Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia

Kristen Meldi,[1] Tingting Qin,[1] Francesca Buchi,[2] Nathalie Droin,[3] Jason Sotzen,[1] Jean-Baptiste Micol,[3,4] Dorothée Selimoglu-Buet,[3] Erico Masala,[2] Bernardino Allione,[5,6] Daniela Gioia,[6,7] Antonella Poloni,[6,8] Monia Lunghi,[6,9] Eric Solary,[3] Omar Abdel-Wahab,[4] Valeria Santini,[2,6] and Maria E. Figueroa[1]

[1]University of Michigan Medical School, Department of Pathology, Ann Arbor, Michigan, USA. [2]University of Florence, AOU Careggi, Hematology, Florence, Italy. [3]INSERM U1009, Gustave Roussy Cancer Center, Villejuif, France. [4]Memorial Sloan Kettering Cancer Center, Human Oncology and Pathogenesis Program and Leukemia Service, New York, New York, USA. [5]Hematology, Ospedale S. Giovanni Battista Molinette, Torino, Italy. [6]Fondazione Italiana per le Sindromi Mielodisplastiche (FISM), Alessandria, Italy. [7]Azienda Ospedaliera SS. Antonio e Biagio e C. Arrigo, Alessandria, Italy. [8]Hematology, Università Politecnica delle Marche, Ancona, Italy. [9]Department of Hematology, Amedeo Avogadro University, Novara, Italy.

Myelodysplastic syndromes and chronic myelomonocytic leukemia (CMML) are characterized by mutations in genes encoding epigenetic modifiers and aberrant DNA methylation. DNA methyltransferase inhibitors (DMTis) are used to treat these disorders, but response is highly variable, with few means to predict which patients will benefit. Here, we examined baseline differences in mutations, DNA methylation, and gene expression in 40 CMML patients who were responsive or resistant to decitabine (DAC) in order to develop a molecular means of predicting response at diagnosis. While somatic mutations did not differentiate responders from nonresponders, we identified 167 differentially methylated regions (DMRs) of DNA at baseline that distinguished responders from nonresponders using next-generation sequencing. These DMRs were primarily localized to nonpromoter regions and overlapped with distal regulatory enhancers. Using the methylation profiles, we developed an epigenetic classifier that accurately predicted DAC response at the time of diagnosis. Transcriptional analysis revealed differences in gene expression at diagnosis between responders and nonresponders. In responders, the upregulated genes included those that are associated with the cell cycle, potentially contributing to effective DAC incorporation. Treatment with CXCL4 and CXCL7, which were overexpressed in nonresponders, blocked DAC effects in isolated normal CD34$^+$ and primary CMML cells, suggesting that their upregulation contributes to primary DAC resistance.

Introduction

Chronic myelomonocytic leukemia (CMML) is a myelodysplastic syndrome/myeloproliferative neoplasm (MDS/MPN) overlap syndrome (1) that was historically classified within MDS (2) until 2001 (3). CMML shares many characteristics with MDS, including dysplasia in one or more myeloid cell lineages and increased risk of transformation into acute myeloid leukemia (AML). However, a distinguishing feature of CMML is the presence of persistent peripheral monocytosis (>1 × 10^9/l). CMML can be subdivided into 2 subtypes on the basis of blast count: CMML1, with less than 10% bone marrow (BM) blasts, and CMML2, which has between 10% and 19% blasts.

Substantial epigenetic abnormalities have been described in both MDS and MDS/MPN. Mutations in epigenome-modifying enzymes are highly prevalent in these disorders, including those responsible for DNA methylation and demethylation — DNA methyltransferase 3A (DNMT3A) (4) and ten-eleven translocation 2 (TET2) (5, 6), respectively — as well as those involved in histone-modifying complexes — additional sex combs-like 1 (ASXL1) (7) and enhancer of zeste homolog 2 (EZH2) (8–11). Although the precise mechanisms through which these mutations drive the aberrant epigenetic changes observed in MDS are still not completely understood, it has been shown that MDS and MDS/MPN are characterized by a DNA hypermethylation that increases with disease severity (12, 13).

MDS and MDS/MPN are resistant to conventional chemotherapies; however, epigenome-modifying drugs can be used successfully as therapeutics to treat these disorders. In particular, the nucleoside analogs azacytidine (AZA) and decitabine (DAC) are commonly used to treat MDS and CMML (14, 15). Both AZA and DAC are DNA methyltransferase inhibitors (DMTis), and while their precise mechanism of action in treating MDS and MDS/MPN remains a point of controversy, they are known to be incorporated into DNA during the S phase, where they covalently trap DNA methyltransferases and target them for proteasomal degradation (16, 17). DMTis can also cause DNA damage (18), and because AZA is mostly incorporated into RNA, it may have additional effects on RNA processing and translation (19). Despite the utility of DAC and AZA, only a subset of MDS and CMML patients respond to them. Only approximately 50% of patients treated with DMTis show a hematological improvement (HI) or better that is associated with a survival benefit (20). Furthermore, as many as 6 months of treatment may be required for the therapeutic benefit of DMTis to become apparent, thus forcing half of the patients

Authorship note: Kristen Meldi, Tingting Qin, Valeria Santini, and Maria E. Figueroa contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: August 27, 2014; **Accepted:** February 9, 2015.
Reference information: *J Clin Invest*. 2015;125(5):1857–1872. doi:10.1172/JCI78752.

Table 1. Clinical characteristics of the FISM CMML patient cohort treated with DAC

Clinical characteristics	Responders	Nonresponders	*P* value
Total no. of patients	20	20	
CMML1, no. (%)	15 (75%)	10 (50%)	NS[A]
CMML2, no. (%)	5 (25%)	10 (50%)	
Male, no. (%)	14 (70%)	14 (70%)	NS[A]
Female, no. (%)	6 (30%)	6 (30%)	
Median age, yr (range)	73.5 (45–84)	70.5 (41–82)	NS[B]
Median survival, mo (range)	26.5 (6–39)	13.5 (2–25)	$P = 0.0004$[C]
Median hemoglobin, no. (range)	10 (7.2–14.9)	9.7 (6.6–13.8)	NS[A]
Median marrow blasts, % (range)	5 (0–18)	7 (0–19)	NS[D]
Median monocytes, % (range)	24 (2–67)	22 (5–45)	NS[D]
Median wbc, % (range)	17.8 (3.7–75.2)	18.9 (2.8–52.5)	NS[A]
Cytogenetics			
Normal	14	14	NS[A]
Abnormal	6	6	
Splenomegaly	9	7	NS[A]
Hepatomegaly	8	5	NS[A]
Lymphadenomegaly	2	3	NS[A]

[A]Fisher's exact test; [B]Student's *t* test; [C]log-rank test; [D]Wilcoxon rank-sum test.

to undergo long periods of treatment before they can be deemed resistant to this therapy. Currently, there are very few means of predicting response versus resistance, and even this is exclusive to AZA (21). Additionally, few alternative treatments exist for patients who fail to respond to DMTis, and their prognosis is extremely poor. Therefore, it is critical that we better understand the molecular profiles associated with sensitivity and resistance to DMTis in order to improve risk stratification strategies as well as shed light on the mechanisms of resistance.

While some studies have suggested that reversal of methylation and/or transcript reexpression of certain loci was associated with clinical response to DMTis (22–28), epigenetic studies to date have failed to identify any strong correlation between response to these agents and the presence of specific baseline DNA methylation profiles (23, 26, 27, 29, 30). We hypothesized that this lack of correlation was due to the promoter-centric nature of assays used over the past decade and that methylation differences associated with potential for therapeutic response were likely present in these patients upon diagnosis at promoter-distal and intergenic regulatory regions. In this study, we report, for the first time to our knowledge, the identification of DNA methylation and expression differences in diagnostic BM specimens from a cohort of CMML patients treated with DAC. These differences, detected through the use of genome-wide next-generation sequencing assays, reveal underlying biological differences between these 2 groups of patients and point to a novel mechanism of resistance to DMTis.

Results

Somatic mutations do not correlate with response to DAC in CMML. Somatic mutations in epigenome-modifying enzymes and other genes are prevalent in MDS and CMML (4–6, 31–35). Recently, it has been reported that mutations in *TET2* and *DNMT3A* are associated with improved response to DMTi therapy in MDS and related disorders (36–38). Despite this, the presence of these mutations did not translate to an improved overall survival rate in any of these studies, indicating that therapeutic response and survival benefit are likely influenced by multiple different factors. Moreover, these findings have not been recapitulated in CMML exclusively (39). To determine whether particular genetic or epigenetic abnormalities are associated with DMTi sensitivity or resistance in this disease, we studied a cohort of primary CMML cases. BM mononuclear cells (BM MNCs) were collected from 40 patients with de novo CMML at the time of their diagnosis. All patients included in this study were enrolled in a clinical trial conducted by the FISM and received single-agent treatment with DAC as frontline therapy (20 mg/m^2/day for 5 days), and response was evaluated after 6 cycles of treatment. Responsive patients ($n = 20$) were defined as those who achieved either complete remission, marrow complete remission, partial remission, or HI, as defined by the 2006 International Working Group (IWG) response criteria for myelodysplasia (40). Patients with either stable disease or progressive disease were considered to have primary resistance to DAC ($n = 20$). As shown in Table 1, there were no significant differences in terms of age, gender, BM monocytosis, blast percentage, cytogenetics, or presence of either splenomegaly or extramedullary lesions between responder and nonresponder patients. Using MiSeq to sequence DNA isolated from the diagnostic BM MNCs, we performed targeted resequencing of the following panel of genes mutated at frequencies greater than 5% in CMML: *SRSF2*, *TET2*, *ASXL1*, *NRAS*, *DNMT3A*, *RUNX1*, *U2AF1*, *TP53*, *JAK2*, *KIT*, *KRAS*, *SF3B1*, *EZH2*, *IDH1*, and *IDH2*. As with previous reports, *SRSF2*, *TET2*, and *ASXL1* were the most frequently mutated genes in this cohort of patients (6, 32, 34, 35, 41–44). However, no somatic mutation was significantly correlated with response to DAC in our cohort (Fisher's exact test, *P* = NS for all mutations) (Figure 1A and Table 2).

We have previously shown, as have others, that distinct DNA methylation profiles in AML and acute lymphoid leukemia (ALL) are strongly correlated with the presence of specific molecular and cytogenetic subtypes (12, 45–48). To determine whether similarly distinct methylation patterns in CMML can be linked to the presence of specific somatic mutations, we examined DNA methylation patterns in the same specimens through enhanced reduced representation bisulfite sequencing (ERRBS) (45), a deep-sequencing method that captures and accurately quantifies DNA methylation at approximately 3 million CpG sites. ERRBS data were available for 39 of the 40 patients (19 nonresponders and 20 responders). The percentage of methylation measured by ERRBS was highly concordant with the findings of the quantitative single-locus DNA methylation validation assay MassARRAY EpiTYPER (ref. 49 and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI78752DS1). Unsupervised clustering analysis of the patients based on their DNA methylation patterns did not reveal a correlation between gene mutations and particular methylation clusters (Figure 1B). In addition, there was no significant difference in the observed

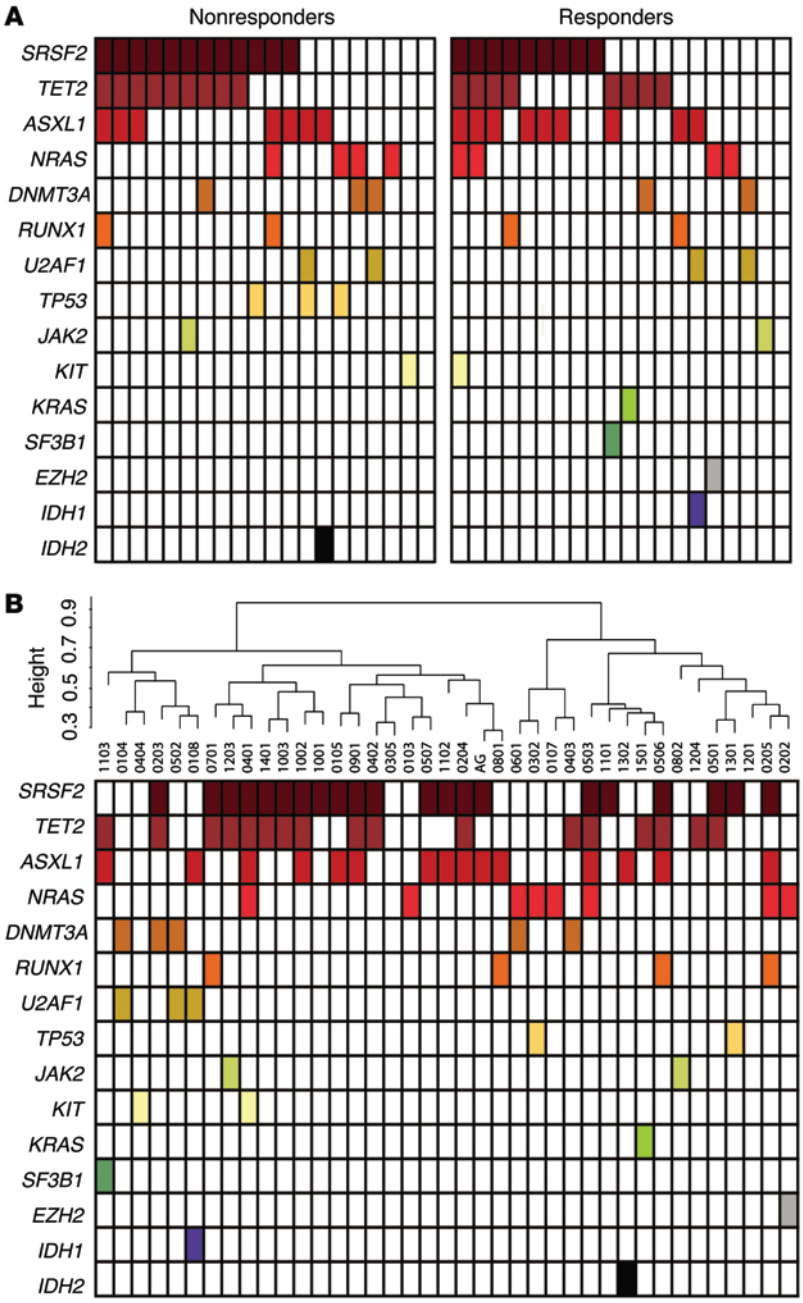

Figure 1. Somatic mutations in CMML do not correlate with DAC response or specific epigenetic clusters. Mutational status of a panel of 15 genes frequently mutated in CMML according to (**A**) therapeutic response to DAC or (**B**) DNA methylation hierarchical clustering.

patient survival time between the 2 top-level methylation clusters (log-rank test, $P = 0.33$).

Next, we performed supervised analyses comparing *TET2*, *ASXL1*, *DNMT3A*, and *SRSF2* WT and mutant cases to identify the differentially methylated regions (DMRs) associated with each of these mutations. As expected, given its role in de novo DNA methylation, we identified a predominantly hypomethylated profile associated with *DNMT3A* mutations (total DMRs: 243; hypomethylated DMRs [hypo-DMRs]: 197; hypermethylated DMRs [hyper-DMRs]: 46) that was targeted mainly at intergenic and intronic regions (Figure 2A). By contrast, *TET2* loss-of-function mutations were associated with the presence of hypermethylation compared with that seen in *TET2* WT cases (total DMRs: 188; hypo-DMRs: 48; hyper-DMRs: 140) (Figure 2B). Mutations in *ASXL1,* another epigenetic modifier, were associated with a specific signature consisting of equal proportions of hyper- and hypo-DMRs (total DMRs: 144, hypo-DMRs: 82, hyper-DMRs: 62). Both hyper- and hypo-DMRs in *ASXL1*-mutant CMML cases were strongly depleted at promoter regions (hyper-DMRs 3% vs. background 21%, $P = 6.79 \times 10^{-5}$; hypo-DMRs 5% vs. background 21%, $P = 4.30 \times 10^{-5}$) and significantly enriched at intergenic regions (hypo-DMRs 54% vs. background 38%, $P = 2.84 \times 10^{-3}$) (Figure 2C). Notably, mutations in the splicing factor *SRSF2* were linked to the strongest DNA methylation differences, with a total of 724 DMRs (hypo-DMRs: 383; hyper-DMRs: 341). In this case, hypermethylated DMRs were strongly enriched at promoter regions (hyper-DMRs 31% vs. background 21%, $P = 1.44 \times 10^{-5}$) and depleted at introns (hyper-DMRs 19% vs. background 33%, $P = 1.50 \times 10^{-8}$) (Figure 2D). While *SRSF2* itself does not have any direct epigenetic function, it is likely that mutations in this gene lead to mis-splicing and the consequent deregulation of other epigenome-modifying genes, resulting in this strong epigenetic signature. Additionally, the observed survival time was not significantly different between the patients with or without individual *DNMT3A*, *TET2*, *ASXL1*, and *SRSF2* mutations (log-rank test, $P = 0.61$, 0.067, 0.93, and 0.58, respectively).

A specific epigenetic profile distinguishes DAC-resistant CMML patients at diagnosis. Previous efforts by many groups, including ours, have failed to identify baseline epigenetic differences between DMTi-sensitive and -resistant patients (12, 27, 30). However, all of these studies were performed using platforms that examined DNA methylation within CpG islands and gene promoters. A growing body of recent evidence suggests that DNA methylation and other epigenetic modifications at enhancers and other distal regulatory regions play a key role in transcriptional regulation and that these regions are often located at a significant distance from the transcription start site of the target gene (50). Therefore, we hypothesized that key epigenetic differences may exist between DAC-sensitive and -resistant patients at diagnosis that are located distally from promoters, targeting enhancers and other distal regulatory regions.

For this purpose, we used the ERRBS assay, a deep-sequencing–based method that targets not only promoter regions but also intronic, exonic, and distal intergenic regions (45). Using the MethylSig package, we performed a direct comparison between the diagnostic DNA methylation profiles of DAC-sensitive and DAC-resistant patients (51). We identified 167 DMRs that displayed a methylation difference of 25% or more between respond-

Table 2. Somatic mutations of the FISM cohort did not correlate with response

Mutation	Nonresponders (*n* = 20)	Responders (*n* = 20)	Total (*n* = 40)	*P* value[A]
SRSF2	60.0% *n* = 12	45.0% *n* = 9	52.5% *n* = 21	0.53
TET2	45.0% *n* = 9	40.0% *n* = 8	42.5% *n* = 17	1.0
ASXL1	35.0% *n* = 7	45.0% *n* = 9	40.0% *n* = 16	0.75
NRAS	20.0% *n* = 4	20.0% *n* = 4	20.0% *n* = 8	1.0
DNMT3A	15.0% *n* = 3	10.0% *n* = 2	12.5% *n* = 5	1.0
RUNX1	10.0% *n* = 2	10.0% *n* = 2	10.0% *n* = 4	1.0
U2AF1	10.0% *n* = 2	10.0% *n* = 2	10.0% *n* = 4	1.0
TP53	15.0% *n* = 3	0.0% *n* = 0	7.5% *n* = 3	0.23
JAK2	5.0% *n* = 1	5.0% *n* = 1	5.0% *n* = 2	1.0
KIT	5.0% *n* = 1	5.0% *n* = 1	5.0% *n* = 2	1.0
KRAS	0.0% *n* = 0	5.0% *n* = 1	2.5% *n* = 1	1.0
SF3B1	0.0% *n* = 0	5.0% *n* = 1	2.5% *n* = 1	1.0
EZH2	0.0% *n* = 0	5.0% *n* = 1	2.5% *n* = 1	1.0
IDH1	0.0% *n* = 0	5.0% *n* = 1	2.5% *n* = 1	1.0
IDH2	5.0% *n* = 1	0.0% *n* = 0	2.5% *n* = 1	1.0

[A]Fisher's exact test.

ers and nonresponders and that were statistically significant at an FDR of less than 0.1. Among these DMRs were regions displaying higher methylation in responders, as well as regions of lower methylation as compared with those in nonresponders (Figure 3A and Supplemental Table 1). Hierarchical clustering of our cohort using these DMRs was sufficient to achieve a perfect segregation of DAC-sensitive and -resistant patients (Figure 3B). These findings indicate that numerous epigenetic differences exist at the time of diagnosis that correlate with a patient's likelihood of responding to DAC treatment.

Response-associated DMRs localize preferentially to distal regulatory regions. Next, we sought to determine whether DMRs were distributed evenly across the genome or whether they were enriched at specific genomic regions. For this, we analyzed both the genomic distribution of DMRs as well as their association with known regulatory regions. Notably, our analysis of the distribution of DMRs relative to coding regions revealed that DMRs were significantly depleted at promoter regions (DMRs 10% vs. background 21%, binomial test $P = 6.70 \times 10^{-5}$), with a concurrent enrichment at intronic regions, thus confirming our initial hypothesis. This distribution was not the same across hyper- and hypo-DMRs. While all DMRs were depleted at promoter regions, hyper-DMRs were significantly enriched at introns (hyper-DMRs 49% vs. background 33%, binomial test $P = 1.29 \times 10^{-3}$), while hypo-DMRs were enriched at intergenic regions (hypo-DMRs 49% vs. 38% background, binomial test $P = 0.03$) (Figure 4A).

Next, we sought to determine the association of DMRs with regulatory regions. For this purpose, we analyzed their relative enrichment at CpG island and enhancer regions. Analysis of CpG islands and CpG shores demonstrated that DMRs were also significantly depleted at CpG islands (DMRs 14% vs. background 25%, binomial test $P = 2.8 \times 10^{-4}$), with enrichment at CpG shores (DMRs 22% vs. background 15%, binomial test $P = 8.79 \times 10^{-3}$). This pattern was conserved across both hyper- and hypo-DMRs (Figure 4B).

Recently, DNA methylation at enhancers was reported to strongly correlate with aberrant gene expression observed in cancer cells (52). We hypothesized that differential DNA methylation at enhancers, rather than at promoters, may be better correlated with differential response to DAC in CMML. Enrichment analysis of all DMRs relative to intragenic and intergenic enhancers revealed that DMRs were enriched for intragenic enhancers (DMRs 25% vs. background 18%, binomial test $P = 0.01$). When this analysis was stratified into hyper- and hypo-DMRs, it became apparent that hyper-DMRs showed the strongest enrichment at enhancer regions and, in particular, at enhancers located within gene bodies (hyper-DMRs 32% vs. background 18%, binomial test $P = 8.14 \times 10^{-4}$). Conversely, hypo-DMRs were not significantly enriched at enhancer regions and were similarly distributed within gene body and intergenic enhancers (Figure 4C).

Finally, we asked whether the DMRs associated with DAC response were specifically enriched within relevant biological pathways. The 167 DMRs were annotated to known genes, and pathway enrichment analysis was performed against the KEGG pathway database. The MAPK signaling pathway, which plays a key role in the cell cycle, apoptosis, cell proliferation, and differentiation, was significantly enriched in DMR-associated genes (hypergeometric test $P = 7.68 \times 10^{-3}$, FDR = 0.084) (Supplemental Figure 2A). There were 7 DMRs that were annotated to MAPK pathway genes, including *STMN1*, *CACNAE1*, *PRKCB*, *MAPT*, *NFATC1*, *CRKL*, and *MKNK2* (Supplemental Table 2). Three of these DMRS — those annotated to *STMN1*, *CACNAE1*, and *MAPT* — were hypermethylated in DAC nonresponders, while *MKNK2*-, *NFATC1*-, *CRKL*-, and *PRKCB*-associated DMRs were hypermethylated in DAC responders (summarized in Supplemental Table 2). To further validate epigenetic deregulation of the MAPK signaling pathway in these patients, we performed MassARRAY EpiTYPER analysis of 3 of the affected MAPK genes in the pathway in a subset of samples (Supplemental Figure 2B). This analysis confirmed the increased methylation in the *STMN1* and *CACNAE1* DMRs in nonresponder patients, as well as validated the increased methylation of the *NFATC1* DMR in responders.

DNA methylation differences can be harnessed for therapeutic response prediction. Given that our data identified, for the first time to our knowledge, the existence of baseline DNA methylation differences between DAC responders and nonresponders prior to DAC treatment, we hypothesized that these unique methylation profiles could be harnessed to predict at the time of diagnosis which patients would be sensitive or resistant to treatment. To test this, we used the percentage of cytosine methylation at each genomic location among patients in the FISM cohort (cohort 1) as potential predictors and applied a machine-learning approach, support vector machine (SVM) (53), to build a classifier (see details in Methods). Twenty-one 25-bp tile regions were identified by feature selection as the predictors with the highest predictability in the SVM classifier (Figure 5A, Supplemental Figure 3A, and Supplemental Table 3). Unsupervised analysis using only the methylation levels at the 21 selected tile regions revealed that they were sufficient to almost

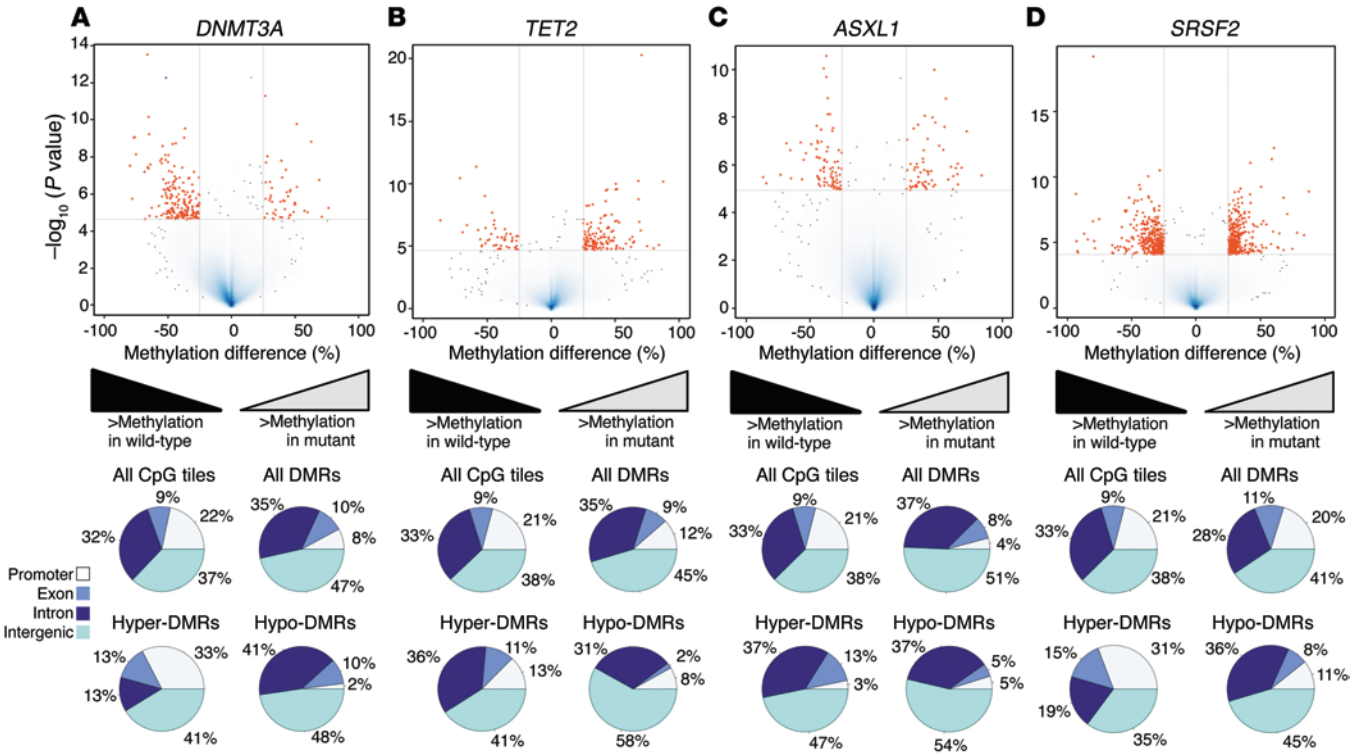

Figure 2. Distinct DNA methylation profiles are associated with recurrent somatic mutations in *DNMT3A*, *TET2*, *ASXL1*, and *SRSF2*. Volcano plots illustrating the methylation differences between *DNMT3A*-mutant (n = 5) (**A**), *TET2*-mutant (n = 17) (**B**), *ASXL1*-mutant (n = 15) (**C**), or *SRSF2*-mutant (n = 21) (**D**) samples versus WT patients (n = 39 for the number of mutated samples). DMRs are indicated by red dots (beta-binomial test, FDR <0.1 and absolute methylation different ≥25%). Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to the RefSeq promoter, exonic, intronic, and intergenic regions.

separate the 39 samples by response (Figure 5B and Supplemental Figure 3, B and C). There was no defined clustering of the patients according to their specific degree of response as shown by multidimensional scaling (MDS) analysis (Supplemental Figure 3C), which is concordant with the fact that the classifier was built to identify an all-or-nothing response versus no response and not to distinguish between types of responses. Ten-fold cross-validation was performed using the cases from cohort 1 to evaluate the predictive performance of the classifier, and the reported area under the receiver operating characteristic curve (ROC-AUC) was 0.99, indicating a strong predictive accuracy for the classifier model (Supplemental Figure 3D). In order to further assess the robustness of the SVM classifier built with the 21 selected features, we performed 3 different random splits of the same cohort 1 into training and test sets. We trained the classifier on each of the 3 sets of randomly selected samples and predicted the responses for the remaining samples in the cohort. The classifier was able to accurately predict response to DAC in 18 of 19 (accuracy = 94.74%) (Table 3), 13 of 14 (accuracy = 92.86%), and 9 of 9 (accuracy = 100%) patients, respectively (Supplemental Figure 4A).

Since validation in an independent cohort of patients is the gold standard for biomarker development, we identified a second cohort of patients in which to test the performance of our SVM classifier. Twenty-eight additional diagnostic CMML specimens from patients enrolled in a clinical trial from the Groupe Francophone des Myelodysplasies (GFM), all of whom had been treated with the same DAC regimen of 20 mg/m^2/day for 5 days, were collected and subjected to ERRBS (Table 4 and Supplemental Table 4). Specimens from this second cohort (cohort 2) of 12 responder and 16 nonresponder patients consisted of sorted monocytes from peripheral blood (PB). The SVM classifier that had been developed using cohort 1 was applied blindly to these samples, without any prior knowledge of the therapeutic response labels for this second cohort. Due to the stochastic nature of ERRBS, the CpG coverage is never identical across all samples, thus leading to missing values for some regions of interest. In effect, only 6 of the 21 features were present in all 28 samples in cohort 2. Therefore, using only these 6 features, we first trained our SVM classifier on the 39 samples of the FISM cohort (cohort 1) and then applied the trained classifier on the GFM cohort (cohort 2). As shown in Table 5 and Supplemental Figure 4B, despite this limitation, the 6-feature classifier was still capable of correctly predicting response in 20 of 28 patients in the GFM cohort (accuracy = 71% and AUC = 0.82). Next, in order to increase the number of features being tested while still retaining a large enough cohort in which to test the predictive accuracy, we used 14 of the 21 features of the SVM classifier to predict response for 19 patients in the GFM cohort. Once again, we used only these 14 features to train the model on cohort 1, which consisted of the initial 39 patients, and then blindly applied the model to the 19 test samples from the GFM cohort (cohort 2). This modified classifier with 14 features was capable of accurately predicting therapeutic outcome for 15 of the 19 patients, which represents an accuracy of 79% and an AUC of 0.83. (Table 5 and Supplemental Figure 4B). Finally, we determined that of the original 21 features, 16 was the

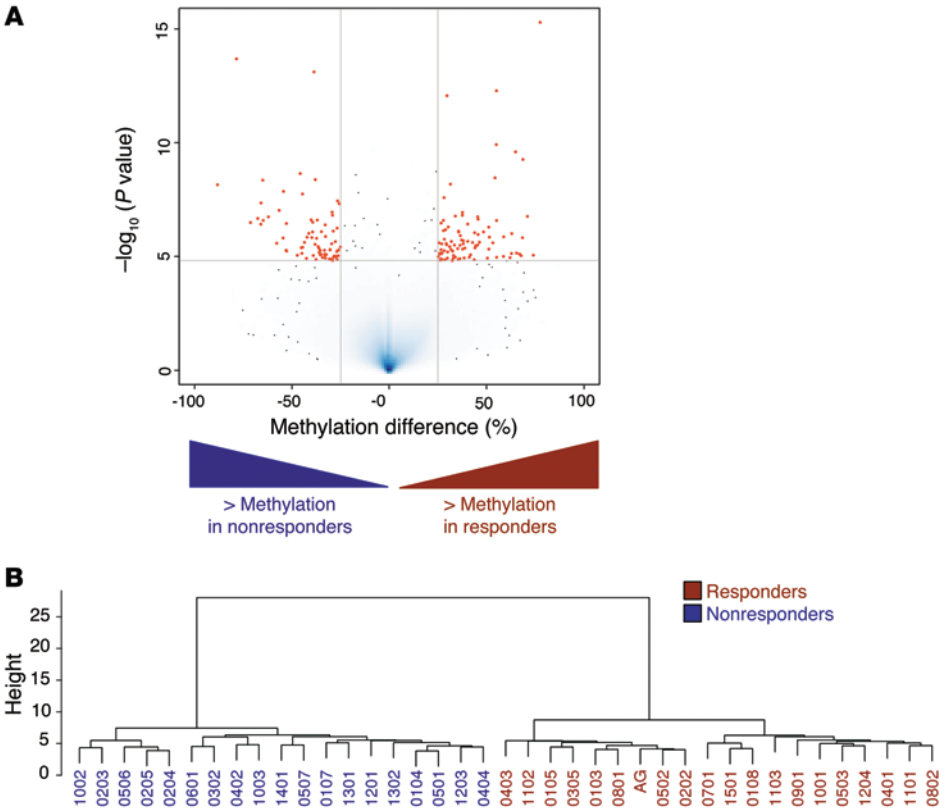

Figure 3. Baseline DNA methylation differences distinguish DAC responders and nonresponders at the time of diagnosis. (**A**) Volcano plot illustrating methylation differences between 20 DAC-sensitive and 19 DAC-resistant patients. Mean methylation difference between the 2 groups is represented on the *x* axis and statistical significance ($-\log_{10}$ *P* value) on the *y* axis. Beta-binomial test identified 167 DMRs, which are indicated by red dots (FDR <0.1 and absolute methylation difference ≥25%). (**B**) Hierarchical clustering of the patients using the 167 DMRs illustrates the power of these genomic regions in segregating the patients into nonresponder (blue) and responder (red) groups.

maximum number of features shared by at least 15 of the cohort-2 patients. We trained the model on cohort 1 using only these 16 shared features and then predicted response for the 15 patients in the independent cohort 2, achieving an accuracy of 87% with an AUC of 0.94 (Table 5 and Supplemental Figure 4B). These findings demonstrate that the SVM classifier developed using the original FISM cohort is general enough to be applied to and accurately predict the therapeutic outcome of fully independent samples (i.e., GFM cohort 2), which is a critical step in the development of a biomarker. Moreover, this robustness was maintained even across different cell types (BM MNCs in cohort 1 vs. PB monocytes in the validation cohort 2), further underscoring the power of the classifier to predict outcome in an independent cohort. While further validation in larger cohorts will be required to fully assess the accuracy of the features reported here, and additional studies of larger cohorts might help refine the selection of features to include those with the strongest accuracy over a large number of patients, our findings demonstrate that the epigenetic differences between responders and nonresponders at diagnosis have the potential to be harnessed as classifiers to predict clinical response to DAC.

DAC sensitivity can be linked to a specific transcriptional program at diagnosis. While it has been previously shown that reduced expression of uridine-cytidine kinase, an enzyme involved in nucleoside metabolism, is associated with response to AZA in MDS (54), we did not find that differential expression of this or other DMTi-metabolizing enzymes was associated with response to DAC in CMML (data not shown). Therefore, we sought to determine whether other transcriptional differences between DAC responders and nonresponders are indicative of response and can provide insight on functional pathways that contribute to DAC resistance. We performed RNA-sequencing (RNA-seq) on samples from 14 patients (8 responders and 6 nonresponders) in the cohort of CMML patients for whom we had high-quality RNA. Prior to performing differential analysis, we validated the ability of our RNA-seq approach to accurately detect quantitative variability by performing quantitative reverse transcriptase PCR (qRT-PCR) on RNAs from 13 of the 14 patients and determining the degree of agreement between the 2 methods ($r = 0.85$, R^2 value = 0.73, $P < 0.0001$) (Supplemental Figure 5A). As shown in Figure 6A, a direct comparison of the 2 groups of patients identified 601 genes with an absolute $\log_2$ fold change greater than 1 and a *P* value of less than 0.05. Notably, this gene signature consisted of a majority of genes overexpressed in DAC-sensitive patients (405 upregulated genes), with only a small proportion of genes downregulated in these patients (Supplemental Table 5).

In order to identify biological differences that might explain the difference between these patients in their therapeutic response to DAC, we performed gene set enrichment analysis (GSEA) (55). Gene sets enriched in DAC-sensitive patients at an FDR of less than 0.1 were involved in proliferation, cell cycle activity, and DNA replication (Figure 6B). Likewise, genes reported as being downregulated in quiescent versus dividing $CD34^+$ cells (56) were found to be upregulated in DAC responders. This enrichment of gene sets involved in the cell cycle and in DNA replication in DAC-sensitive patients is consistent with the requirement for DAC incorporation into the DNA during the S phase.

Primary resistance to DAC is associated with overexpression of ITGβ3 and the chemokines CXCL4 and CXCL7. As mentioned above, only a small fraction of genes were found to have at least

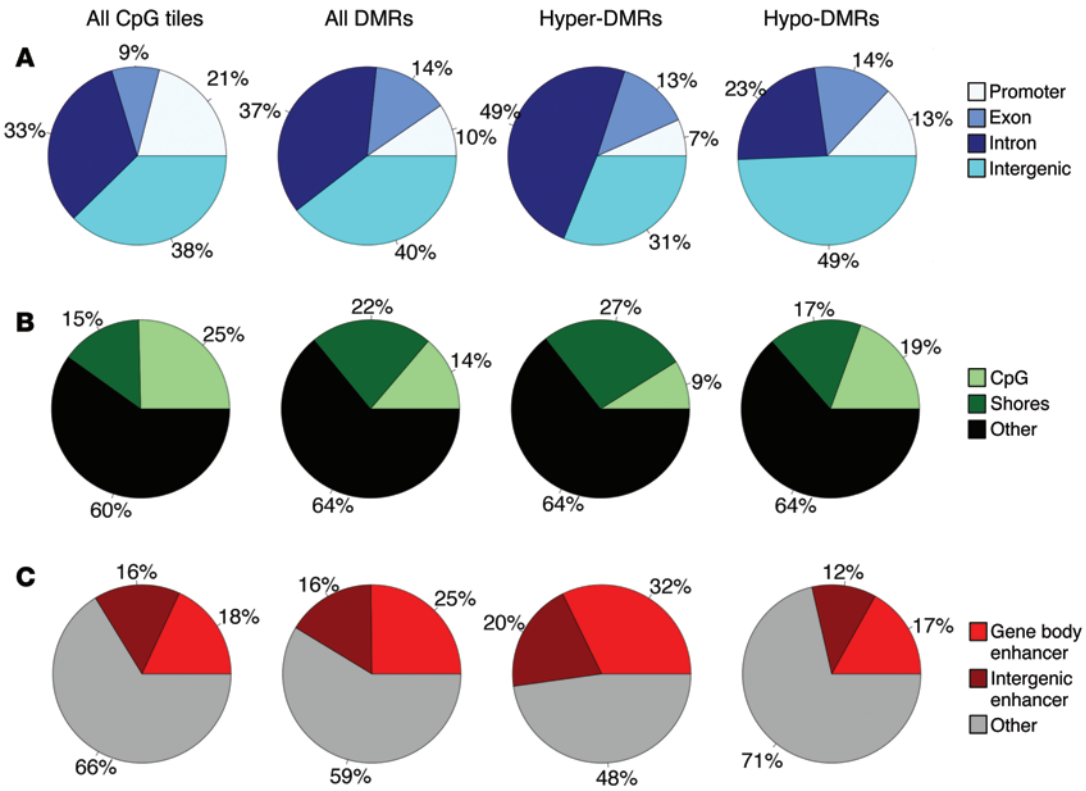

Figure 4. DMRs are enriched at distal intergenic regions and enhancers. (**A**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to RefSeq promoter, exonic, intronic, and intergenic regions. (**B**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to CpG islands, CpG shores, and regions beyond CpG shores. (**C**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to enhancers within gene bodies, enhancers within intergenic regions, and nonenhancer regions.

a 2-fold overexpression in DAC-resistant patients. Among these, 3 genes that have previously been implicated in chemoresistance and leukemogenesis were overexpressed in nonresponders: *CXCL4* (also known as *PF4*), *CXCL7* (also known as *PPBP*), and integrin β3 (*ITGB3*) (Figure 6C). Thus, we hypothesized that overexpression of these genes might be a potential mechanism through which CMML acquires resistance to DAC. First, as shown in Figure 7A, we validated the overexpression of these genes in DAC-resistant patients by qRT-PCR. Notably, there was a statistically significant linear correlation between the levels of CXCL4 and CXCL7 expression by both RNA-seq ($r = 0.9350$, $R^2 = 0.87$, $P < 0.0001$) and qRT-PCR ($r = 0.9865$, $R^2 = 0.9731$, $P < 0.0001$), suggesting that these factors act in concert in the BM microenvironment (Figure 7B). While both chemokines were originally thought to be produced exclusively by megakaryocytes, there is evidence that monocytes (57, 58) and other cells within the BM also produce CXCL4 and CXCL7 (refs. 59, 60, and Supplemental Figure 5, B and C). To further confirm the overexpression of these chemokines in nonresponder patients as well as to determine the cellular source and localization of the proteins in the BM, IHC was performed on a subset of paraffin-embedded BM biopsies taken at diagnosis from responders and nonresponders. As shown in Figure 7, C and D, CXCL4 was primarily localized to megakaryocytes, while CXCL7 staining was stronger in an MNC population compatible with a monocytic origin. Importantly, there was increased CXCL4 and CXCL7 staining in BM from nonresponder patients as compared with that in BM from responders, confirming the presence of CXCL4 and CXCL7 proteins in the BM microenvironment, which, like mRNA levels, are increased in DAC-resistant patients.

Previous studies have implicated serum levels of CXCL4 and CXCL7 as potential prognostic markers in MDS (61, 62). To determine whether serum levels of CXCL4 and CXCL7 could potentially serve as biomarkers for DAC response, we quantified the serum concentrations of these chemokines by ELISAs in 35 of 40 CMML patients (Supplemental Figure 6). There was no significant difference in serum CXCL4 or CXCL7 levels between responders and nonresponders. In addition, we found no significant correlation between BM mRNA levels and serum protein levels for these 2 chemokines, indicating that serum levels of these chemokines are not reflective of mRNA expression in the BM and mirroring previous observations documented for other chemokines in the BM and serum of AML patients (63, 64).

CXCL4 and CXCL7 abrogate the effect of DAC on hematopoietic cells. It has been previously reported that both CXCL4 and CXCL7 can reduce the chemosensitivity of BM cells to 5-fluorouracil in vitro (65), and CXCL4 has been implicated in cell cycle arrest (66) and quiescence (67, 68), which might be a mechanism through which it acts to prevent sufficient incorporation of DAC into cells of nonresponders. Therefore, we hypothesized that an overabundance of CXCL4 and CXCL7 in the BM microenvironment acts to overcome the effects of DAC. To test this, we cultured primary human $CD34^+$ cells for 3 days in vitro with CXCL4 (50 ng/ml), CXCL7 (50 ng/ml),

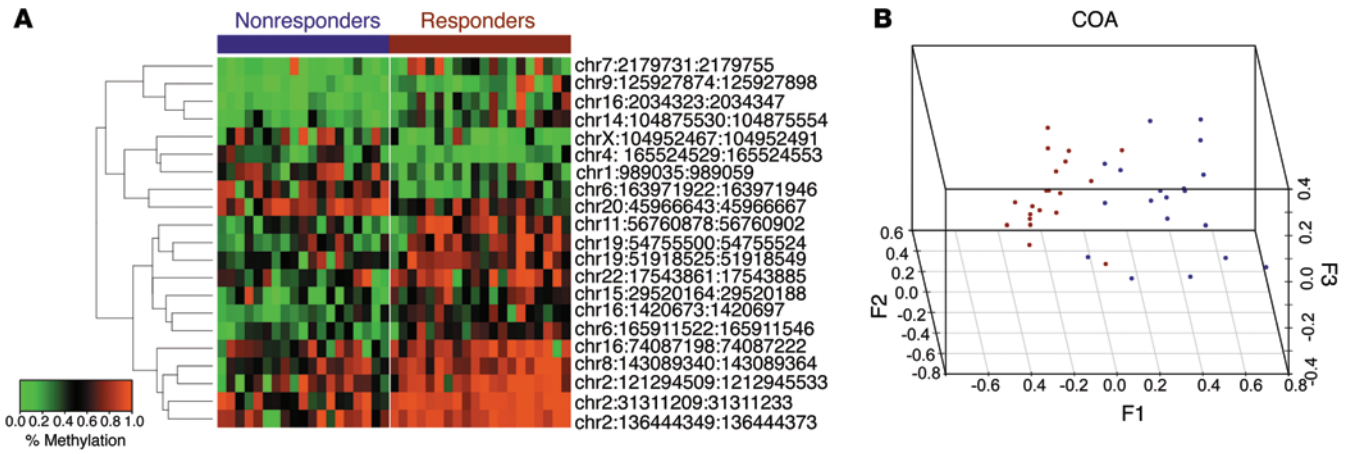

Figure 5. Methylation profiles can be harnessed to classify patients according to DAC response at diagnosis. (**A**) Heatmap of 21 CpG tiles selected as the SVM classifier predictors. DAC-sensitive patients are indicated with the red bar and nonresponders with the blue bar. (**B**) Correspondence analysis (COA) using only the 21 CpG tiles included in the classifier could segregate the majority of the CMML cohort according to DAC response (responders are represented by red dots and nonresponders by blue dots).

or a combination of both chemokines in either the presence or absence of low-dose DAC (10 nM) and then plated them in methylcellulose to test their clonogenic potential. The chemokines and low-dose DAC did not affect cell proliferation during the in vitro liquid culture period (Supplemental Figure 7A). Moreover, as previously reported, low-dose DAC did not reduce cell viability or induce apoptosis after 3 days in culture (Supplemental Figure 7, B and C, and ref. 69). However, 3 days of treatment with 10 nM DAC significantly reduced colony formation. Addition of either CXCL4 or CXCL7 alone did not have a significant impact on DAC-induced colony inhibition. However, concomitant treatment of $CD34^+$ cells with CXCL4 and CXCL7 completely abolished the suppressive effect of DAC on colony formation (Figure 8A).

Finally, we tested the ability of CXCL4 and CXCL7 to induce resistance in primary CMML cells. BM MNCs from diagnostic specimens collected from 3 patients were placed in liquid culture and treated for 72 hours with 10 nM DAC in the presence or absence of 50 ng/ml CXCL4, CXCL7, or a combination of both. Viability was assessed after 72 hours. Unlike normal $CD34^+$ cells, which did not show diminished viability with 10 nM DAC (Supplemental Figure 5B), treatment of primary CMML cells with low-dose DAC led to a significant decrease in viability in all 3 patients ($P < 0.01$). However, concomitant treatment of CMML cells with CXCL4, CXCL7, or their combination abrogated the effect of DAC on all 3 patients (Figure 8B). Combined, these data support the hypothesis that the presence of excess CXCL4 and CXCL7 in the marrow microenvironment contributes to induction of DAC resistance in CMML cells.

Discussion

While DMTis remain the only FDA-approved therapy for the majority of MDS and nonproliferative CMML patients, prognosis following DMTi treatment failure is extremely poor, with median survival for these patients barely reaching 6 months and approximately 50% of patients never even achieving a response in the first place (20, 70). This relatively low rate of therapeutic response is further complicated by the slow kinetics of DMTis, which may take as long as 6 to 12 months to show efficacy, thus committing the majority of patients to receive a drug to which they will ultimately be deemed resistant. Therefore, we set out to study the epigenetic and transcriptional characteristics associated with response to DAC in a cohort of CMML patients in order to identify molecular features that allow risk stratification at the time of diagnosis and, additionally, to explain the mechanisms behind the primary resistance to this agent. To better understand the molecular and mechanistic basis for DMTi response and effectively risk-stratify patients at diagnosis, we performed next-generation sequenc-

Table 3. Prediction performance of the SVM classifier trained on 20 randomly selected samples and applied to the remaining 19 samples in the FISM cohort (accuracy = 94.74%)

Patient ID	Original label	Prediction
1002	NR	NR
0402	*NR*	*R*
0501	NR	NR
0502	R	R
0103	R	R
0105	R	R
0205	NR	NR
0202	R	R
1301	NR	NR
1302	NR	NR
1101	R	R
0204	NR	NR
0507	NR	NR
0802	R	R
0404	NR	NR
0108	R	R
1103	R	R
0901	R	R
0701	R	R

NR, nonresponder; R, responder. Italics indicate an incorrect prediction.

Table 4. Clinical characteristics of the GFM CMML cohort treated with DAC

Clinical characteristics	Responders	Nonresponders	*P* value
Total no. of patients	12	16	
CMML1, no. (%)	2 (17%)	10 (62.5%)	$P = 0.0235^A$
CMML2, no. (%)	10 (83%)	6 (37.5%)	
Male, no. (%)	9 (75%)	13 (81%)	NS[A]
Female, no. (%)	3 (25%)	3 (19%)	
Median age, yr (range)	72.5 (61–88)	71 (55–85)	NS[B]
Median survival, mo (range)	39 (8–95)	14.5 (5–67)	NS[C]
Median hemoglobin, % (range)	9.1 (6.7–13.3)	9.05 (8–12.2)	NS[A]
Median marrow blasts, % (range)	14 (3–20)	9 (4–19)	NS[D]
Median monocytes, % (range)	23 (2–47)	15.5 (3–34)	NS[D]
Median wbc, % (range)	18.9 (4.9–77.5)	24.95 (4.1–81.7)	NS[A]
Cytogenetics			
Normal	7	11	NS[A]
Abnormal	5	5	

[A]Fisher's exact test; [B]Student's *t* test; [C]log-rank test; [D]Wilcoxon rank-sum test.

ing assays to study both the epigenome and the transcriptome of a uniformly treated cohort of CMML patients who differed in their response to DAC. The use of this improved technology, with extended genomic coverage and better dynamic range, allowed us to detect, for the first time to our knowledge, the presence of DNA methylation and gene expression differences present at the time of diagnosis that distinguish DMTi-sensitive and -resistant patients. The enrichment of these DMRs at distal enhancers, as well as the depletion of promoter-associated DMRs identified in this baseline epigenetic signature, underscores the importance of analyzing DNA methylation changes beyond promoter regions and explains the lack of statistically significant differential methylation observed in previous studies that were confined solely to promoter methylation analysis (12, 27, 30).

Moreover, our observation that the genomic locations predominantly affected by differential DNA methylation are distal regulatory regions adds more data to the strong evidence that emphasizes the critical role of long-range epigenetic gene regulation. Techniques to examine 3D chromatin architecture, such as chromosome conformation capture (3C) (71) and its subsequent iterations 4C (72, 73), 5C (74), and Hi-C (75), have indicated that gene regulation often occurs at very distant locations, in part through DNA looping at distal enhancers. In fact, only a small percentage (~7%) of gene-looping events have been reported to involve the nearest gene transcription start site (50). This argues for the critical role of distal, nonpromoter regulatory regions in controlling gene expression. If the differential methylation at nonpromoter regions does impact the expression of long-range target genes, this may explain why several previous studies have struggled to correlate differential DNA methylation with gene expression changes using nearest-gene annotations (30, 76).

We found that the MAPK pathway was significantly enriched in DMRs, with both gains and losses of methylation in responders and nonresponders within this pathway. These DMRs were localized to both intra- and intergenic genomic regions annotated for 7 genes involved in the MAPK pathway. While in-depth functional analysis of these DMRs will be required in additional experiments that are beyond the scope of our study, our findings support the results by others suggesting the importance of aberrant MAPK pathway signaling in contributing to MDS/MPN (77, 78), as well as to drug resistance and cell cycle progression in leukemic cells (79, 80). Furthermore, while it is known that multiple genes in the MAPK pathway can be mutated in CMML (81), our results indicate that the epigenetic alterations of genes in this pathway may also be present in CMML patients.

While previous reports on MDS and related malignancies have linked the presence of certain mutations — specifically, *TET2* (36–38) and *DNTM3A* (37) — to an increased rate of response to DMTis, we could not find any correlation between the mutational status of these and other genes commonly mutated in CMML and response to DAC in our FISM cohort. This finding is in concordance with those of a previous report on CMML (39), which likewise failed to detect a correlation between response to DAC and mutational status, indicating that the impact of mutational status may be different in CMML patients compared with that in MDS patients or in mixed cohorts consisting of MDS patients as well as patients with other myeloid malignancies, including AML (37, 38) and MDS/MPN (37). Furthermore, the studies demonstrating better *TET2*- and *DNMT3A*-associated responses involved patients treated with AZA alone (38) or cohorts including both AZA- and DAC-treated patients (36, 37), which may also contribute to the differing result obtained in our study on patients who received DAC exclusively.

Conversely, DNA methylation status was indeed different at diagnosis between DAC-sensitive and DAC-resistant patients, and we demonstrate that these differences can risk-stratify patients at the time of diagnosis using an epigenetic classifier that exploits these identified methylation differences. Moreover, the SVM classifier developed in this study performed with 87% accuracy on an independent cohort, even when only a subset of the original features were included and 2 different cell types were used in the training and validation cohorts (BMN MNCs vs. PB monocytes). Thus, while the classifier reported here will require further extensive validation in larger, independent cohorts, the present study demonstrates not only that DNA methylation differences exist between patients with different responses to DAC but that these DNA meth-

Table 5. Summary of the prediction performance of the independent validation cohort (GFM) in 3 scenarios using an increasing number of shared features of the 21 features preselected from the FISM cohort

Number of features used	Correct predictions/ Total patients	Accuracy (%)
16	13/15	87%
14	15/19	79%
6	20/28	71%

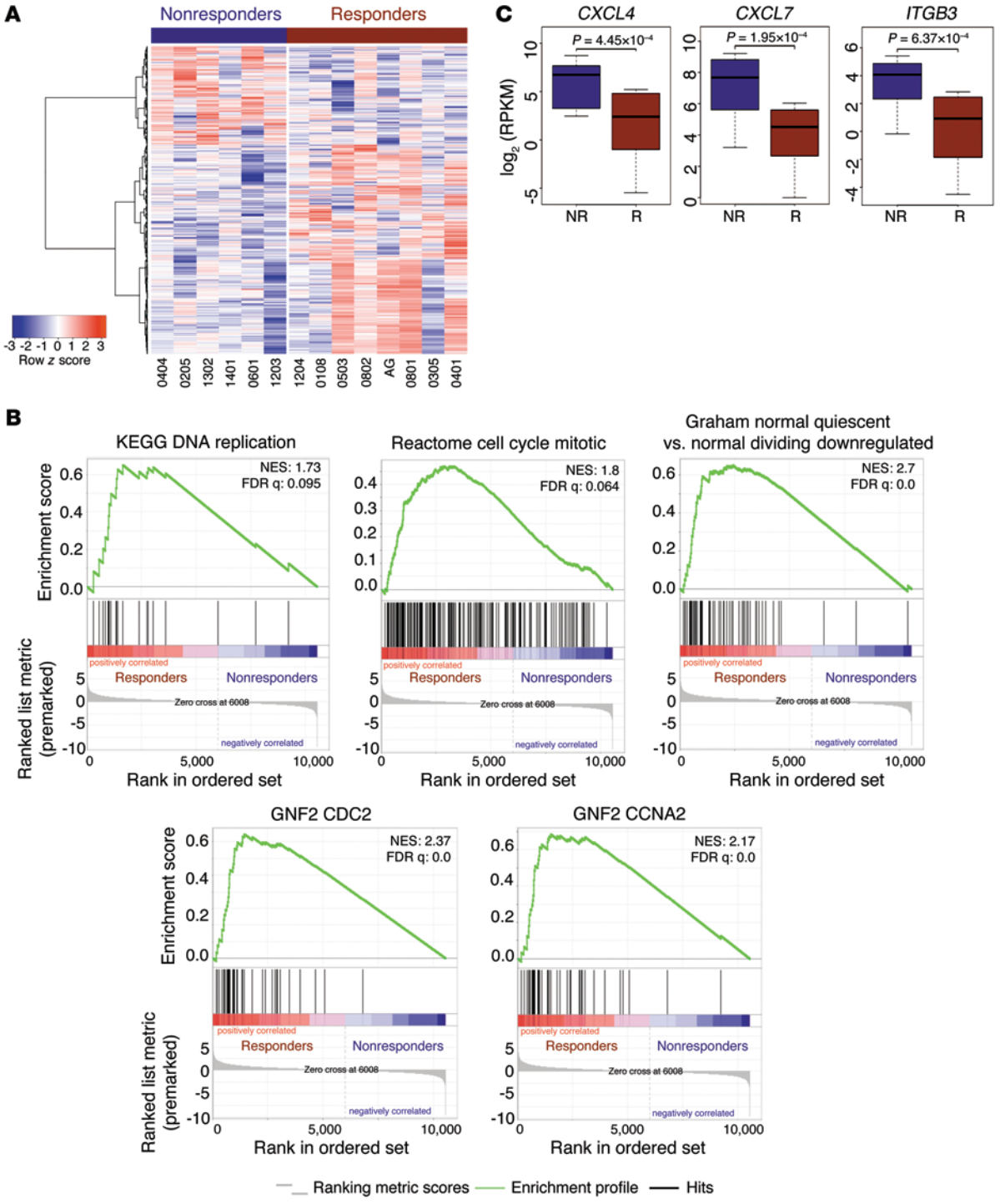

Figure 6. A specific transcriptional program is associated with response to DAC. (**A**) Heatmap illustrates gene expression differences between 8 DAC-sensitive (indicated by the red bar) and 6 DAC-resistant patients (indicated by the blue bar). Genes represented in the heatmap were identified by a GLM likelihood ratio test ($P < 0.05$ and absolute $\log_2$ fold change >1). (**B**) Enrichment plots for GSEA using the expression difference–ranked gene list showing enrichment for cell cycle–related gene sets. NES, normalized enrichment score. (**C**) Box plots showing gene expression differences for *CXCL4*, *CXCL7*, and *ITGB3* (red box plots denote responders; blue box plots denote nonresponders). *P* values were obtained from a GLM likelihood ratio test.

ylation differences are sufficiently robust to be harnessed for use in the clinic as accurate classifiers. These classifiers have the potential to prevent patients who are unlikely to respond to DAC from receiving prolonged, unwarranted treatments with this drug and instead permit them to be quickly transitioned to alternative therapies.

In addition to epigenetic differences, our study also revealed baseline differences at the transcriptional level that correlated with response to DAC. Analysis of this response-associated signature demonstrated a strong enrichment for gene sets involved in cell cycle regulation among the genes upregulated in DMTi-sensitive

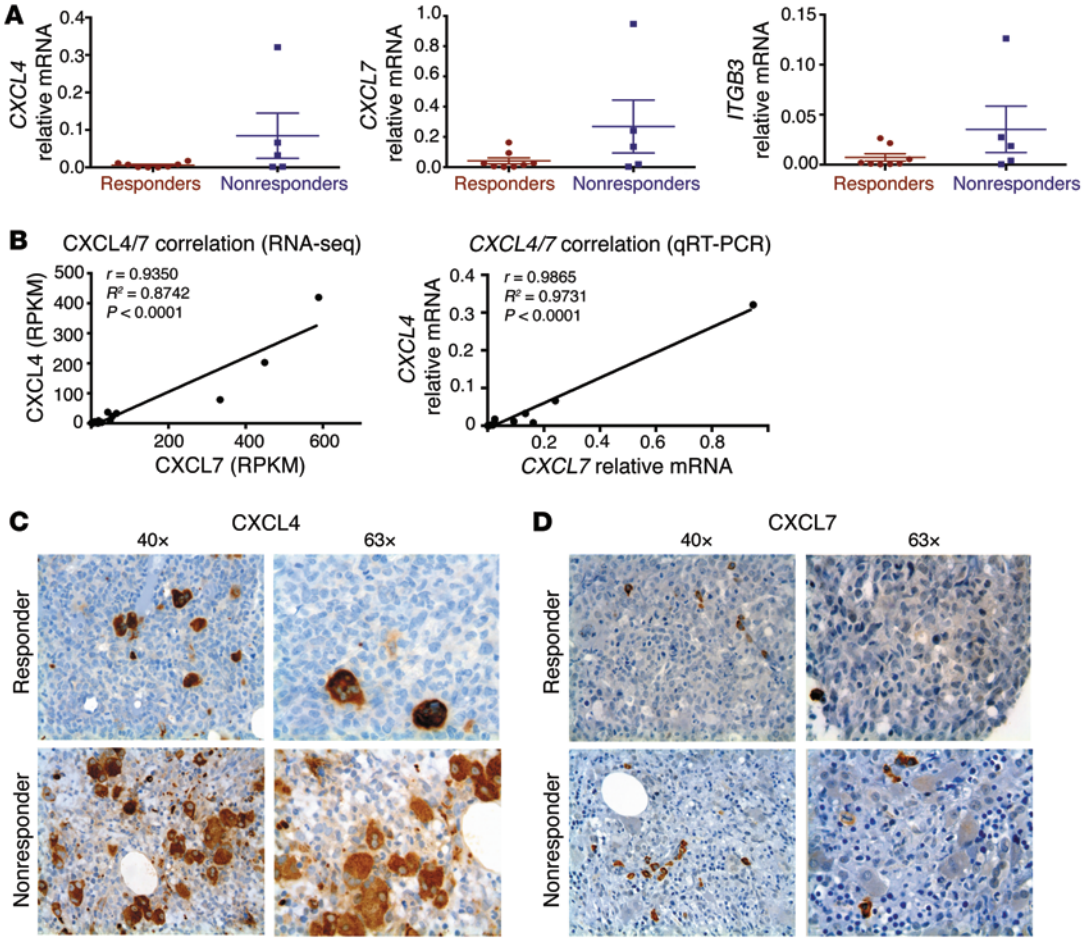

Figure 7. CXCL4 and CXCL7 are upregulated in the BM of nonresponders. (**A**) qRT-PCR showing validation of overexpression of *CXCL4*, *CXCL7*, and *ITGB3* in nonresponders; each point represents the mean of triplicate wells for each patient sample; the line and error bars indicate the group mean and SD, respectively. (**B**) Pearson's correlation analysis of expression levels of CXCL7 and CXCL4 by RNA-seq and qRT-PCR. (**C** and **D**) Representative IHC images for CXCL4 (**C**) and CXCL7 (**D**) in diagnostic BM biopsies in DAC responders and nonresponders. Original magnification, ×40 (**C** and **D**, left panels), ×63 (**C** and **D**, right panels). Representative images from duplicate experiments are shown.

patients. This finding is in line with the need for DAC to be incorporated into the DNA during cell cycle activity in order to exert its effects. By contrast, fewer genes were upregulated in resistant patients. Among these overexpressed genes, we found CXCL4 and CXCL7, two chemokines that have been previously implicated in mediating cell cycle arrest (66), quiescence (67, 68), and reduced chemosensitivity of BM cells to 5-fluorouracil in vitro (65). We therefore focused our efforts on studying the impact of these chemokines on response to DAC. In vitro treatment of both normal $CD34^+$ cells or primary CMML MNCs with CXCL4 and CXCL7 blocked the effect of DAC on these cells, indicating that overexpression of these 2 genes may indeed lead to primary resistance to DAC and opening the possibility for future targeting of the downstream signaling cascades in order to overcome the effect of these chemokines.

Methods

Sample collection and processing

FISM cohort. BM specimens were collected before treatment from 40 patients with CMML. BM MNCs were isolated through Ficoll density centrifugation and viably frozen in 10% DMSO and 90% FBS. Patients with advanced CMML were enrolled in the nonrandomized clinical trial conducted by the FISM (NCT01251627; https://clinicaltrials.gov/) and were given DAC (20 mg/m²/day i.v.) for 5 days every 28 days for at least 6 cycles prior to being classified as responders or nonresponders, with response defined as HI or better according to IWG 2006 criteria (40). The clinical characteristics of the patients are summarized in Table 1. Genomic DNA and total RNA were isolated using the AllPrep DNA/RNA kit (QIAGEN) according to the manufacturer's instructions.

GFM cohort. The patients were enrolled in the EudraCT 2008-000470-21 GFM trial (NCT01098084; https://www.clinicaltrials.gov/) and received DAC (20 mg/m²/day i.v.) for 5 days every 28 days for at least 3 cycles. Blood samples were collected using EDTA-containing tubes, mononucleated cells were isolated on Ficoll-Hypaque, and monocytes were enriched using the AutoMacs system (Miltenyi Biotec) through negative selection with microbeads conjugated with antibodies against CD3, CD7, CD16, CD19, CD56, CD123, and glycophorin A, then further enriched by positive selection with microbeads conjugated with a monoclonal mouse anti-human CD14 antibody (Miltenyi Biotec). Genomic DNA was extracted from the monocytes

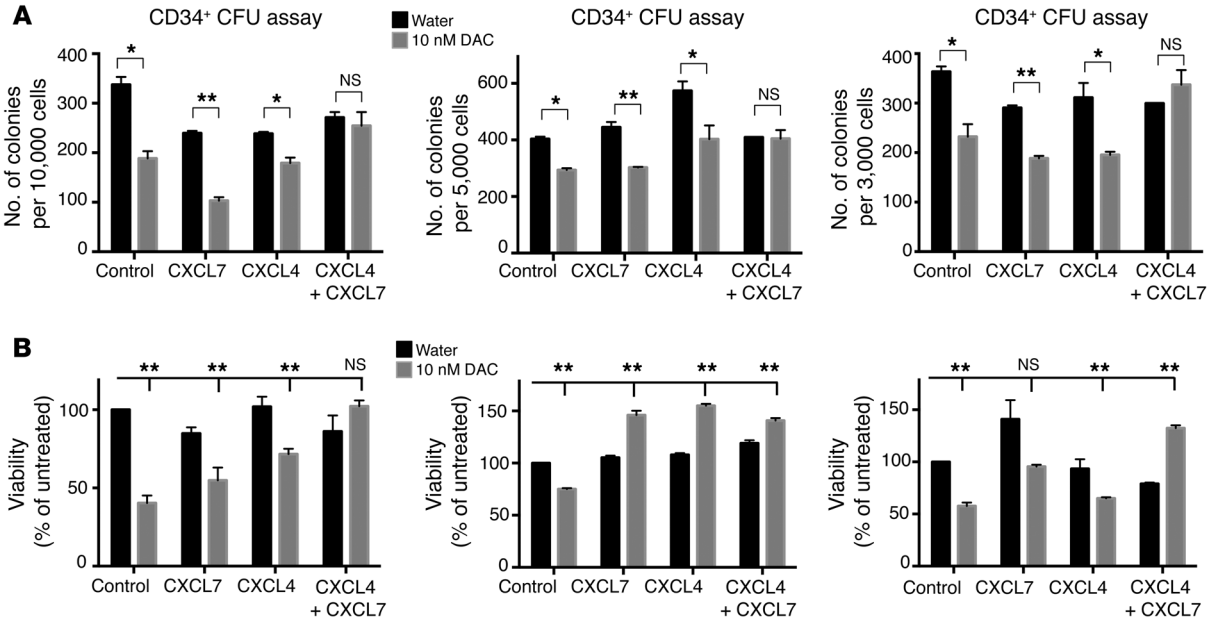

Figure 8. CXCL4 and CXCL7 promote resistance to DAC in CD34⁺ and primary CMML specimens. (**A**) Colony formation was inhibited by DAC but restored with the combination of CXCL4 and CXCL7. CD34⁺ cells were treated with 1 dose of CXCL4, CXCL7, or both (50 ng/ml each) or with vehicle (PBS containing 0.1% BSA) and daily 10-nM doses of DAC for 3 days. After 3 days of in vitro treatment with DAC, cells were plated in methylcellulose and incubated for 12 to 15 days before colonies were counted. Data represent the mean ± SD. Treatment with 10 nM DAC significantly decreased colony formation but failed to do so in the presence of CXCL7 and CXCL4 together. Shown in the 3 panels are the results of 3 independent experiments. Error bars represent the SD. (**B**) CXCL4 and CXCL7 abrogated the effect of DAC on the viability of primary CMML MNCs. CMML MNCs were treated in vitro for 72 hours with 10 nM DAC alone or in the presence of 50 ng/ml CXCL4, CXCL7, or both. Data represent the mean ± SD. Treatment with DAC alone significantly reduced the viability of these cells, but this effect was lost when CXCL4 or CXCL7 was added to the culture. All data represent independent experiments performed in 3 different CMML patients. Error bars represent the SD. $*P < 0.05$ and $**P < 0.01$ by unpaired 2-tailed Student's *t* test.

using the Norgen Biotek kit (Thorold) kit according to the manufacturer's instructions. The clinical characteristics of the patients are summarized in Table 4.

Mutational sequencing

Target capture. Capture of the target regions (exons plus splice junctions) was carried out using a custom-designed HaloPlex Target Enrichment kit (Agilent Technologies) following the HaloPlex Target Enrichment System-Fast Protocol, version D.5.

Sequencing. DNA (500 ng) from each sample was quantified with a Qubit Fluorometer (Invitrogen) and used in the capture reaction. Each sample had a unique index. Libraries were quantified by Qubit, pooled, and run in an Illumina HiSeq 2500 rapid-run flow cell using the on-board cluster method for paired-end sequencing (2 × 100 bp reads).

Analysis. Sequencing results were demultiplexed and converted to a FASTQ format using Illumina BCL2FASTQ software. The reads were adapter and quality trimmed with Trimmomatic (82) and then aligned to the human genome (UCSC build hg19) using the Burrows-Wheeler Aligner (83). Further local indel realignment and base-quality score recalibration were performed using the Genome Analysis Toolkit (GATK) (84). Single-nucleotide variation and indel calls were generated with the GATK HaplotypeCaller. ANNOVAR (85) was used to annotate variants with functional consequence on genes as well as to identify the presence of these variants in dbSNP 137, the 1000 Genomes Project, ESP6500 (National Heart, Lung, and Blood Institute [NHLBI] GO Exome Sequencing Project), and COSMIC 67.

Genome-wide DNA methylation by ERRBS

High-molecular-weight genomic DNA (25 ng) was used to perform the ERRBS assay as previously described (45) and was sequenced on an Illumina HiSeq 2000. Reads were aligned against a bisulfite-converted human genome (hg18) using Bowtie and Bismark (86). Downstream analysis was performed using R statistical software (version 3.0.3) (87), Bioconductor 2.13 (88), and the MethylSig 0.1.3 package (51). Only genomic regions with coverage ranging from 10 to 500 times were used for the downstream analysis. DMRs were identified by first summarizing the methylation status of the genomic regions into 25-bp tiles and then identifying regions with an absolute methylation difference of 25% or more and an FDR of less than 10%. DMRs were annotated to the RefSeq genes (NCBI) using the following criteria: (a) DMRs overlapping with a gene were annotated to that gene; (b) intergenic DMRs were annotated to all neighboring genes within a 50-kb window; and (c) if no gene was detected within a 50-kb window, then the DMR was annotated to the nearest transcription start site (TSS).

Methylation classifier

SVM (53) was applied using R package e1071 (89) to classify the 2 groups of patients (responders and nonresponders), in which the percentage of methylation of the 25-bp tiles was used as a predictor. The probability mode and sigmoid kernel were used in the SVM function, otherwise the default parameters were applied. We performed 2-step feature selections for the SVM classifier: (a) 25-bp tiles were prefiltered by nominal *P* values of less than 0.05 and by an absolute methylation difference greater than 20%, calculated using

the MethylSig package (51); (b) greedy forward-feature selection was applied on the remaining tiles. Briefly, we assessed and prioritized the predictability of each of the filtered tiles in the SVM model and then sequentially evaluated the combinatorial predictability of the tiles by adding 1 tile from the prioritized tiles to the classifier at a time. The final predictors of the SVM classifier were selected from the set of tiles that could optimally predict patient response. The predictability was assessed on the basis of 10-fold cross-validation. Specifically, we randomly partitioned the 39 samples for which ERRBS libraries were available into 10 complementary subsets, training the SVM model on 9 of the 10 subsets (called the training set) and predicting the classes (responder or nonresponder) on the 1 left-out subset (called the validation set or testing set). To reduce variability, 10 rounds of cross-validation were performed using different partitions, and the validation results were summarized over the rounds. During each round of validation, the probability of each sample being predicted as a responder was recorded, and then the ROC-AUC across 10 rounds was calculated with the R package ROCR (90), and this calculation was used as the assessment of the predictability. Complete code is provided in the Supplemental Methods.

EpiTYPER MassARRAY

Validation of CpG methylation of select genomic regions was performed by MALDI-TOF using EpiTYPER MassARRAY (Sequenom) (49) on bisulfite-converted genomic DNA from a subset of DAC responders and nonresponders. The primers used to amplify these genomic regions and the resultant amplicon sequences are listed in Supplemental Table 6.

RNA-seq

RNA-seq was performed on RNA samples from 14 patients (8 responders and 6 nonresponders) who had high-quality RNA (RNA integrity number >6 as determined by the Agilent 2100 Bioanalyzer). RNA-seq libraries were prepared using the Illumina TruSeq RNA Sample Prep Kit (version 2) according to the manufacturer's instructions. A set of synthetic RNAs from the ERCC (91) at known concentrations were mixed with each of the cDNA libraries. Four separate samples were multiplexed into each lane and sequenced on a HiSeq 2000. The quality of reads obtained was evaluated using FastQC (Babraham Bioinformatics; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The sequenced libraries were aligned to the human genome (hg18) or to the ERCC spike-in reference sequence using TopHat, version 2.0.8 (92), with default parameters.

RNA-seq analysis

HTSeq (0.5.4p5) (93) was used to generate the count matrix with the following parameters: "htseq-count --mode=union --stranded=no" using the following 2 gene transfer format (GTF) annotation files, respectively: (a) the hg18 RefSeq gene GTF file downloaded from the UCSC genome browser for endogenous gene assembly; (b) the ERCC spike-in transcript GTF file downloaded from the official website (http://www.lifetechnologies.com/order/catalog/product/4456740) for ERCC spike-in assembly. The endogenous gene counts were normalized by ERCC spike-in library size, and the differential expression analysis was performed using the edgeR (version 3.4.2; Bioconductor) (94) generalized linear model (GLM). Genes with an absolute $\log_2$ fold change greater than 1 and a P value of less than 0.05 were reported.

qRT-PCR

To validate the RNA-seq results, RNA from selected nonresponder and responder patients was reverse transcribed using the Verso cDNA synthesis kit (Thermo Scientific) with random hexamer primers, according to the manufacturer's instructions. qRT-PCR was performed on the resulting cDNA in triplicate using intron-spanning and -flanking primer sets with Fast SYBR Green Master Mix and the StepOne Plus PCR System (Applied Biosystems) according to the manufacturer's instructions. The primer sequences are listed in Supplemental Table 7.

ELISAs

ELISAs for CXCL4 and CXCL7/NAP2 on serum from the CMML patients were performed using the corresponding ELISA kits (RAB0402 and RAB0135) from Sigma-Aldrich according to the manufacturer's directions. For CXCL4, the serum was diluted 1:500 in the sample dilution buffer provided in the kit.

IHC

For immunostaining, 3-μm-thick formalin-fixed, paraffin-embedded BM sections were deparaffinized in xylenes and hydrated in graded alcohols. Antigen retrieval was performed in EDTA (1 mM, pH 8.0) for two 15-minute cycles at maximum power in a microwave oven, and slides were then incubated with a CXCL4 antibody (1:300, catalog 500-P05; PeptroTech) or a CXCL7 antibody (1:50, catalog orb13423; Biorbyt). Immunostaining was performed with the BenchMark histostainer (Ventana Medical Systems, Roche) using a peroxidase detection kit with DAB substrate according to standard procedures. Sections were then counterstained with hematoxylin.

Cell culture and colony-forming assays

$CD34^+$ cells were isolated from cryopreserved BM MNCs from femoral head specimens using the CD34 MicroBead Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions. For CMML cells, the cryopreserved BM MNCs were rapidly thawed at 37°C and treated with DNAse to prevent cell clumping. Cells were plated in prestimulation media containing IMDM with 20% BIT (STEMCELL Technologies); IL-6 (20 ng/ml); SCF (100 ng/ml); TPO (100 ng/ml); and FLT3L (10 ng/ml) (PeproTech) and recovered overnight. The following day, either CXCL4 (50 ng/ml; PeproTech); CXCL7 (50 ng/ml; PeproTech); a combination of both chemokines (50 ng/ml each); or vehicle (PBS containing 0.1% BSA) was added as well as freshly prepared DAC (10 nM) (Sigma-Aldrich) or vehicle (water). DAC was replenished daily for a total of 3 days. Live cell numbers and viability were determined by trypan blue exclusion.

For colony assays, an equal number of live, treated $CD34^+$ cells were plated in duplicate in H4435 Enriched MethoCult (STEMCELL Technologies). Colonies were counted after 12 to 15 days.

Apoptosis assays

Apoptosis was assessed using the Tali Apoptosis Kit with annexin V Alexa Fluor 288 and propidium iodide according to the manufacturer's instructions and was measured on a Tali Image-Based Cytometer (all from Life Technologies).

Accession numbers

FISM cohort ERRBS and RNA-seq data are deposited in the NCBI's Gene Expression Omnibus (GEO) database (GEO GSE61163). GFM cohort ERRBS data are also deposited in the GEO database (GEO GSE63787).

Statistics

For the analysis of clinical parameters, Fisher's exact test was used for CMML type and sex; unpaired, 2-tailed Student's *t* tests were used for clinical parameters with a normal distribution; Wilcoxon signed-rank tests were used when the samples were not normally distributed; and the log-rank test was used for survival. A *P* value of less than 0.05 was considered significant. Somatic mutations between nonresponders and responders was evaluated using Fisher's exact test, and significance was considered at a *P* value of less than 0.05. For in vitro cell culture and colony-forming experiments, unpaired, 2-tailed Student's *t* tests were used for comparisons, and significance was considered at a *P* value of less than 0.05. For correlation analysis between the RNA-seq and qPCR results, Pearson's correlation was performed, and the *r* values and *P* values are indicated in the figures. The ERRBS and RNA-seq analyses were performed using a beta binomial test for differential methylation and a generalized linear model likelihood ratio test for differential gene expression. These methods were implemented through specific algorithms that are described in detail in their respective sections above.

Study approval

The current study was approved by the IRB of the University of Michigan Medical School and the ethics committee of the University of Florence, AOU Careggi-Firenze. The GFM clinical trial EudtraCT (2008-000470-21; https://eudract.ema.europa.eu/) was approved by the ethics committee of the Centre Hospitalier Universitaire de Dijon (Dijon, France). All samples were obtained from patients enrolled in clinical trials, and written, informed consent was obtained from these patients at the time of their enrollment in the study. Samples used in the current study were deidentified prior to use at the University of Michigan.

Acknowledgments

The current study was supported through a Translational Research Program Award from the Leukemia and Lymphoma Society (6016-14, to M.E. Figueroa). K. Meldi was partially supported by a Sass Foundation Postdoctoral Fellowship Award and an NIH T32 Training Grant (T32 CA 9676-22) through the University of Michigan Cancer Biology Program. O. Abdel-Wahab is supported by an NIH K08 Clinical Investigator Award (1K08CA160647-01), a US Department of Defense Postdoctoral Fellow Award in Bone Marrow Failure Research (W81XWH-12-1-0041), the Josie Robertson Investigator Program, and a Damon Runyon Clinical Investigator Award, with support from the Evans Foundation. E. Solary is supported by grants from the French National Cancer Institute and the Ligue Nationale Contre le Cancer. F. Buchi is supported by a Ministero dell'Istruzione dell'Universita e della Ricerca (MIUR)/University of Florence Postdoctoral Fellowship. E. Masala is supported by an FISM Postdoctoral Fellowship. We thank Gianna Baroni and Giulia Raugei for performing IHC.

Address correspondence to: Maria E. Figueroa, Department of Pathology, University of Michigan Medical School, 109 Zina Pitcher Pl., BSRB 2019, Ann Arbor, Michigan 48176, USA. Phone: 734.763.1865; E-mail: marfigue@med.umich.edu. Or to: Valeria Santini, University of Florence AOU Careggi, Hematology, Largo Brambilla 3, 50134 Firenze, Italy. Phone: 39.055.7947296; E-mail: santini@unifi.it.

1. Swerdlow SH, et al., eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 4th ed. Lyon, France: International Agency for Cancer (IARC); 2008.
2. Bennett JM, et al. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol*. 1982;51(2):189–199.
3. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. 2002;100(7):2292–2302.
4. Walter MJ, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 2011;25(7):1153–1158.
5. Tefferi A, et al. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN AML. *Leukemia*. 2009;23(7):1343–1345.
6. Abdel-Wahab O, et al. Genetic characterization of TET1, TET2, TET3 alterations in myeloid malignancies. *Blood*. 2009;114(1):144–147.
7. Gelsi-Boyer V, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes chronic myelomonocytic leukaemia. *Br J Haematol*. 2009;145(6):788–800.
8. Khan SN, et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia*. 2013;27(6):1301–1309.
9. Ernst T, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8):722–726.
10. Makishima H, et al. Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia*. 2010;24(10):1799–1804.
11. Nikoloski G, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*. 2010;42(8):665–667.
12. Figueroa ME, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood*. 2009;114(16):3448–3458.
13. Jiang Y, et al. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood*. 2009;113(6):1315–1325.
14. Kantarjian H, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer*. 2006;106(8):1794–1803.
15. Silverman LR, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 2002;20(10):2429–2440.
16. Ghoshal K, et al. 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, nuclear localization signal. *Mol Cell Biol*. 2005;25(11):4727–4741.
17. Patel K, Dickson J, Din S, Macleod K, Jodrell D, Ramsahoye B. Targeting of 5-aza-2′-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Res*. 2010;38(13):4313–4324.
18. Palii SS, Van Emburgh BO, Sankpal UT, Brown KD, Robertson KD. DNA methylation inhibitor 5-Aza-2′-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 3B. *Mol Cell Biol*. 2008;28(2):752–771.
19. Cihak A, Weiss JW, Pitot HC. Effects of 5-azacytidine on hepatic polyribosomes maturation of ribosomal RNA. *Acta Biol Med Ger*. 1974;33(5–6):859–865.
20. Griffiths EA, Gore SD. DNA methyltransferase and histone deacetylase inhibitors in the treatment of myelodysplastic syndromes. *Semin Hematol*. 2008;45(1):23–30.
21. Itzykson R, et al. Prognostic factors for response and overall survival in 282 patients with higher-risk myelodysplastic syndromes treated with azacitidine. *Blood*. 2011;117(2):403–411.
22. Gore SD, et al. Combined DNA methyltransferase and histone deacetylase inhibition in the treatment of myeloid neoplasms. *Cancer Res*. 2006;66(12):6361–6369.
23. Daskalakis M, et al. Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2′-deoxycytidine (decitabine) treatment. *Blood*. 2002;100(8):2957–2964.
24. Kantarjian H, et al. Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome chronic myelo-

monocytic leukemia. *Blood*. 2007;109(1):52–57.
25. Mund C, Hackanson B, Stresemann C, Lubbert M, Lyko F. Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Res*. 2005;65(16):7086–7090.
26. Blum W, et al. Phase I study of decitabine alone or in combination with valproic acid in acute myeloid leukemia. *J Clin Oncol*. 2007;25(25):3884–3891.
27. Shen L, et al. DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes. *J Clin Oncol*. 2010;28(4):605–613.
28. Follo MY, et al. Reduction of phosphoinositide-phospholipase C β1 methylation predicts the responsiveness to azacitidine in high-risk MDS. *Proc Natl Acad Sci U S A*. 2009;106(39):16811–16816.
29. Issa JP, et al. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2′-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood*. 2004;103(5):1635–1640.
30. Fandy TE, et al. Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies. *Blood*. 2009;114(13):2764–2773.
31. Bejar R, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011;364(26):2496–2506.
32. Jankowska AM, et al. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood*. 2011;118(14):3932–3941.
33. Kosmider O, et al. TET2 gene mutation is a frequent adverse event in chronic myelomonocytic leukemia. *Haematologica*. 2009;94(12):1676–1681.
34. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64–69.
35. Patnaik MM, et al. Spliceosome mutations involving SRSF2, SF3B1, U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, prognostic relevance. *Am J Hematol*. 2013;88(3):201–206.
36. Bejar R, et al. TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. *Blood*. 2014;124(17):2705–2712.
37. Traina F, et al. Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. *Leukemia*. 2014;28(1):78–87.
38. Itzykson R, et al. Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes low blast count acute myeloid leukemias. *Leukemia*. 2011;25(7):1147–1152.
39. Braun T, et al. Molecular predictors of response to decitabine in advanced chronic myelomonocytic leukemia: a phase 2 trial. *Blood*. 2011;118(14):3824–3831.
40. Cheson BD, et al. Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. *Blood*. 2006;108(2):419–425.
41. Itzykson R, et al. Prognostic score including gene mutations in chronic myelomonocytic leukemia. *J Clin Oncol*. 2013;31(19):2428–2436.
42. Patnaik MM, et al. Mayo prognostic model for WHO-defined chronic myelomonocytic leukemia: ASXL1 and spliceosome component mutations outcomes. *Leukemia*. 2013;27(7):1504–1510.
43. Meggendorfer M, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120(15):3080–3088.
44. Kohlmann A, et al. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, RUNX1. *J Clin Oncol*. 2010;28(24):3858–3865.
45. Akalin A, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet*. 2012;8(6):e1002781.
46. Figueroa ME, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553–567.
47. Figueroa ME, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest*. 2013;123(7):3099–3111.
48. Bullinger L, et al. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood*. 2010;115(3):636–642.
49. Ehrich M, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A*. 2005;102(44):15785–15790.
50. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012;489(7414):109–113.
51. Park Y, Figueroa ME, Rozek LS, Sartor MA. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics*. 2014;30(17):2414–2422.
52. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol*. 2013;14(3):R21.
53. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297.
54. Valencia A, et al. Expression of nucleoside-metabolizing enzymes in myelodysplastic syndromes and modulation of response to azacitidine. *Leukemia*. 2014;28(3):621–628.
55. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
56. Graham SM, Vass JK, Holyoake TL, Graham GJ. Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal chronic myeloid leukemia sources. *Stem Cells*. 2007;25(12):3111–3120.
57. Schaffner A, Rhyn P, Schoedon G, Schaer DJ. Regulated expression of platelet factor 4 in human monocytes — role of PARs as a quantitatively important monocyte activation pathway. *J Leukoc Biol*. 2005;78(1):202–209.
58. Pillai MM, Iwata M, Awaya N, Graf L, Torok-Storb B. Monocyte-derived CXCL7 peptides in the marrow microenvironment. *Blood*. 2006;107(9):3520–3526.
59. Bagger FO, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. *Nucleic Acids Res*. 2013;41(Database issue):D1034–D1039.
60. Bagger FO, et al. HemaExplorer: a web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis. *Blood*. 2012;119(26):6394–6395.
61. Aivado M, et al. Serum proteome profiling detects myelodysplastic syndromes and identifies CXC chemokine ligands 4 and 7 as markers for advanced disease. *Proc Natl Acad Sci U S A*. 2007;104(4):1307–1312.
62. Chen C, Bowen DT, Giagounidis AA, Schlegelberger B, Haase S, Wright EG. Identification of disease- and therapy-associated proteome changes in the sera of patients with myelodysplastic syndromes and del(5q). *Leukemia*. 2010;24(11):1875–1884.
63. Reikvam H, Fredly H, Kittang AO, Bruserud O. The possible diagnostic and prognostic use of systemic chemokine profiles in clinical medicine-the experience in acute myeloid leukemia from disease development and diagnosis via conventional chemotherapy to allogeneic stem cell transplantation. *Toxins (Basel)*. 2013;5(2):336–362.
64. Kittang AO, Hatfield K, Sand K, Reikvam H, Bruserud Ø. The chemokine network in acute myelogenous leukemia: molecular mechanisms involved in leukemogenesis and therapeutic implications. *Curr Top Microbiol Immunol*. 2010;341:149–172.
65. Han ZC, et al. Platelet factor 4 and other CXC chemokines support the survival of normal hematopoietic cells reduce the chemosensitivity of cells to cytotoxic agents. *Blood*. 1997;89(7):2328–2335.
66. Gupta SK, Singh JP. Inhibition of endothelial cell proliferation by platelet factor-4 involves a unique action on S phase progression. *J Cell Biol*. 1994;127(4):1121–1127.
67. Dudek AZ, Nesmelova I, Mayo K, Verfaillie CM, Pitchford S, Slungaard A. Platelet factor 4 promotes adhesion of hematopoietic progenitor cells binds IL-8: novel mechanisms for modulation of hematopoiesis. *Blood*. 2003;101(12):4687–4694.
68. Bruns I, et al. Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. *Nat Med*. 2014;20(11):1315–1320.
69. Tsai HC, et al. Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*. 2012;21(3):430–446.
70. Prebet T, et al. Outcome of high-risk myelodysplastic syndrome after azacitidine treatment failure. *J Clin Oncol*. 2011;29(24):3322–3327.
71. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306–1311.
72. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 2006;38(11):1348–1354.
73. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- interchromosomal interactions. *Nat Genet*. 2006;38(11):1341–1347.
74. Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic

elements. *Genome Res*. 2006;16(10):1299–1309.
75. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–293.
76. Wong YF, Jakt LM, Nishikawa S. Prolonged treatment with DNMT inhibitors induces distinct effects in promoters and gene-bodies. *PLoS One*. 2013;8(8):e71099.
77. Chung E, Hsu CL, Kondo M. Constitutive MAP kinase activation in hematopoietic stem cells induces a myeloproliferative disorder. *PLoS One*. 2011;6(12):e28350.
78. Katsoulidis E, et al. Role of the p38 mitogen-activated protein kinase pathway in cytokine-mediated hematopoietic suppression in myelodysplastic syndromes. *Cancer Res*. 2005;65(19):9029–9037.
79. Abrams SL, et al. The Raf/MEK/ERK pathway can govern drug resistance, apoptosis and sensitivity to targeted therapy. *Cell Cycle*. 2010;9(9):1781–1791.
80. Steelman LS, et al. Dominant roles of the Raf/MEK/ERK pathway in cell cycle progression, prevention of apoptosis and sensitivity to chemotherapeutic drugs. *Cell Cycle*. 2010;9(8):1629–1638.
81. Zhang L, et al. BRAF kinase domain mutations are present in a subset of chronic myelomonocytic leukemia with wild-type RAS. *Am J Hematol*. 2014;89(5):499–504.
82. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
84. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–498.
85. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
86. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27(11):1571–1572.
87. RC Team. *R: A Language And Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
88. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
89. The Comprehensive R Archive Network. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. TCRAN Web site. http://cran.r-project.org/package=e1071. Accessed February 20, 2015.
90. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–3941.
91. Jiang L, et al. Synthetic spike-in standards for RNA-seq experiments. *Genome Res*. 2011;21(9):1543–1551.
92. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31(1):46–53.
93. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–169.
94. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.