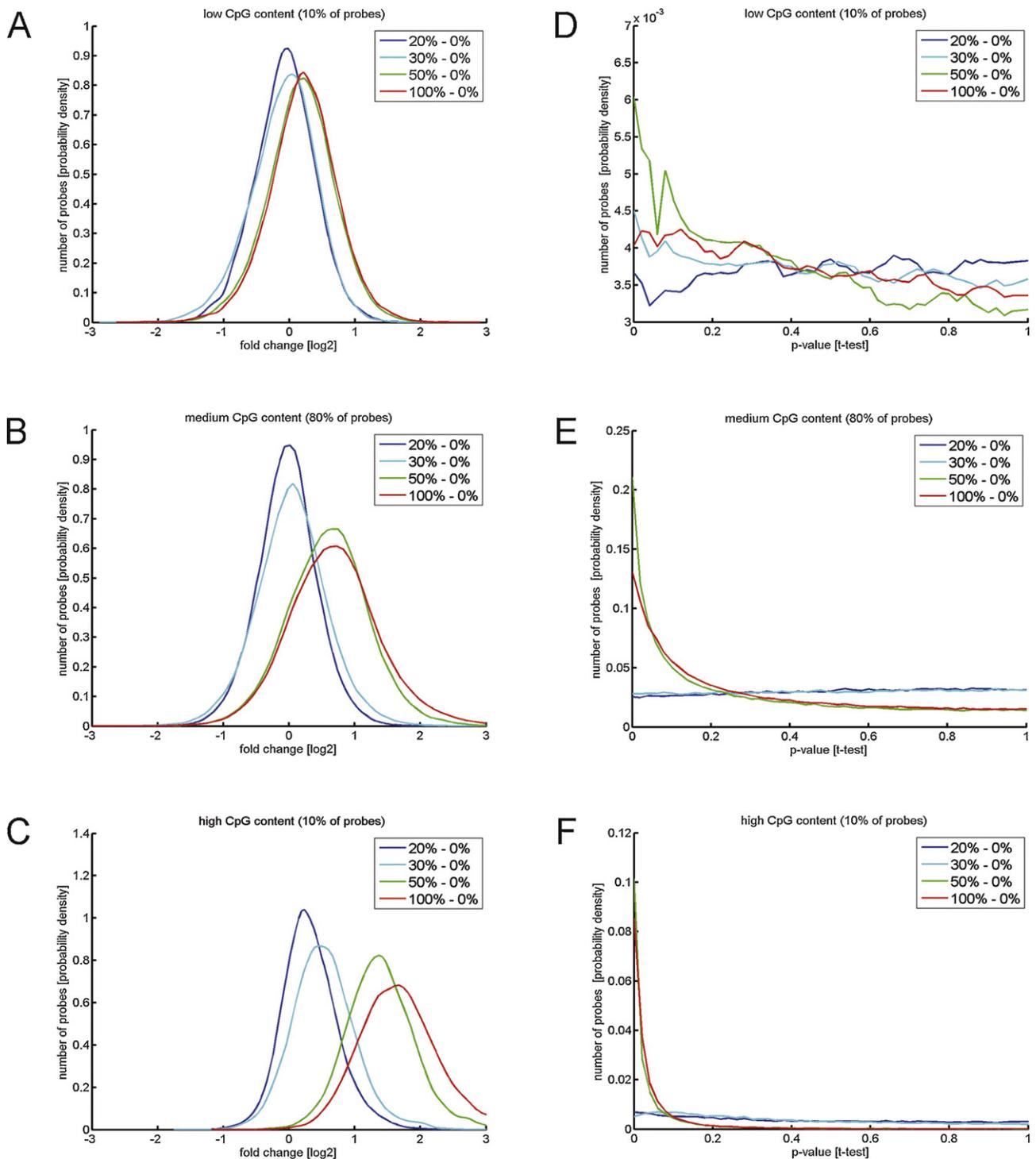
## 4. Discussion

Several methods are currently used to study global DNA methylation differences. Using the HELP assay, Figueroa et al. identified biologically and clinically relevant AML subgroups and described a 15 gene DNA methylation-classifier capable of predicting overall survival, demonstrating the potential of epigenetic markers for use in patients for whom clinical biomarkers are not currently available [41]. Moreover, that study demonstrated DNA methylation-specific AML subtypes, which showed a characteristic epigenetic signature when compared with normal bone marrow CD34+ cells, a comparison frequently used, also because of the homogeneity in contrast to unseparated whole bone-marrow.

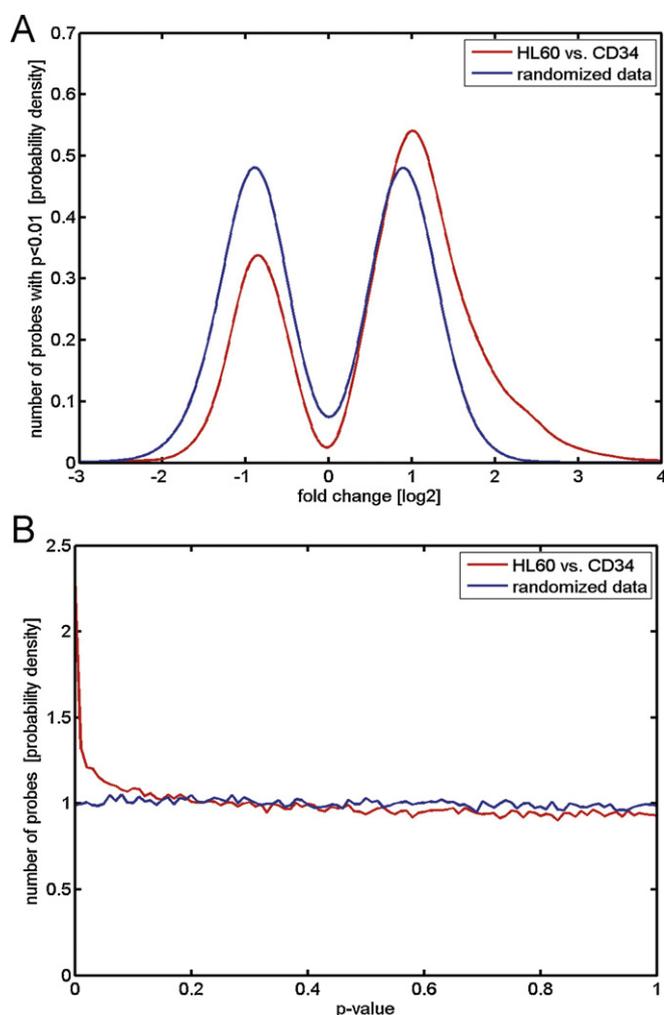
The use of methylation-sensitive restriction enzymes (Hpa II and Not I) as used in the HELP-assay [41] or methylation-specific (McrBC) limit these approaches to profile genomic regions that contain these restriction site motifs [17,42]. The immunoprecipitation of methylated DNA is not dependent on the genomic distribution of restriction enzymes recognition sequences and utilizes single stranded and fragmented DNA making this method less biased and potentially compatible with archival specimens.

Here we demonstrate a genome-wide approach for the identification of differentially methylated genes using MeDIP in combination with a promoter-tiling array. The advantage of the promoter tiling array used in this study is that it includes multiple probes per gene to improve the sensitivity of assay [43].

The 5'-methylcytosine antibody enriched sequences of CpG-rich promoter regions in HL-60 known to be highly methylated, which underlines that enrichment of methylated DNA is associated with the CpG density in that particular sequence. The in vitro methylated dilution series (0–100% methylation level) showed enrichment of two uniformly methylated genes with increasing methylation level, where sufficient enrichment was recognized only at methylation level above 20%. It is possible that the 5mC antibody precipitates methylated cytosines in CpG dinucleotides with at least 30% methylation grade. In the dilution series, all but one (at 20% methylation)



**Fig. 4. Histogram of distribution pattern of fold change and p-values of probes in according to CpG density.** (A–C) The fold-change is plotted on a log<sub>2</sub>-scale, i.e. a value of zero corresponds to a factor of one, and a value of one to a factor of two. For probes with low CpG density (A) the measurements are slightly different from the background obtained for 0% methylation. The fold-change distribution for the comparisons 50% vs. 0% and 100% vs. 0% shows a slight shift toward positive values. For the comparisons 30% vs. 0% and 20% vs. 0% methylation, there are almost no detectable effects. For medium CpG levels (B), methylation differences are detectable. The fold-changes calculated for the comparisons 50% vs. 0% and 100% vs. 0% are clearly shifted toward positive values. For high CpG levels (C), even the comparison of 30% vs. 0% (light blue) yields detectable positive fold-changes. (D–F) The p-values calculated by a two-sample t-test are significant only for the comparisons 50% vs. 0% and 100% vs. 0%. For the comparisons 30% vs. 0% to 100% vs. 0%, the p-value distribution is flat indicating very low detectable effects for probes with low CpG density (D). For medium CpG levels (E), the comparisons 50% vs. 0% and 100% vs. 0% yield significant results, i.e. the p-values are shifted toward zero while the p-values are almost uniformly distributed between zero and one for 20% and 30% methylated DNA. For high CpG levels (F), the comparisons 50% vs. 0% and 100% vs. 0% are highly significant. The p-values calculated for the comparison of 30% vs. 0% (light blue) methylation are slightly significant. The histograms based on the analysis of the dilution series reflect the data based on probes of chromosome 10. (For interpretation of the references to color in the artwork, the reader is referred to the web version of the article.)



**Fig. 5.** Histogram of distribution pattern of fold change and  $p$ -values in the microarray data of HL-60 and CD34<sup>+</sup> cells. The mean fold-changes (red columns) were calculated for each probe. The null-distribution (blue) has been calculated by randomization of the class labels. (A) Positive fold-changes are over-represented in HL-60 cells compared to CD34. To remove probes which are not differentially methylated, only fold-changes with significant  $p$ -value ( $p < 0.01$ ) are displayed. (B) The  $p$ -values obtained by a two-sample  $t$ -test (red) are shifted toward zero. (For interpretation of the references to color in the artwork, the reader is referred to the web version of the article.)

technical replicate of a sample group accumulate in the principal component analysis (PCA), an unsupervised method to discriminate biological difference, so that a separation of the different groups was determined (from 0% to 100%). The separation in different groups here is based on the difference in the signal intensities of the probes derived from the precipitated DNA.

Differences in the precipitated DNA between the healthy CD34<sup>+</sup> and the AML cell line HL-60 were shown in the PCA as well, so that the two cell populations were separated based on the differences in the signal intensities and differences in their DNA methylation profile. The two replicates of HL-60 spreading away from the others do not affect the outcome of the data, and were therefore not excluded from further analysis. The differences of these two microarray data sets of the same cell line may issue from sample preparation just before hybridization as the differences in the outcome are little. However, the separation between both cell populations appears to be clear cut. It is of note that although IP DNA was amplified before hybridization, the technical replicates showed comparable distribution in the PCA, indicating that MeDIP was reproducible and that

a potential bias through whole-genome amplification was only of minor concern.

In theory, the MeDIP assay can detect all methylated CpG dinucleotides in the genome. Here we show that MeDIP-Chip is clearly biased toward sequences with high CpG density as well as repeated sequences tend to be overrepresented due to the 5mC antibody. Therefore, the analysis of regions with low CpG density is expected to be less sensitive [22].

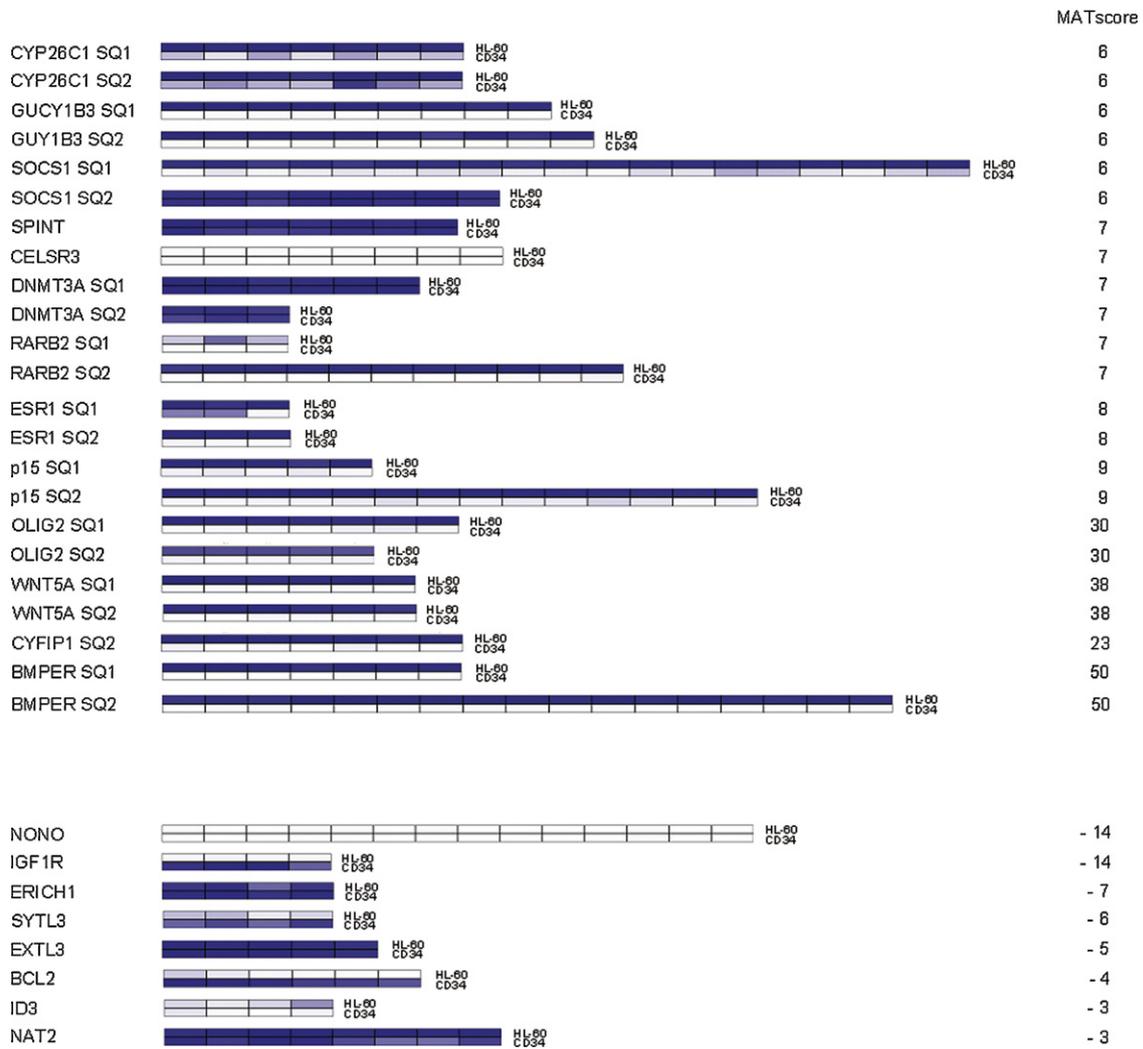
Since less strong but relevant effects of transcriptional gene regulation already occur at sequences with medium CpG density, these regions of methylated CpGs are possibly underrepresented when using MeDIP. Furthermore, MeDIP requires several methylated CpGs since the 5mC antibody only binds to >4 nearby methylated CpG sites [44]. Thus far, only little is known about the degree of methylation of the individual CpG within the precipitated fragment detected by antibodies [45].

In contrast to Weber et al., the group of Butcher et al. [27] showed that MeDIP not only enriches for CpG islands but also regions with very low CpG density and CpG island shores, as most tissue-specific differential DNA methylation is located at CpG island shores [46]. The results of Nair et al. obtained are similar – the group compared MeDIP and MBD, a technique to capture methylated DNA using a methyl-CpG binding domain-based protein, in combination with the Affymetrix promoter tiling array. They revealed that both enrichment techniques are sensitive for the detection of methylated CpGs, with MeDIP primarily enriching for DNA regions with low CpG density, while the MBD technique enriches regions of higher CpG density [47]. Our experiments with artificially methylated DNA dilution series showed that CpGs with methylation levels above 30% in the dilution series were captured by the 5mC antibody, which underlines the study of Butcher et al. [27]. The analysis of the in vitro methylated DNA revealed that highly methylated sequences (starting from 50% methylated dilution series) yield considerably higher signal intensities. Moreover, probes with high CpG density show higher fold changes than probes with medium or low CpGs. However, probes with low and medium CpG density were able to detect significant signal intensities, demonstrating the sensitivity of the 5′mC-antibody to precipitate DNA sequences with low CpG density as well.

Interestingly, the distribution pattern of probes with medium and high CpG density revealed a considerable shift of fold changes only for highly methylated dilution series (50% and 100%) to positive values ( $p$ -value < 0.05). Since one restraint of affinity based methods is that methylated CpG-rich regions are more strongly enriched than methylated CpG-poor sequences, we accurately assess the CpG density of the probes and classified them in low, medium and high CpG density.

In the dilution series experiment, fold changes larger than 2 are almost exclusively observed for methylation difference of 50% and 100%. If this observation for an in vitro setting can be generalized to an in vivo situation, fold-changes larger than 2 in an application would correspond to a strong increase of the methylation level. It is important to note, that the results of the in vitro methylated dilution series do not correlate with DNA samples of healthy donors or AML patients as the CpGs in the used samples are uniformly methylated and therefore do not represent the actual methylation level and distribution in humans. As the distribution of methylated cytosines in vivo is mosaic-like, it is difficult to estimate the methylation level of individual CpGs to be immunoprecipitated by the 5mC antibody.

The MAT algorithm used to identify differentially methylated regions in the tested cell lines assigned a MAT score for the detected regions. The analysis revealed 6832 regions with positive MATscore for HL-60, suggesting hypermethylation. To test this assumption we analyzed randomly selected promoter regions via pyrosequencing to quantify CpG methylation level. In fact, we determined



**Fig. 6. Pyrosequencing data of MeDIP-Chip results.** Pyrosequencing map of analyzed promoter regions derived from microarray data analysis. A subset of the identified differentially methylated regions was analyzed for the CpG methylation level by pyrosequencing. The methylation level of each CpG (presented by a box) in an amplicon is visualized by a color gradient (white boxes indicate no methylation and dark blue boxes high methylation). The corresponding MATscore of the analyzed region is shown on the right. On this map, 21 of 23 pyrosequenced genes with positive MATscore showed hypermethylation and 4 of 8 genes showed hypomethylation in HL-60 in contrast to CD34+ cells. (For interpretation of the references to color in the artwork, the reader is referred to the web version of the article.)

hypermethylation in 21 of 23 analyzed target sequences in HL-60 and hypomethylation in 4 of 8 genes, respectively. But because by pyrosequencing only sequences of about 150 bp can be analyzed, while data analysis done with PARTEK GS software considers promoter regions in a sliding window of 600 bp, there is a lack of comparability of the two methods. Moreover, due to the bisulfite treatment of DNA, unmethylated cytosines are converted to thymine, resulting in long poly-T stretches that make it difficult to sequence designated target regions.

The data analysis also revealed 8985 regions with negative MATscores indicating hypomethylation. As reported recently [48], MeDIP-seq of bone marrow samples of AML patients revealed not only that alterations in leukemia-associated differentially methylated regions include gene promoters, gene bodies and CpG island shores but also the non-promoter genomic features with significant hypomethylation pattern of certain interspersed repeats being associated with AML cytogenetic subtypes.

However, to investigate DNA methylation differences between a malignant cell population (here: HL-60 cells) and a healthy counterpart (here: CD34+ cells) the MeDIP-Chip is a robust platform. Fishers exact test confirmed that both methods yield the same tendency, i.e. the sign of the MAT score is statistically not

independent from the methylation level as measured by pyrosequencing ( $p=0.0009$ ).

In summary we found high coherence between MeDIP-Chip data and pyrosequencing data and were able to detect genes that are known to be methylated in AML, thus demonstrating that immunoprecipitation of methylated DNA coupled with a promoter tiling array is a timely method to generate comparative genome-wide profiles of DNA methylation. In ongoing studies we are applying this method to primary AML samples.

#### Conflicts of interest statement

The authors reported no potential conflicts of interest.

#### Acknowledgments

This work was supported by the Deutsche Krebshilfe [grant no. 108397, 110213]. ML is supported by the Deutsche Forschungsgemeinschaft (DFG, grant no. Lu 429/7-1 and CRC992/C04). We thank Prof. Dirk Schübeler (Friedrich Miescher Institute for Biomedical

Research, Basel, Switzerland) for the critical reading and comments on this manuscript.

**Contributors.** DP, JT, ML and BH designed the study and wrote the paper. AY performed experiments and wrote the paper. CK analyzed experiments and wrote the paper. MA and GK performed experiments.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.leukres.2012.09.014.

## References

- Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell* 2007;128:669–81.
- Vucic EA, Wilson IM, Campbell JM, Lam WL. Methylation analysis by DNA immunoprecipitation (MeDIP); microarray analysis of the physical genome. In: Walker JM, editor. *Methods in molecular biology*. New York, USA: Humana Press; 2009. p. 141–53.
- Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum Mol Genet* 2001;10:687–92.
- Fraga MF, Esteller M. Epigenetics and aging: the targets and the marks. *Trends Genet* 2007;23:413–8.
- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6–21.
- Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 2003;33:245–54.
- Gonzalzo ML, Liang G, Spruck CH, Zingg J-M, Rideout WM, Jones PA. Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res* 1997;57:594–9.
- Piotrowski A, Benetkiewicz M, Menzel U, de Ståhl TD, Mantripragada K, Grigoriadis G, et al. Microarray-based survey of CpG islands identifies concurrent hyper- and hypomethylation patterns in tissues derived from patients with breast cancer. *Genes Chromosomes Cancer* 2006;45:656–67.
- Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007;16:R50–9.
- Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983;301:89–92.
- Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 2002;3:415–28.
- Chekun VF, Lukyanova NY, Kovalchuk O, Tryndyk VP, Pogribny IP. Epigenetic profiling of multidrug-resistant human MCF-7 breast adenocarcinoma cells reveals novel hyper- and hypomethylated targets. *Mol Cancer Ther* 2007;6:1089–98.
- Glasspool RM, Teodoridis JM, Brown R. Epigenetics as a mechanism driving polygenic clinical drug resistance. *Br J Cancer* 2006;94:1087–92.
- Plimack ER, Kantarjian HM, Issa J-P. Decitabine and its role in the treatment of hematopoietic malignancies. *Leukemia Lymphoma* 2007;48:1472–81.
- Callinan PA, Feinberg AP. The emerging science of epigenomics. *Hum Mol Genet* 2006;15:R95–101.
- Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddloh JA, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 2008;18:780–90.
- Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, et al. Comparative isochizomer profiling of cytosine methylation: the HELP assay. *Genome Res* 2006;16:1046–55.
- Gebhard C, Schwarzfischer L, Pham T-H, Schilling E, Klug M, Andreessen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res* 2006;66:6118–28.
- Schilling E, Rehli M. Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* 2007;90:314–23.
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 2006;126:1189–201.
- Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 2007;39:457–66.
- Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005;37:853–62.
- Cross SH, Charlton JA, Nan X, Bird AP. Purification of CpG islands using a methylated DNA binding column. *Nat Genet* 1994;6:236–44.
- Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M, Segal E, et al. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat Genet* 2006;38:149–53.
- Koga Y, Pelizzola M, Cheng E, Krauthammer M, Sznol M, Ariyan S, et al. Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Res* 2009;19:1462–70.
- Jia J, Pekowska A, Jaeger S, Benoukraf T, Ferrier P, Spicuglia S. Assessing the efficiency and significance of methylated DNA immunoprecipitation (MeDIP) assays in using in vitro methylated genomic DNA. *BioMed Central* 2010;3:240.
- Butcher LM, Beck S. AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay. *Methods* 2010;52:223–31.
- Collins SJ, Gallo RC, Gallagher RE. Continuous growth and differentiation of human myeloid leukaemic cells in suspension culture. *Nature* 1977;270:347–9.
- Dalton Jr WT, Ahearn MJ, McCredie KB, Freireich EJ, Stass SA, Trujillo JM. HL-60 cell line was derived from a patient with FAB-M2 and not FAB-M3. *Blood* 1988;71:242–7.
- Fatemi M, Pao MM, Jeong S, Gal-Yam EN, Egger G, Weisenberger DJ, et al. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res* 2005;33:e176.
- Downey T. Analysis of a multifactor microarray study using partek genomics solution. *methods in enzymology*. San Diego, USA: Academic Press; 2006. pp. 256–270.
- Sadikovic B, Andrews J, Carter D, Robinson J, Rodenhiser DI. Genome-wide H3K9 histone acetylation profiles are altered in benzo(a)pyrene-treated MCF7 breast cancer cells. *J Biol Chem* 2008;283:4051–60.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, et al. Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci* 2006;103:12457–62.
- Droit A, Cheung C, Gottardo R. rMAT – an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics* 2010;26:678–9.
- Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protocols* 2007;2:2265–75.
- Fazi F, Zardo G, Gelmetti V, Travaglini L, Ciolfi A, Di Croce L, et al. Heterochromatic gene repression of the retinoic acid pathway in acute myeloid leukemia. *Blood* 2007;109:4432–40.
- Hackanson B, Bennett KL, Brena RM, Jiang J, Claus R, Chen S-S, et al. Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia. *Cancer Res* 2008;68:3142–51.
- Kroeger H, Jelinek J, Estéicio MRH, He R, Kondo K, Chung W, et al. Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse. *Blood* 2008;112:1366–73.
- Melki JR, Vincent PC, Clark SJ. Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. *Cancer Res* 1999;59:3730–40.
- Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* 2010;17:13–27.
- Fazzari MJ, Grealley JM. Epigenomics: beyond CpG islands. *Nat Rev Genet* 2004;5:446–55.
- Huang Y-W, Huang TH-M, Wang L-S. Profiling DNA methylomes from microarray to genome-scale sequencing. *Technol Cancer Res Treat* 2010;9:107–218.
- Weng Y-I, Huang THM, Yan PS. Methylated DNA immunoprecipitation and microarray-based analysis: detection of DNA methylation in breast cancer cell lines; molecular endocrinology. In: Walker JM, editor. *Methods in molecular biology*. New York, USA: Humana Press; 2009. p. 165–76.
- Sengenès J, Daunay A, Charles M-A, Tost J. Quality control and single nucleotide resolution analysis of methylated DNA immunoprecipitation products. *Anal Biochem* 2010;407:141–3.
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 2009;41:178–86.
- Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* 2011;6:34–44.
- Saied MH, Marzec J, Khalid S, Smith P, Down TA, Rakyan VK, et al. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. *PLoS ONE* 2012;7.