35 Speck NA, Gilliland DG. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer* 2002; **2**: 502–513.
36 Swerdlow S, Campo E, Harris N, Jaffe E, Pileri S, Stein H *et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 4th edn. WHO Press: Lyon, 2008.
37 Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009; **114**: 937–951.
38 Abdel-Wahab O, Levine RL. Mutations in epigenetic modifiers in the pathogenesis and therapy of acute myeloid leukemia. *Blood* 2013; **121**: 3563–3572.
39 Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F *et al.* A novel hierarchical prognostic model of AML solely based on molecular mutations. *Blood* 2012; **120**: 2963–2972.
40 Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.
41 Tang JL, Hou HA, Chen CY, Liu CY, Chou WC, Tseng MH *et al.* AML1/RUNX1 mutations in 470 adult patients with *de novo* acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood* 2009; **114**: 5352–5361.
42 Stone RM. Acute myeloid leukemia in first remission: to choose transplantation or not? *J Clin Oncol* 2013; **31**: 1262–1266.
43 Paschka P, Schlenk RF, Gaidzik VI, Habdank M, Kronke J, Bullinger L *et al.* IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J Clin Oncol* 2010; **28**: 3636–3643.
44 Dufour A, Schneider F, Metzeler KH, Hoster E, Schneider S, Zellmeier E *et al.* Acute myeloid leukemia with biallelic CEBPA gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. *J Clin Oncol* 2010; **28**: 570–577.
45 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; **19**: 1639–1645.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)

Longitudinal Analysis of DNA Methylation in CD34+ Hematopoietic Progenitors in Myelodysplastic Syndrome

Yan-Fung Wong, Chris N. Micklem, Masataka Taguchi, Hidehiro Itonaga, Yasushi Sawayama, Daisuke Imanishi, Shinichi Nishikawa, Yasushi Miyazaki and Lars Martin Jakt

Stem Cells Trans Med published online August 13, 2014

STEM CELLS TRANSLATIONAL MEDICINE®

CANCER STEM CELLS

Longitudinal Analysis of DNA Methylation in CD34+ Hematopoietic Progenitors in Myelodysplastic Syndrome

YAN-FUNG WONG,[a,b,c] CHRIS N. MICKLEM,[a,d] MASATAKA TAGUCHI,[b] HIDEHIRO ITONAGA,[b] YASUSHI SAWAYAMA,[b] DAISUKE IMANISHI,[b] SHINICHI NISHIKAWA,[a,e] YASUSHI MIYAZAKI,[b] LARS MARTIN JAKT[a,f]

[a]Laboratory for Stem Cell Biology, RIKEN Center for Developmental Biology, Kobe, Japan; [b]Department of Hematology, Atomic Bomb Disease and Hibakusya Medicine Unit, Atomic Bomb Disease Institute, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; [c]The Danish Stem Cell Centre (DanStem), University of Copenhagen, Copenhagen, Denmark; [d]Imperial College London, London, United Kingdom; [e]All About Science Japan, Kobe, Japan; [f]Department of Systems Medicine, Mitsunada Sakaguchi Laboratory, Keio University School of Medicine, Institute of Integrated Medical Research, Tokyo, Japan

Correspondence: Yan-Fung Wong, Ph.D., University of Copenhagen, DanStem, Faculty Of Health Sciences, Blegdamsvej 3B, Copenhagen 2200, Denmark. Telephone: 45-2963-2154; E-Mail: yan.fung.wong@sund.ku.dk; or Lars Martin Jakt, Ph.D., Department of Systems Medicine, Mitsunada Sakaguchi Laboratory, Keio University School of Medicine, Room 2N5, Institute of Integrated Medical Research, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan. Telephone: 81-90-97022835; E-Mail: mjakt@z2.keio.jp

Received February 26, 2014; accepted for publication June 26, 2014.

1066-5099/2014/$20.00/0

http://dx.doi.org/10.5966/sctm.2014-0035

Key Words. Azacytidine • DNA methylation • MDS • Myelodysplastic syndromes

ABSTRACT

Myelodysplastic syndrome (MDS) is a disorder of hematopoietic stem cells (HSCs) that is often treated with DNA methyltransferase 1 (*DNMT1*) inhibitors (5-azacytidine [AZA], 5-aza-2′-deoxycytidine), suggesting a role for DNA methylation in disease progression. How DNMT inhibition retards disease progression and how DNA methylation contributes to MDS remain unclear. We analyzed global DNA methylation in purified CD34+ hematopoietic progenitors from MDS patients undergoing multiple rounds of AZA treatment. Differential methylation between MDS phenotypes was observed primarily at developmental regulators not expressed within the hematopoietic compartment and was distinct from that observed between healthy hematopoietic cell types. After AZA treatment, we observed only limited DNA demethylation at sites that varied between patients. This suggests that a subset of the stem cell population is resistant to AZA and provides a basis for disease relapse. Using gene expression data from patient samples and an in vitro AZA treatment study, we identified differentially methylated genes that can be activated following treatment and that remain silent in the CD34+ stem cell compartment of high-risk MDS patients. Haploinsufficiency in mice of one of these genes (*NR4A2*) has been shown to lead to excessive HSC proliferation, and our data suggest that suppression of *NR4A2* by DNA methylation may be involved in MDS progression. STEM CELLS TRANSLATIONAL MEDICINE *2014;3:1–11*

INTRODUCTION

Hematopoietic stem cells (HSCs) are multipotent progenitor cells from which all differentiated blood cell types arise during the process of hematopoiesis [1]. Myelodysplastic syndrome (MDS) is a heterogeneous spectrum disorder in HSCs characterized by ineffective and dysplastic hematopoietic differentiation [2]. It is mainly a disease of the older population, with ~80% of cases in those older than 60 years of age, with the natural course ranging from chronic cytopenia to slow or rapid progression into acute myeloid leukemia (AML) [3]. Allogenic hematopoietic stem cell transplantation is the most effective therapy, particularly for AML; however, for elderly patients, the risks of transplantation outweigh the benefits [4]. Instead the hypomethylating drugs, 5-azacytidine (AZA) and 5-aza-2′-deoxycytidine (DAC) have provable therapeutic effects and form the basis of currently recommended treatment [5, 6]. Although the exact mechanisms of action are not well understood, the efficacy of AZA and DAC implicate the role of DNA methylation in the progression of MDS.

MDS is characterized by the expansion of the CD34+ bone marrow (BM) population and is similar to AML, with the two disorders primarily distinguished by the severity of the symptoms (particularly bone marrow blast counts). However, the clinical course of AML and MDS differ, and global analysis of DNA methylation has indicated that the MDS genome is more extensively hypermethylated than in de novo AML [7]. Consequently, AML and MDS are likely to arise from different molecular causes and to require different treatment. Genes implicated in epigenetic control, including *ASXL1* [8], *TET2* [9], *EZH2* [10], and *DNMT3A* [11], are frequently mutated in MDS patients, and DNA methylation at single genes [11, 12] or combinations of multiple genes [13] has been correlated with outcome. Gene silencing by hypermethylation at promoters of tumor-associated genes or molecules involved in signaling may be important in the pathogenesis of MDS [14], and whole-genome studies of DNA methylation are beginning to reveal the complexity of the disease at the epigenetic level [7, 15, 16]. We performed a genomewide methylation analysis of purified bone marrow CD34+ blast cells

taken from MDS patients undergoing AZA treatment. Our results demonstrate that both disease-associated hypermethylation and AZA-induced demethylation occur at restricted loci that are enriched for regulatory regions.

It has been suggested recently that mammalian promoters can be divided into three types [17], and our data support the argument that disease- and differentiation-associated DNA methylation occur at type III and type I promoters, respectively. Our results also indicate that a portion of the CD34+ compartment is resistant to AZA treatment, and this population may serve as a source of tumorigenic cells that facilitate tumor reversion. Making use of previous data sets, we were also able to identify a small set of genes that may be involved in both the progression of MDS and the response to AZA.

Materials and Methods

Patients

All experimental protocols were reviewed and approved by the ethics committees of Nagasaki University Hospital and the RIKEN Center for Developmental Biology. Fresh bone marrow samples were collected from eight patients who provided informed consent. All patients had diagnosis of MDS according to the World Health Organization classification system [18] and were assessed by the original [19] and revised [20] International Prognostic Scoring System (supplemental online Table 1). Four patients received AZA treatment at the schedule approved by the U.S. Food and Drug Administration (75mg/m^2 per day for 7 days every 4 weeks). Responses were scored according to International Working Group 2006 criteria for MDS (supplemental online Figs. 1–4; supplemental online Tables 1, 2). Two patients (identifiers MDS2 and MDS6) were classified as having the mild refractory cytopenia with multilineage dysplasia (RCMD) form of MDS at the start of treatment. Both patients responded to AZA treatment and achieved complete remission (CR) [21] after 4 cycles (MDS6) or 5 cycles (MDS2) of AZA, although neither showed a decrease in BM blast counts (supplemental online Fig. 5). MDS2 blast counts increased gradually during the course of treatment, and patient MDS2 progressed to AML 21 months after beginning treatment. In contrast, the two patients with the severe refractory anemia with excess blasts type 2 (RAEB-2) form (identifiers MDS3 and MDS4) both showed an initial and rapid decrease in blast counts and achieved BM CR and partial hematologic improvement. However, after the initial improvement, patient MDS4 rapidly progressed to full-blown AML and was taken off AZA treatment. Blast counts for patient MDS3 reached a minimum after seven rounds of treatment, after which counts gradually increased; however, this patient had not progressed to AML and remained under AZA therapy at the time this report was written. Similarly, patient MDS6 continued to respond to treatment.

Cell Cultures

Mononuclear cells (MNCs) were isolated from bone marrow samples using Ficoll-Hypaque (GE Healthcare, Stockholm, Sweden, http://www.gehealthcare.com). The frozen MNCs were thawed and cultured for 16 hours in hematopoietic culture medium (serum-free X-Vivo 10 containing 50 ng/ml stem cell factor; Peprotech, Rocky Hill, NJ, http://www.peprotech.com), 10 ng/ml thrombopoietin (Peprotech), and 20 ng/ml Flt3/4 ligand (R&D Systems Inc., Minneapolis, MN, http://www.rndsystems.com) in 5% CO_2 at 37°C at saturated humidity. CD34+ cells were purified by magnetic bead separation using the human CD34 MicroBead Kit (Miltenyi Biotec, Bergisch Gladbach, Germany, http://www.miltenyibiotec.com) and the AutoMACS Pro separator (Miltenyi Biotec) per the manufacturer's instructions. Post-sorting purity was higher than 95%. Purified CD34+ 2M-101C cells (lots 090125A [C1] and 0000312727 [C2]), CD14+ 2W-400C cells (lot 2F3400B), and CD56+ cells (lot 061223A) were purchased from Lonza (Walkersville, MD, http://www.lonza.com) and used as nonleukemic cells for methylation analyses.

Methylcytosine Fractionation and Promoter-Array Experiments

Genomic DNA was extracted from purified CD34+ cells (DNeasy Blood and Tissue Kit; Qiagen, Hilden, Germany, http://www.qiagen.com). McrBC-based methylcytosine fragmentation was performed, as described previously [16, 22], with 100 ng purified genomic DNA. CHROMA SPIN 1000 (Clontech, Mountain View, CA, http://www.clontech.com) columns were used to remove DNA molecules with fewer than 300 base pairs. Whole-genome amplification polymerase chain reaction (PCR; GenomePlex Complete Whole Genome Amplification Kit; Sigma-Aldrich, St. Louis, MO, http://www.sigmaaldrich.com) was performed for 14 cycles with 10 ng of eluted DNA samples.

DNA labeling and hybridization was performed according to the supplied protocols (Agilent Microarray Analysis of Methylated DNA Immunoprecipitation Version 1.0; Agilent Technologies, Palo Alto, CA, http://www.agilent.com). For each patient sample, 2.5 μg of McrBC- and mock-digested DNA were labeled with Cy5 and Cy3, respectively. Equal amounts of labeled samples were mixed and applied to SurePrint G3 Human Promoter Microarrays (G4873A; Agilent). Methylation levels were estimated from the log (base 2) of the ratio of the intensity of signal from the undigested to digested DNA. Data were analyzed with Agilent Genomic Workbench 6.0, and statistical analyses (supplemental online data) were performed using Bioconductor [23] and custom R code.

Bisulfite Sequencing Analysis

DNA (1 μg) from each sample was modified using sodium bisulfite (EpiTect Bisulfite kits; Qiagen), as suggested by the manufacturers. Gene primers for amplifying *OTP*, *MAB21L1*, *C17orf53*, *EN1*, *WT1*, and *NR4A2* were developed using MethPrimer [24]. PCR was performed using EpiTaq HS enzyme (Takara, Otsu, Japan, http://www.takara.co.jp), according to the manufacturer's instructions. The PCR products were gel purified (QIAquick Gel Extraction Kit; Qiagen) and cloned into the TA vector (TOPO-TA cloning; Invitrogen, Carlsbad, CA, http://www.invitrogen.com). Ten clones from each sample were selected, and the sequences were determined using an ABI PRISM 3700 Genetic Analyzer (Applied Biosystems, Foster City, CA, http://www.appliedbiosystems.com). Data were summarized with a Web-based software, QUMA [25] (http://quma.cdb.riken.jp/top/quma_main_j.html).

Primers for pyrosequencing *MPO*, *NR4A2*, and *OTP* were designed using PyroMark Assay Design 2.0 software (Qiagen). Pyrosequencing was conducted using a PyroMark Q24 instrument (Qiagen), and methylation levels were quantitated with

PyroMark Q24 1.010 software. Relative peak height differences were used to calculate the percentage of 5-methylcytosines at each site. All primer sequences for bisulfite sequencing are shown in supplemental online Table 3.

Gene Expression Analysis

Total RNA was purified from CD34+ cells using the ZR RNA MicroPrep Kit (Zymo Research, Orange, CA, http://www.zymoresearch.com). The amount and length distribution of the total RNA were measured using an RNA 6000 Pico Kit (Agilent Biotechnology). Sequencing libraries were prepared using the SMARTer Ultra Low RNA Kit for Illumina sequencing (Clontech) with 10-ng purified RNA samples and 12 cycles of PCR enrichment. The amplified cDNA was fragmented using the Covaris shearing system (Covaris Inc., Woburn, MA, http://covarisinc.com), and the DNA sequencing library was analyzed using a massively parallel sequencer (HiSEquation 1000; Illumina Inc., San Diego, CA, http://www.illumina.com). Raw sequence reads were trimmed using Trimmomatic (to remove sequencing primers) [26] and mapped to the human reference genome (hg19) using TopHat (version 2.0.3) [27] with default parameters. Fragments per kilobase of exon per million fragments mapped were calculated using Cufflinks (version 2.0.1) [28].

RESULTS

DNA Methylation Marks MDS Progression in the CD34+ Stem Cell Compartment

We assessed methylation levels in CD34+ cells isolated from 13 MDS bone marrow samples using McrBC methyl-cytosine fragmentation [16] in combination with whole-genome promoter-tiling microarrays. We also included $CD14^+$ monocytes, $CD56^+$ natural killer (NK) cells, and CD34+ HSCs collected from healthy adult blood donors (nontumor [NT] samples) to assess the normal range of DNA methylation in blood. To obtain an overview, we subjected the data to principal component analysis (PCA) (Fig. 1A) trained on all technical replicates. NT and MDS samples were separated in the first dimension, and high-risk RAEB-2 and low-risk RCMD samples were segregated in the second dimension. In order to summarize the direction of change at differentially methylated regions (DMRs), we calculated a relative methylation score for each region (supplemental online data; File 1 in supplemental online Appendix 1). Among the NT samples, the methylation levels were highest in NK cells, but in general, normal cell types had lower methylation than MDS samples (although many individual DMRs were more heavily methylated in NK cells or MNCs). DMRs from RAEB-2 samples were hypermethylated compared with those from RCMD samples. These observations indicate that the distance between RAEB-2 and RCMD in the PCA is due to hypermethylation at specific DMRs in RAEB-2 and that, in general, MDS cells are hypermethylated compared with normal cells.

Differential Methylation at Distinct Locations and Genes

To investigate the biological roles of differential methylation, we identified genes containing DMRs between NT samples (NT-DMRs) (supplemental online Data Set 1; File 2 in supplemental online Appendix 1) and between the two tumor types (TT-DMRs) (supplemental online Data Set 2; File 3 in supplemental online Appendix 1). Genes containing NT-DMRs were highly enriched ($p = 1.4 \times 10^{-20}$) in Gene Ontology (GO) categories related to immune response and hematopoietic function (Table 1), which is consistent with a role for demethylation in the activation of hematopoietic genes during hematopoiesis [29]. In contrast, TT-DMRs were found primarily within genes associated with regulation of developmental processes ($p = 8.9 \times 10^{-53}$) (Table 1). Of the top 100 scoring genes (supplemental online Data Set 2), more than half were transcription factors and virtually all (~90%) have known developmental roles. Differential methylation was validated at the *OTP*, *MAB21L1*, *C17orf53*, *EN1*, *WT1*, and *NR4A2* loci by bisulfite sequencing (Fig. 2).

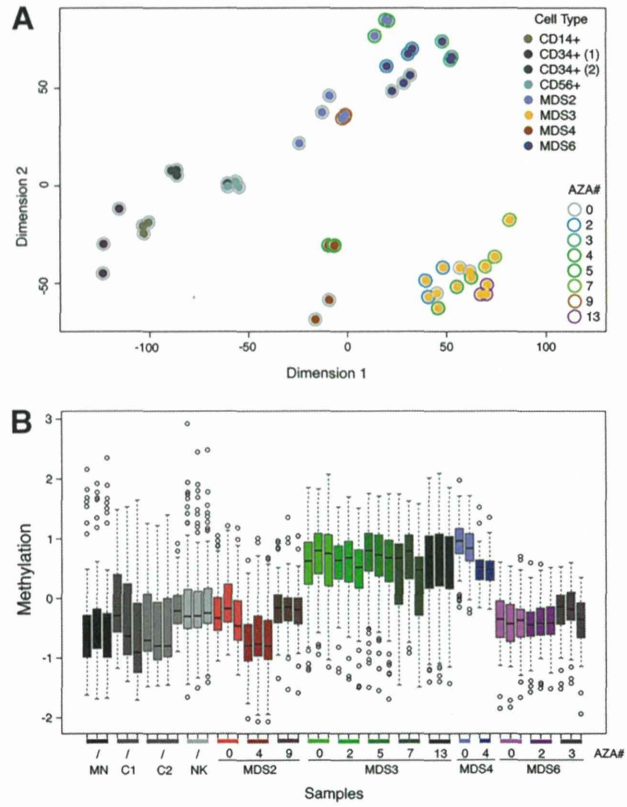

Figure 1. DNA methylation profiles of CD34+ hematopoietic progenitors from healthy donors and MDS patients. **(A):** Dimensions 1 and 2 of a principal component analysis of DNA methylation data from $CD14^+$ MN, $CD56^+$ NK, and CD34+ cells (C1 and C2) from healthy donors and MDS patients at different cycles of AZA treatment (AZA#). MDS samples are separated along dimension 1, whereas dimension 2 clearly segregates refractory cytopenia with multilineage dysplasia (patients MDS2 and MDS6) from refractory anemia with excess blasts type 2 (RAEB-2; patients MDS3 and MDS4). **(B):** Box plot of relative mean methylation scores at the top 100 scoring regions identified as differentially methylated between replicate groupings. Samples are along the *x*-axis, with differential methylation (methods) indicated on the y-axis. Colors indicate technical replicate groupings, with nonleukemic samples in shades of gray. AZA treatment is indicated by shading of the color, with darker shading indicating longer treatment. Each plot indicates values in a single technical replicate. All experiments were performed in triplicate except for C2 (with four sets) and MDS4 (with two sets). Methylation in RAEB-2 samples is clearly higher than in other samples, but a clear AZA response at these DMRs can be observed only for MDS2 and MDS4. Abbreviations: AZA, 5-azacytidine; MDS, myelodysplastic syndrome; MN, monocytes; NK, natural killer cells.

Table 1. GO analysis for differential methylation regions

GOBPID	Term	*p* value	Count	Size
Nontumor				
GO:0006955	Immune response	1.43×10^{-20}	128	993
GO:0002376	Immune system process	1.04×10^{-18}	180	1,725
GO:0001775	Cell activation	1.37×10^{-17}	100	738
GO:0002682	Regulation of immune system process	1.11×10^{-16}	106	832
GO:0050776	Regulation of immune response	5.21×10^{-16}	77	516
GO:0045321	Leukocyte activation	4.86×10^{-15}	77	537
GO:0048583	Regulation of response to stimulus	1.88×10^{-14}	203	2,238
GO:0002252	Immune effector process	2.28×10^{-13}	64	431
GO:0002684	Positive regulation of immune system process	3.56×10^{-11}	67	517
GO:0046649	Lymphocyte activation	7.01×10^{-11}	61	455
Tumor type				
GO:0044707	Single-multicellular organism process	8.91×10^{-53}	489	5,275
GO:0032501	Multicellular organismal process	4.47×10^{-52}	498	5,460
GO:0048731	System development	8.77×10^{-51}	365	3,327
GO:0007275	Multicellular organismal development	1.10×10^{-49}	398	3,864
GO:0048856	Anatomical structure development	9.92×10^{-48}	391	3,817
GO:0007399	Nervous system development	2.44×10^{-47}	237	1,687
GO:0032502	Developmental process	2.32×10^{-46}	420	4,326
GO:0007389	Pattern specification process	3.78×10^{-43}	107	424
GO:0048513	Organ development	4.12×10^{-43}	282	2,366
GO:0009887	Organ morphogenesis	2.12×10^{-42}	148	797

Abbreviations: GO, Gene Ontology; GOBPID, Gene Ontology Biological Process ID.

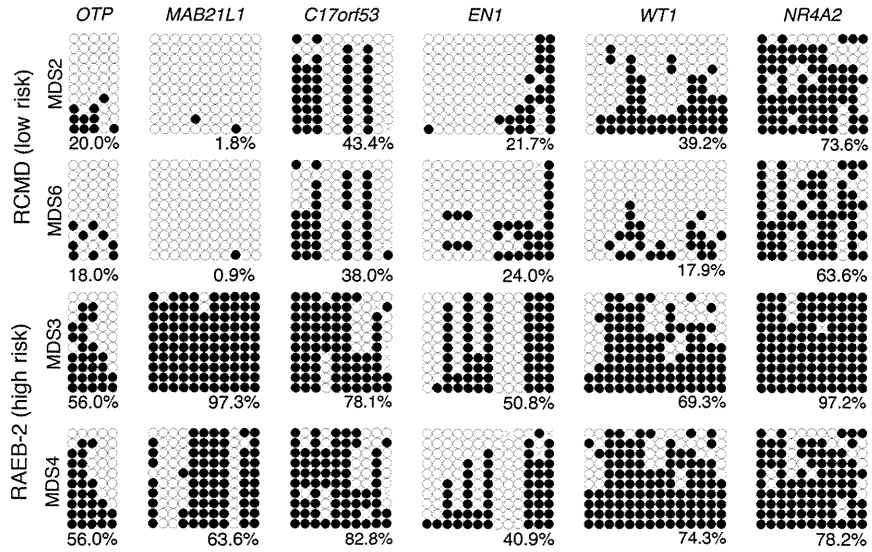

Figure 2. Bisulfite sequencing within tumor type differentially methylated regions. Bisulfite sequencing within the *OTP, MAB21L1, C17orf53, EN1, WT1*, and *NR4A2* genomic regions identified as differentially methylated between RCMD and RAEB-2 by microarray analysis. Abbreviations: MDS, myelodysplastic syndrome; RAEB-2, refractory anemia with excess blasts type 2; RCMD, refractory cytopenia with multilineage dysplasia.

To determine the expression of differentially methylated genes, we analyzed RNA-sequencing data obtained from four of the patient samples (MDS2-AZA0, MDS3-AZA0, MDS3-AZA4, and MDS4-AZA4). We selected the top 1,000 genes on the basis of differential methylation among all technical replicates (ALL), RAEB2 and RCMD (TT), NT samples, and treatment stages for individual MDS patients and compared the distributions of expression levels (Fig. 3). Distributions for TT-DMR genes were

massively shifted to the left, with the distribution completely losing its high-level peak. Effects of gene selection for other types of DMRs were much less severe, but treatment DMRs for MDS2 also selected genes with lower levels of expression. In contrast, selection of genes by DMRs for MDS3 and NT samples shifted the expression in the opposite direction, with more genes expressed at high levels than in the full distribution (dashed line). To address the expression of differentially methylated (DM) genes in the extended hematopoietic system, we used published gene expression data (GSE24759) [30] covering 38 hematopoietic cell types. Again, TT-DMR genes were expressed at low levels in all 38 cell types (supplemental online Fig. 6). These results demonstrate that genes that gain methylation during MDS progression tend not to be expressed or are expressed at low levels in the hematopoietic system.

We next considered the location of DMRs with respect to transcriptional start sites (TSSs) and CpG islands. All classes of DMRs were predominantly centered upstream of their TSSs, with the frequency decreasing with distance from the TSS, and we did not observe any obvious differences between DMR classes (supplemental online Fig. 7). In contrast, probes lying within DMRs were either depleted (NT, $p \approx 1 \times 10^{-40}$) or enriched (TT, $p \approx 1 \times 10^{-110}$) for CpG islands (supplemental online Fig. 8). However, only ~35% of TT-DMRs overlapped with CpG islands (supplemental online Fig. 8C, 8D), and the observed enrichment appears to be due to the presence of multiple CpG islands within TT-DMR gene regions (Files 4 and 5 in supplemental online Appendix 1) rather than locus-specific overlaps. Treatment-stage DMRs (for patients MDS2, MDS3, and MDS6) had intermediate properties, with patient MDS3 depleted for CpG islands, whereas the others were enriched.

Cancer-induced DNA hypermethylation has been associated with promoters carrying both active (H3K4me3) and repressive (H3K27me3) histone modifications (bivalent genes) in human embryonic stem cells (hESCs) [31]. To determine whether this is also true for the MDS DMRs, we analyzed histone modification data for eight different cell types and nine different histone modifications (Files 4–8 in supplemental online Appendix 1) obtained from the ENCODE Project Consortium [32]. In hESCs TT-DMRs were strongly enriched for both H3K27me3 and H3K4me3 as well as for H3K4me1 and H3K4me2. Interestingly, TT-DMRs were also enriched for H3K27me3 in most of the cell types analyzed. In contrast, NT-DMRs showed obvious selection effects only in the hematopoietic GM12878 cell line, in which strong enrichment for H3K27ac, H3K4me1, H3K4me2, and H3K4me3 was observed. Again, treatment-stage DMRs showed intermediate patterns, with MDS2 and MDS6 DMRs being similar to TT-DMRs, and MDS3 DMRs being similar to NT-DMRs.

The H3K4me1 mark has been associated with enhancer sequences, with active enhancers marked by H3K27ac and nonactive enhancers marked by H3K27me3 [33]. These observations suggest that TT-DMRs are associated with latent hESC enhancers that are more completely silenced in differentiated cell types. In contrast, the NT-DMRs represent enhancers that are more likely to be active in hematopoietic cell types.

To characterize the activity of these enhancer elements across a wider variety of tissues, we used a published analysis [33] of the Fantom5 data set [34] that identified enhancers by bidirectional expression in 808 different samples, representing a wide variety of cell types. Although only a small proportion of Fantom5 enhancers overlapped with DMRs, we found that differential methylation in enhancers was enriched for all replicate groupings except for TT-DMRs ($p = 4 \times 10^{-9}$ to 1.3×10^{-84}) (supplemental online Table 4; supplemental online Fig. 9). We then calculated the statistical enrichment specifically for active enhancers across the 808 Fantom5 samples. All replicate groupings except TT were strongly associated with enhancer activity in hematopoietic cell types. We found marginal enrichment for TT differential methylation in a single sample (adult cerebellum) and apparent depletion of statistical enrichment for hematopoietic cells (supplemental online Fig. 10). Again, these observations indicate that TT- and NT-DMRs are distinct and are enriched for regulatory elements that are either active (NT) or inactive (TT) within hematopoietic cells.

Prolonged AZA Treatment Results in Patient-Specific Demethylation

To characterize changes in DNA methylation in MDS patients following AZA administration, we quantified methylation at regions with methylation that changed during the course of AZA treatment. For both patient MDS2 (RCMD) and patient MDS3 (RAEB-2), we were able to identify clear effects, and we focused on these patients who had received treatment for more than 18 months. In neither case was demethylation correlated with a reduction in blast counts (Fig. 4; supplemental online Fig. 5). MDS2 demethylation was observed after four rounds of AZA treatment (Fig. 4A), but methylation reverted to pretreatment levels after nine rounds. Demethylation could also be seen in patient MDS3 (Fig. 4B); however, this occurred at a later stage and at different DMRs. These observations were confirmed by gene-specific methylation analysis of the *WT1* and *MPO* loci (Fig. 5).

In order to characterize the biological function of genes containing treatment DMRs, we identified GO categories of the top 1,000 genes for MDS2 and MDS3 DMRs (supplemental online Tables 5, 6). MDS3 regions were associated with genes with immune functions ($p = 1 \times 10^{-11}$), whereas MDS2 regions were enriched for developmental function ($p = 1 \times 10^{-22}$) in a similar manner to the TT-DMRs (Table 1). This is consistent with the expression levels of associated genes (Fig. 3) and with histone modification and enhancer expression at the DMRs.

NR4A2 Was Hypermethylated in High-Risk MDS Patients and Can Be Reactivated by AZA Treatment

To identify genes containing hypermethylated DMRs that have the potential to be activated as a result of DNA demethylation, we compared AZA- and DAC-induced perturbation observed in an in vitro model system [16] with the TT-DMR scores (Fig. 6A). This identified 16 candidates showing differential methylation in MDS patients as well as gene activation after in vitro treatment. These genes are hypermethylated in RAEB-2 patients, and five of these genes (*TRIB2*, *NR4A2*, *POU4F1*, *HDC*, and *OTP*) had lower expression in CD34+ cells from MDS3 and MDS4 samples than from MDS2 samples (supplemental online Fig. 11). Of these genes, we selected *NR4A2* and *OTP* (which have large numbers of probes in their DMRs) to further investigate methylation using pyrosequencing (Fig. 6B, 6C). In addition to the previously described MDS samples, we included samples from another four MDS patients, three of whom were classified as RAEB-2 and one with the mild refractory cytopenia with unilineage dysplasia (RCUD) form.

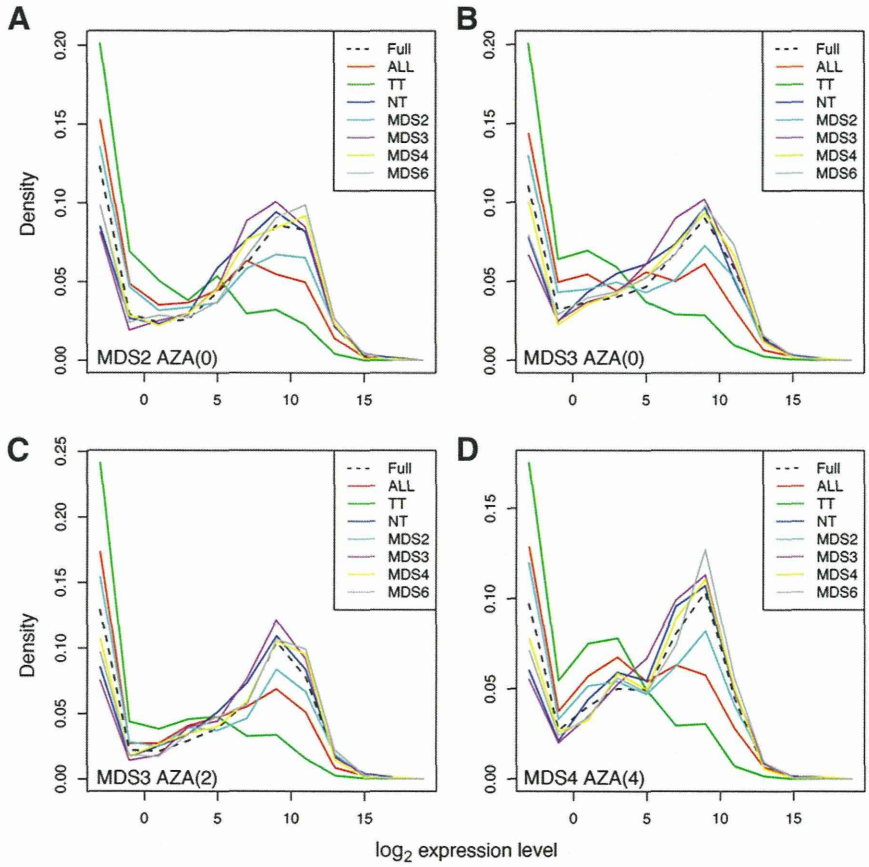

Figure 3. Expression in MDS samples of genes containing differentially methylated regions. Expression estimates in four patient samples (MDS2 and MDS3 control groups and MDS3 and MDS4 samples after two or four cycles of AZA treatment, respectively) were obtained by RNA sequencing. The top 1,000 genes containing regions differentially methylated between all replicate groups, RCMD and RAEB2 (tumor type), nontumor technical replicates, and between-technical replicates groups within patient groupings (indicating treatment or disease-progression differentially methylated regions [DMRs]) were identified, and the distributions of expression levels in these samples were determined. The dashed black line indicates the full distribution of expression levels for all genes. The *x*-axis shows $\log_2$ expression levels (with 0 counts set to 0.1), and the *y*-axis shows density. TT-DMRs select genes with low expression levels (green line), whereas NT (blue) and MDS3 (purple) DMRs select genes with slightly elevated expression levels. Abbreviations: AZA, 5-azacytidine; MDS, myelodysplastic syndrome; NT, nontumor; TT, tumor type.

In general, we found higher methylation levels in the RAEB-2 samples, but such differences were specific to individual CpG positions. For *NR4A2*, a clear separation of methylation levels was observed only at one of the five CpG positions; at other sites, RAEB-2 methylation was generally higher than that for RCMD, but the difference was not as pronounced. High methylation levels were also observed at most *NR4A2* sites in the NK cells, and that finding correlates well with our array-based data (File 3 [page 45] and File 4 [page 37] in supplemental online Appendix 1). Again, similar tendencies were observed at the *OTP* locus, with clear separation found at two CpG loci neighboring a site that showed no difference between samples. At the *OTP* locus, one of the RAEB-2 samples (patient MDS7) did not show hypermethylation at any of the sites, indicating that hypermethylation in these sites is not a requirement of RAEB-2. Somewhat surprisingly, we found an intermediate level of methylation at these sites in the RCUD sample.

Discussion

We performed global DNA methylation measurements in the CD34+ fraction of MDS patients undergoing treatment with AZA. This allowed us to observe differences between individual patients, MDS subclasses, and treatment effects. We believe that this is the first global study of DNA methylation in the CD34+ compartment of MDS patients.

We used promoter rather than CpG-island arrays to interrogate DNA methylation because previous studies indicate that differential methylation occurs primarily outside of CpG islands [35]. In fact, these arrays not only cover promoters but also often include complete genes, and DMRs were found throughout gene bodies. A PCA was sufficient to segregate leukemic from nonleukemic samples and RAEB-2 from RCMD samples in the two primary components, with samples from individual patients clearly segregated from each other. This emphasizes the large variability among individual tumors and suggests that DNA methylation can be used to classify tumor severity. Although our global analysis was limited to samples derived from 4 patients with 2 different subtypes ($n = 2$ for each subtype) of MDS, the extreme statistical associations with external data and ontologies argue that these observations are unlikely to be limited to the patients studied. It is possible that the differences we observed between the

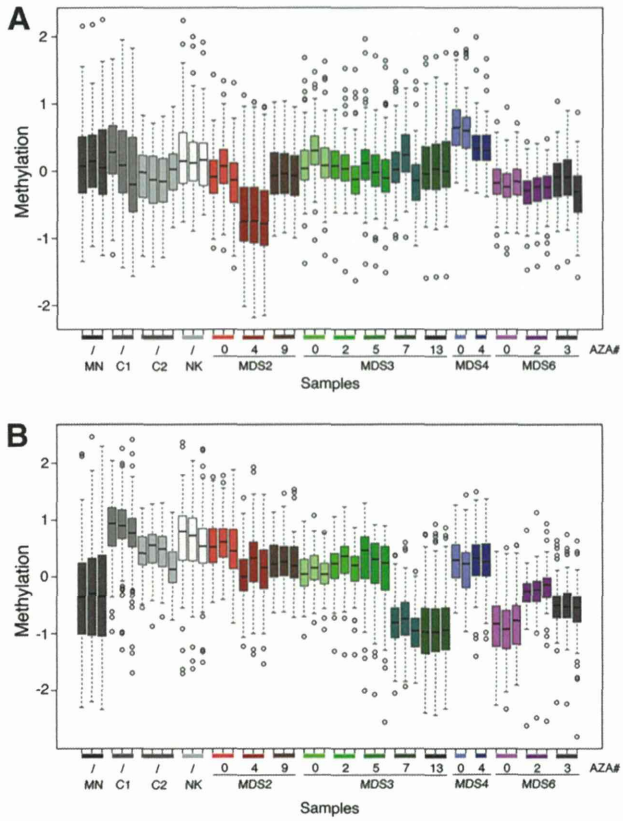

Figure 4. Demethylation after AZA. Methylation levels at the top 100 regions identified as differentially methylated within the MDS2 **(A)** and MDS3 **(B)** samples. The *y*-axis indicates mean differential methylation observed across the identified regions. The *x*-axis indicates samples grouped by the technical replicate groupings. AZA# indicates the number of AZA treatment cycles. Nonleukemic samples (MN, CD34+ cells at C1 and C2, NK) are shown in shades of gray. Patient samples are shown in individual colors, with the shading level indicating the number of AZA treatments, as indicated on the *x*-axis. Boxes indicate the interquartile range, with the median shown by the central bar. Whiskers extend to 1.53 the interquartile range. Demethylation of MDS2 samples can be observed after four rounds of treatment, whereas demethylation in MDS3 samples was observed only after the seventh round of AZA treatment. No appreciable demethylation of MDS3 differentially methylated regions (DMRs) was observed outside of MDS3, and these DMRs were variably methylated within the nonleukemic samples but were not generally hypermethylated within the MDS samples. Abbreviations: AZA, 5-azacytidine; C1 and C2, CD34+ cells from healthy donors; MDS, myelodysplastic syndrome; MN, monocytes; NK, natural killer cells.

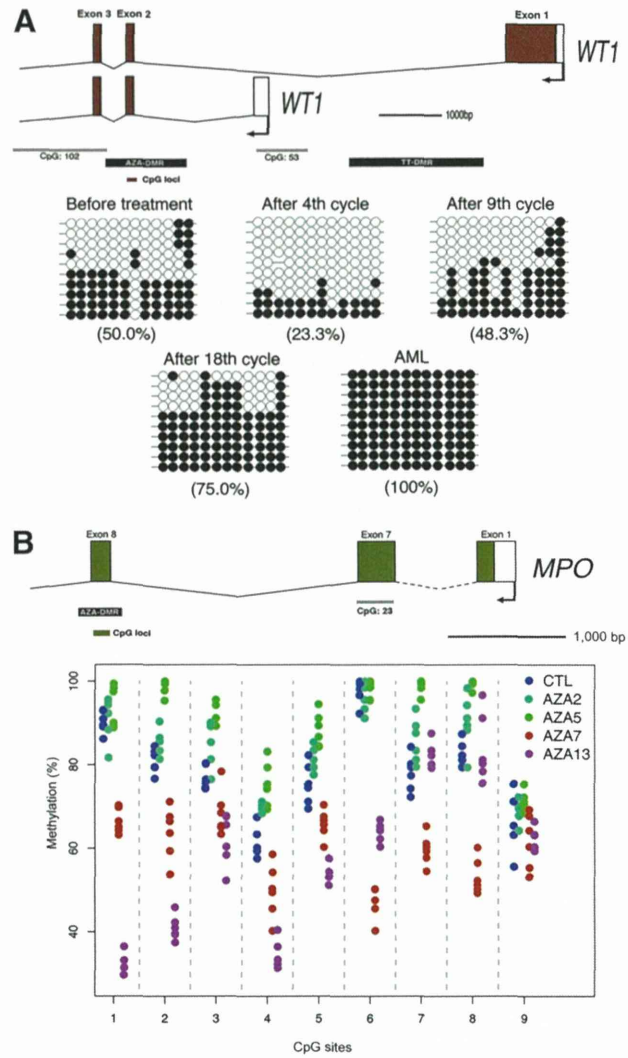

Figure 5. Longitudinal methylation profile at the *WT1* and *MPO* genes. **(A):** Bisulfite sequencing of 12 CpG sites (chr11: 32,450,101–32,450,179) in MDS2 CD34+ samples collected before and during AZA treatment (after 4th, 9th, and 18th cycles) and after progression to AML are shown. Although initial demethylation was observed after four rounds of treatment, methylation increased gradually after this point despite repeated AZA treatment. **(B):** Pyrosequencing of 9 CpG sites (chr17:56,352,875–56,353,055) in MDS3 CD34+ samples collected before and during AZA treatment (after 2nd, 5th, 7th, and 13th cycles). Demethylation was observed after seven rounds of treatment. Abbreviations: AML, acute myeloid leukemia; AZA, 5-azacytidine; bp, base pairs; CTL, control; DMR, differentially methylated region; MDS, myelodysplastic syndrome; NT, nontumor; TT, tumor type.

RAEB-2 and RCMD patients are not related to the stage of the disease but rather to some other, unknown factor that divides the groups. Regardless, the enrichment of functional classes, specific histone modifications, and genes with low expression indicate a general biological phenomenon that is likely to reflect specific properties of MDS.

Using different replicate groupings, we identified general DMRs (all technical replicates), tumor type DMRs (RAEB-2 vs. RCMD), treatment effect DMRs (patients MDS2, MDS3, MDS4, and MDS6) and nonleukemic (NT) DMRs. General and tumor-type DMRs were clearly hypermethylated in RAEB-2 patients compared with both RCMD and nonleukemic samples. General and tumor-type DMRs were predominantly associated with developmental regulators (particularly transcription factors), whereas MDS3 and nonleukemic DMRs were strongly enriched for hematopoietic genes. This relationship could also be observed by considering expression levels, with genes associated with TT-DMRs being expressed at low levels throughout the hematopoietic system. A similar observation has been made in breast cancer cells [36]. Taken together, These observations suggest that most hypermethylation is a consequence, rather than a cause, of disease progression.

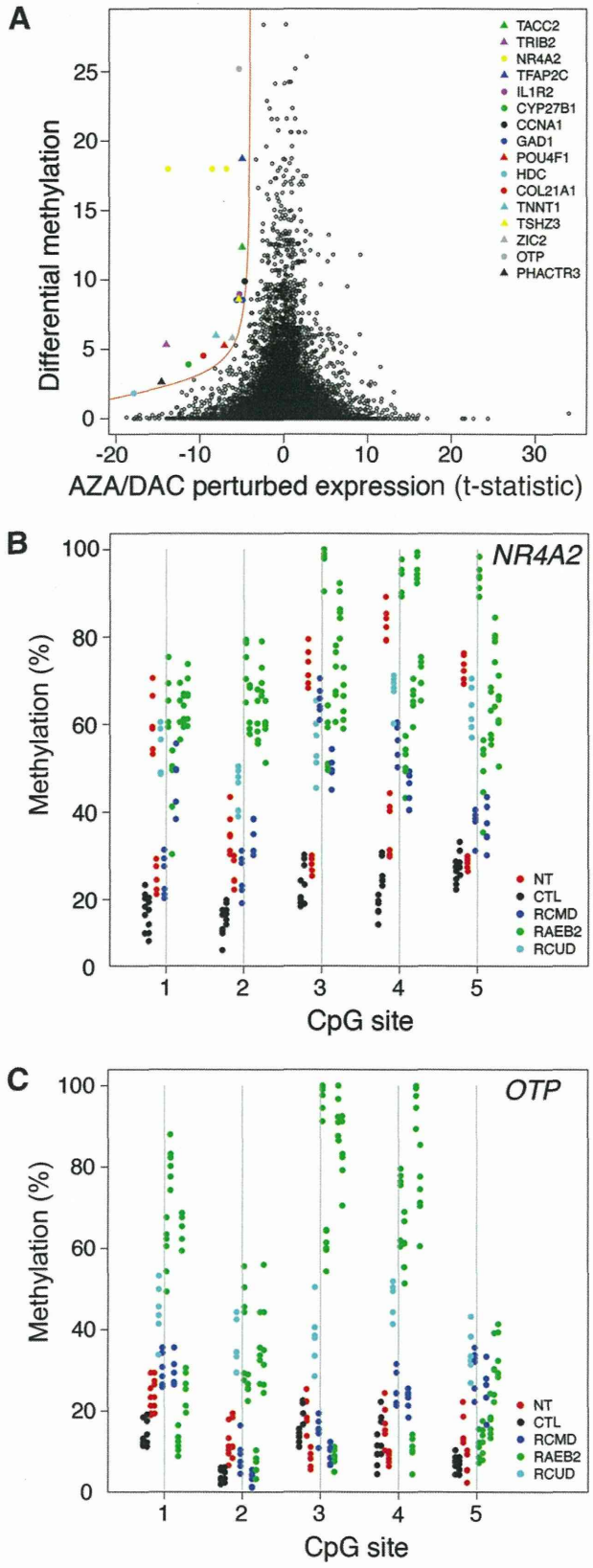

Figure 6. Identification of inducible differentially methylated genes. **(A):** The perturbation of gene expression after long-term in vitro AZA or DAC treatment (t-statistic) was plotted against aggregate TT differential methylation (supplementary methods) calculated for each gene. A line of the form $x = y / (1 + y^2)^{1/2}$ was used as a threshold to select a total of 16 genes. Color indicates the identities of the selected genes. Multiple points for the same gene indicate the presence of several probe sets representing that gene. One gene stands out from the rest, *NR4A2* (yellow), as both highly activated by AZA treatment and strongly differentially methylated between RAEB-2 and RCMD. **(B, C):** Methylation at the *NR4A2* **(B)** and *OTP* **(C)** gene loci. Five CpG positions were analyzed by pyrosequencing for each gene, using six experimental replicates. The *x*-axis indicates the CpG sites analyzed (1–5), with the point offsets indicating the sample identities: CTL1, CTL2, natural killer cells, monocytes, MDS2, MDS3, MDS4, MDS6, MDS7, MDS8, MDS9, and MDS10. All samples were obtained prior to AZA treatment. The classes of myelodysplastic syndrome are indicated by color, as indicated. Abbreviations: AZA, 5-azacytidine; bp, base pairs; CTL, control, normal CD34+ cells; DAC, 5-aza-2′-deoxycytidine; NT, nontumor; RAEB-2, refractory anemia with excess blasts type 2; RCMD, refractory cytopenia with multilineage dysplasia; RCUD, refractory cytopenia with unilineage dysplasia; TT, tumor types.

DMRs varied widely in size but generally could be distinguished clearly from the neighboring sequences (Files 1–3 in supplemental online Appendix 1). This argues that differential methylation is not caused by a general increase in DNA methylation but rather by increased methylation at specific locations either as the result of recruitment of DNA methyltransferases by sequence-specific transcription factors or through the recognition of specific histone modifications. In fact, it has been suggested previously that DMRs coincide with cis-acting elements [37], and our analyses not only support this but also suggest that cancer-associated DNA methylation occurs at a specific subtype of regulatory regions that are generally marked by H3K27me3 and that are associated with genes that regulate developmental processes. The polycomb repressor complex is known to direct H3K27me3 methylation [38], and the presence of the mark at bivalent promoters has previously been linked to hypermethylation in a wide variety of cancers [31]. Our observations suggest that the persistence of this mark in differentiated cell types is one of the factors that determine the location of cancer-associated methylation.

The large-scale analysis of TSSs throughout the genome and different cell types has suggested that gene promoters can be divided into three distinct classes: (a) promoters with well-defined TSSs that do not contain CpG islands and that are specifically expressed in differentiated cell types; (b) promoters with broad TSS regions that generally contain a single CpG island and are expressed ubiquitously; and (c) promoters with broad TSS regions, with long or multiple CpG islands in cis, that are regulated by multiple distal regions and are specifically expressed during development [17]. All our analyses (GO, histone modification, enhancer activity, gene expression and CpG island overlap) suggest that TT- and NT-DMRs are associated with type III and type I promoter-containing genes, respectively. Interestingly, treatment-stage and disease-progression DMRs had intermediate properties; MDS2 and MDS6 DMRs were more similar to the TT-DMRs, whereas MDS3 DMRs had similar properties to the NT-DMRs. It is tempting to speculate that this difference in methylation is related to the clinical outcome of treatment, but the data are too limited to support this; however, the data may be suitable for designing more targeted experiments that can be carried out for larger cohorts.

Somewhat counterintuitively, by restricting our analysis to CD34+ cells, we may have made it impossible to observe