



HSC commitment–associated epigenetic signature is prognostic in acute myeloid leukemia

Boris Bartholdy,¹ Maximilian Christopeit,¹ Britta Will,¹ Yongkai Mo,² Laura Barreyro,¹ Yiting Yu,² Tushar D. Bhagat,² Ujunwa C. Okoye-Okafor,¹ Tihomira I. Todorova,¹ John M. Greally,³ Ross L. Levine,⁴ Ari Melnick,⁵ Amit Verma,^{2,6,7,8} and Ulrich Steidl^{1,6,7,8}

¹Department of Cell Biology, ²Department of Developmental and Molecular Biology, and

³Center for Epigenomics and Division of Computational Genetics, Department of Genetics, Albert Einstein College of Medicine, New York, New York, USA.

⁴Human Oncology and Pathogenesis Program and Leukemia Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. ⁵Division of Hematology/Oncology, Department of Medicine, Weill Cornell Medical College, New York, New York, USA.

⁶Division of Hematologic Malignancies, Department of Medicine (Oncology), Einstein/Montefiore Medical Center, New York, New York, USA.

⁷Albert Einstein Cancer Center and ⁸Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, New York, New York, USA.

Acute myeloid leukemia (AML) is characterized by disruption of HSC and progenitor cell differentiation. Frequently, AML is associated with mutations in genes encoding epigenetic modifiers. We hypothesized that analysis of alterations in DNA methylation patterns during healthy HSC commitment and differentiation would yield epigenetic signatures that could be used to identify stage-specific prognostic subgroups of AML. We performed a nano HpaII-tiny-fragment-enrichment-by-ligation-mediated-PCR (nanoHELP) assay to compare genome-wide cytosine methylation profiles between highly purified human long-term HSC, short-term HSC, common myeloid progenitors, and megakaryocyte-erythrocyte progenitors. We observed that the most striking epigenetic changes occurred during the commitment of short-term HSC to common myeloid progenitors and these alterations were predominantly characterized by loss of methylation. We developed a metric of the HSC commitment–associated methylation pattern that proved to be highly prognostic of overall survival in 3 independent large AML patient cohorts, regardless of patient treatment and epigenetic mutations. Application of the epigenetic signature metric for AML prognosis was superior to evaluation of commitment-based gene expression signatures. Together, our data define a stem cell commitment–associated methylome that is independently prognostic of poorer overall survival in AML.

Introduction

In the pathogenesis of acute myeloid leukemia (AML), genes encoding epigenetic modifiers are frequently mutated (1, 2). Some of these mutations have been attributed prognostic value in AML (3). Additionally, aberrant DNA cytosine methylation in AML blasts has led to the identification of AML subtypes, independent of features usually associated with AML (4). Differentiation of murine HSC to progenitor cells is associated with distinct changes in DNA cytosine methylation (5–7). In turn, targeted disruption of DNA cytosine methylation patterns disturbs regulation of differentiation of murine hematopoietic stem and progenitor cells (HSPC) and affects HSPC function (8–10). This suggests that methylation plays an active role in the differentiation program. In the murine hematopoietic system, dynamic changes of DNA methylation have been described during multipotent progenitor cell differentiation (5) and HSC commitment (7), with pronounced demethylation in erythroid progenitors during differentiation (6, 7). Severely perturbed hematopoiesis (8–11) and myeloid transformation (12–14) are common hallmarks of mouse models with targeted disruptions in a growing number of

enzymes known to contribute to the homeostasis of DNA cytosine methylation. Little is known about changes in DNA cytosine methylation during early human hematopoiesis. Identification of stage- and locus-specific epigenetic mechanisms of leukemic transformation would require a detailed genome-wide map of DNA cytosine methylation patterns and dynamics during the stepwise maturation of HSC.

Here, we delineate DNA cytosine methylation profiles of highly purified long-term HSC (LTHSC), short-term HSC (STHSC), common myeloid progenitors (CMP), and megakaryocyte-erythrocyte progenitors (MEP) from the BM of healthy human individuals. We show that an epigenetic stem cell commitment–specific signature is highly prognostic for overall survival (OS) of patients with AML. The epigenetic signature is enriched for binding sites of known hematopoietic transcription factors and microRNA loci.

Results

Most DNA cytosines are methylated in human HSPC. To characterize DNA cytosine methylation in early human hematopoiesis, we studied the distribution of and changes in methylation during *in vivo* physiologic differentiation from LTHSC, immunophenotypically characterized as lineage depleted (Lin[−], CD34⁺, CD38[−], CD90⁺, to STHSC (Lin[−], CD34⁺, CD38[−], CD90[−]), to CMP (Lin[−], CD34⁺, CD38⁺, CD123⁺, CD45RA[−]) to MEP (Lin[−], CD34⁺, CD38[−], CD123[−], CD45RA[−]) (15–21). We took advantage of a novel method

Authorship note: Boris Bartholdy and Maximilian Christopeit, as well as Amit Verma and Ulrich Steidl, contributed equally to this work.

Conflict of interest: The authors have declared that no conflict of interest exists.

Citation for this article: *J Clin Invest.* 2014;124(3):1158–1167. doi:10.1172/JCI1264.



combining 8-parameter high-speed FACS of primary human BM cells with an optimized nano HpaII-tiny-fragment-enrichment-by-ligation-mediated-PCR (nanoHELP) (Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI71264DS1) assay (22–26). This approach permitted examination of single individuals' stem cells isolated as biological replicates, i.e., without pooling of samples prior to analysis. We were thus able to analyze DNA cytosine methylation in rare, highly purified human HSPC populations (Supplemental Table 1). Globally, we found that the majority of DNA cytosines in human LTHSC, STHSC, CMP, and MEP (76%–81% of loci) from healthy individuals were methylated (Supplemental Figure 2A and Supplemental Table 2). We quantitatively compared methylation across all loci between the stages of differentiation. Interestingly, we found a highly significant reduction in median DNA cytosine methylation specifically at the stem cell-to-progenitor (STHSC-to-CMP) transition ($P < 2.2 \times 10^{-16}$, Mann-Whitney test; Supplemental Figure 2B).

Dynamic changes in DNA cytosine methylation during HSC commitment. To characterize the dynamics of cytosine methylation during HSC commitment, we investigated changes in the methylation status at the level of individual loci and compared methylation in LTHSC to methylation in STHSC, STHSC to CMP, and CMP to MEP. The comparison between LTHSC and STHSC showed 509 significantly differentially methylated loci ($P < 0.05$). Demethylation was observed in 40% (205/509) of these loci during transition from the LTHSC to the STHSC compartment, whereas 60% (304/509) were more methylated in STHSC compared with LTHSC. At the transition from STHSC to CMP, the step of definitive commitment of HSC, we observed a total of 793 differentially methylated loci. However, in stark contrast to the nearly balanced hypo- and hypermethylation of loci between LTHSC and STHSC, 95% (757/793) of differentially methylated loci in STHSC were more methylated than in CMP, whereas only 5% (36/793) were less methylated. The transition from CMP to MEP was accompanied by balanced hypo- and hypermethylation, with 127 (52%) loci showing higher and 116 (48%) loci showing lower methylation in the CMP compartment (Figure 1A and Supplemental Table 3). Changes occurred without apparent focus throughout the genome (Figure 1B). Our findings show that in human HSC, demethylation particularly occurs at the commitment step from STHSC to CMP.

A stem cell commitment-associated epigenetic signature distinguishes human HSC and progenitor cell subsets. To identify loci with most significant methylation changes across the assessed differentiation stages, we performed significance analysis for microarrays (SAM) on loci that showed differentiation-specific methylation changes independently of the variations between biological replicates, in analogy to a recently published approach (7). This resulted in a set of 561 loci that distinguished between the 4 investigated stages of human HSPC development (Figure 1C). The most prominent distinction was observed at the transition from stem cells (STHSC) to progenitor cells (CMP), consistent with the analysis of changes in DNA cytosine methylation during stem cell commitment described in Figure 1A. The signature mainly consisted of loci that were significantly demethylated during commitment from STHSC to CMP (516/561 loci, 92.0%). Interestingly, this stem cell commitment-associated epigenetic signature was enriched in loci associated with several genes that are commonly implicated not only in human HSC function and commitment but also in leukemogenesis, such as *CEBPA* (27–29), *E2F1* (30), *KRAS* (31, 32), and *WEE1* (33)

as well as a noncoding transcript, *MIRLET7B* (ref. 34 and Supplemental Table 4). Given emerging evidence that microRNAs play an essential role in both normal hematopoiesis and leukemogenesis (35–38) we assessed additional microRNA transcripts in the vicinity of the methylation probes on the array. Using miRBase, we indeed found a number of microRNAs that were associated with significant epigenetic changes (Supplemental Table 4). Ingenuity pathway analysis using the significant constituents of this methylation signature revealed an enrichment of genes involved in the function and disease characteristics “Differentiation of blood cells” ($P = 8.39 \times 10^{-43}$), “Lymphohematopoietic cancer” ($P = 3.2 \times 10^{-12}$), specifically “AML” ($P = 2.47 \times 10^{-6}$), and “Cell transformation” ($P = 1.41 \times 10^{-12}$). The top 5 canonical pathways were “AML signaling,” “Molecular mechanisms of cancer,” “Glioblastoma multiforme signaling,” “Pancreatic adenocarcinoma signaling,” and “Glucocorticoid receptor signaling” (Figure 1D and Supplemental Table 5). Taken together, marked changes in DNA cytosine methylation during human HSC commitment occur at genomic loci involved in hematopoietic differentiation and in hematological malignancies.

Stem cell commitment-associated epigenetic signature is prognostic for OS in AML. Pathway analysis of the epigenetic signature showed an enrichment of genes implicated in systemic disorders of hematopoietic development. We therefore sought to determine whether the methylation status of this set of 561 stem cell commitment-associated loci derived from healthy human HSPC was affected in AML, a disease associated with epigenetic dysregulation in HSPC (1). We developed a signature score based on the methylation of the 561 loci defined by the stem cell commitment-associated epigenetic signature. Additionally, we analyzed data from clinical trials of patients with AML. We identified 3 published independent cohorts of patients for which DNA methylation data and gene expression data as well as data on OS and mutational characteristics were available (4, 39–42). To assess the prognostic impact of this epigenetic signature, we associated OS of patients with their score. We first tested this approach on 1 cohort from a prospective randomized clinical trial that compared 2 different doses of daunorubicin (41). In the cohort receiving the standard, lower-dose daunorubicin, a low stem cell commitment-associated epigenetic signature score was associated with increased OS (hazard ratio [HR] = 1.537, 95% CI = 1.086–2.245, $P = 0.0165$, log-rank test; Figure 2, A and B). Patients in the group with lower epigenetic signature scores showed a median OS of 19.0 months, compared with 10.8 months in the group with higher epigenetic scores. Next, we applied the stem cell commitment-associated epigenetic signature score to the group of patients that received a higher dose of daunorubicin (41). The association of the stem cell commitment-associated epigenetic signature score with OS was also observed in this cohort of patients (HR = 1.691, 95% CI = 1.169–2.552, $P = 0.0062$; Figure 2, C and D). Median OS in the group with low epigenetic signature score was 25.4 months, compared with 13.2 months in the high-scoring group. Of note, the significant association of high epigenetic signature score with poor outcome persisted upon combination of the 2 treatment arms of this trial (Figure 2E, median OS 11.1 months for patients with a high versus 22.8 months for patients with a low score, HR = 1.609, 95% CI = 1.258–2.143, $P = 0.0003$). To independently assess the association of the loci from the stem cell commitment-associated epigenetic signature with clinical outcome, we performed *globaltest* analysis (43) using these loci as covariates. This confirmed the significant association of the 561-loci classifier with OS ($P = 0.000697$). In a

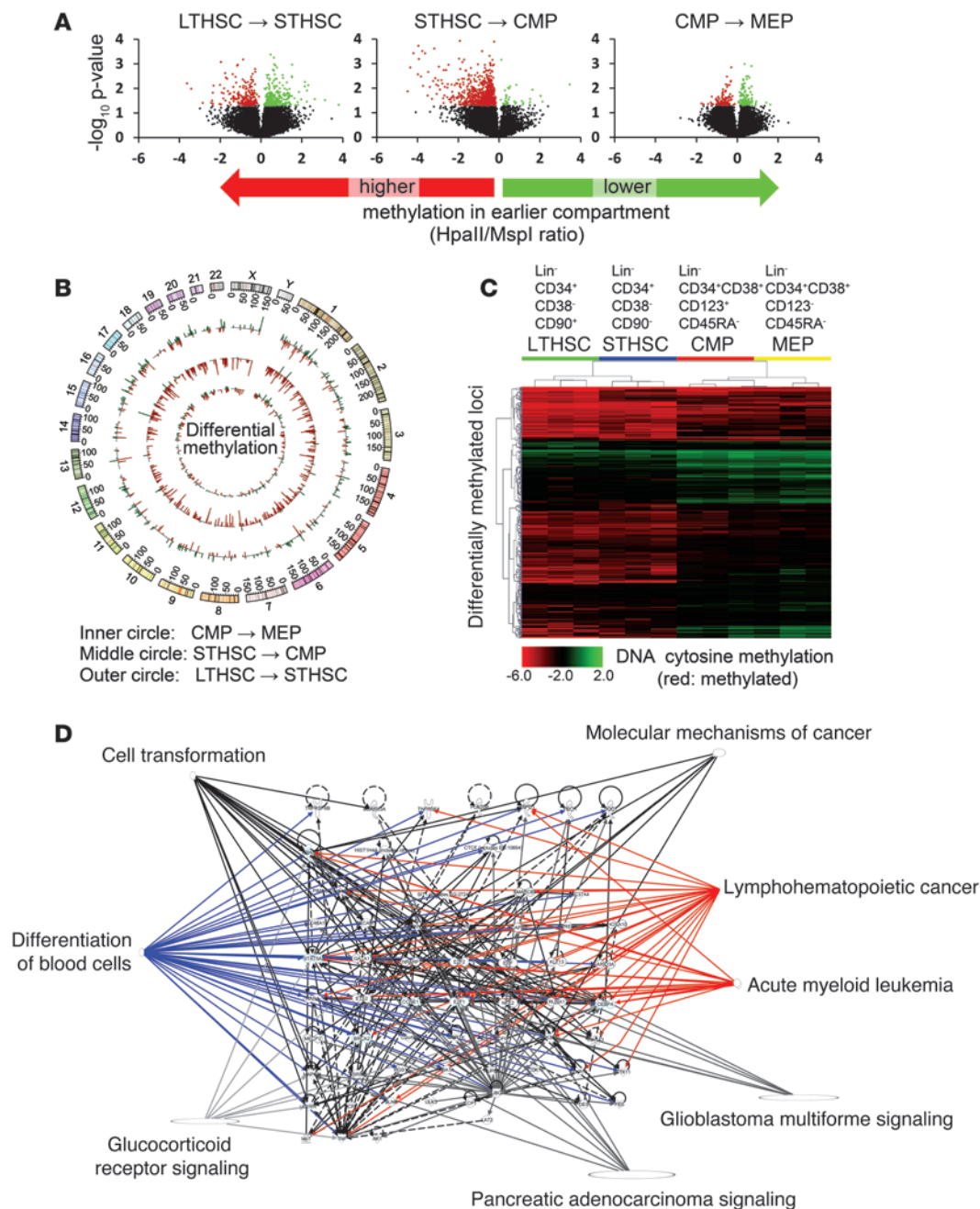


Figure 1

Hypomethylation during HSC commitment to hematopoietic progenitors. **(A)** Genome-wide changes in DNA methylation during HSC commitment. Red dots represent loci with significantly lower methylation at the developmentally later stage, i.e., loci demethylated during the respective transition ($P < 0.05$, t test). **(B)** Significant changes in DNA cytosine methylation at the transition from LTHSC to STHSC (outer circle), STHSC to CMP (middle circle), and CMP to MEP (inner circle) are plotted in relation to the genomic position. Chromosomes are plotted along the ideogram. Red bars denote significantly demethylated loci; green bars denote significant increase in methylation at the respective loci. **(C)** SAM was used to define an epigenetic signature based on loci that undergo the most significant methylation changes during HSPC differentiation. The epigenetic signature (561 loci) distinguishes HSPC subsets in hierarchical clustering analysis. Log₂-transformed HpaII/MspI ratios (color code next to the heat map) of all 561 loci are shown for the 4 analyzed differentiation stages (indicated above the heat map) from 3 healthy human individuals. Trees result from Euclidean clustering of this signature. Associated genes are listed in Supplemental Table 4. **(D)** Ingenuity pathway analysis highlights functional implications of gene enrichment analysis of the epigenetic stem cell commitment-associated signature. 62 genes significantly associated ($P < 0.05$ after Benjamini-Hochberg correction) with Ingenuity “Top bio functions” were entered into pathway generation. Top 5 “Canonical pathways” (“AML signaling,” “Molecular mechanisms of cancer,” “Glioblastoma multiforme signaling,” “Pancreatic adenocarcinoma signaling,” “Glucocorticoid receptor signaling”) and the top 3 characteristics of “Function and disease” (“Differentiation of blood cells” [$P = 8.39 \times 10^{-43}$], “Lymphohematopoietic cancer” [$P = 3.2 \times 10^{-12}$], “AML” [$P = 2.47 \times 10^{-6}$], and “Cell transformation” [$P = 1.41 \times 10^{-12}$]) are depicted.

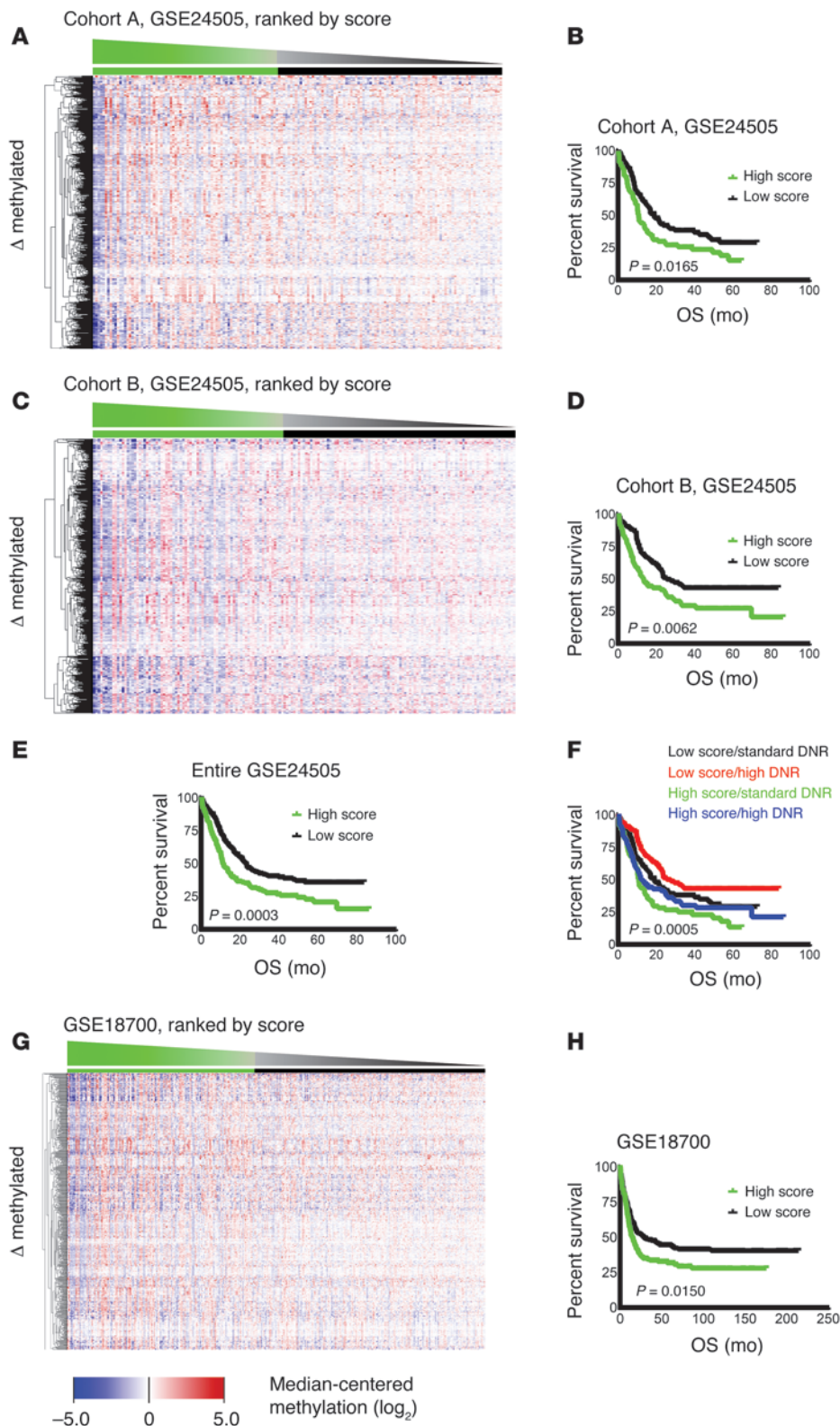
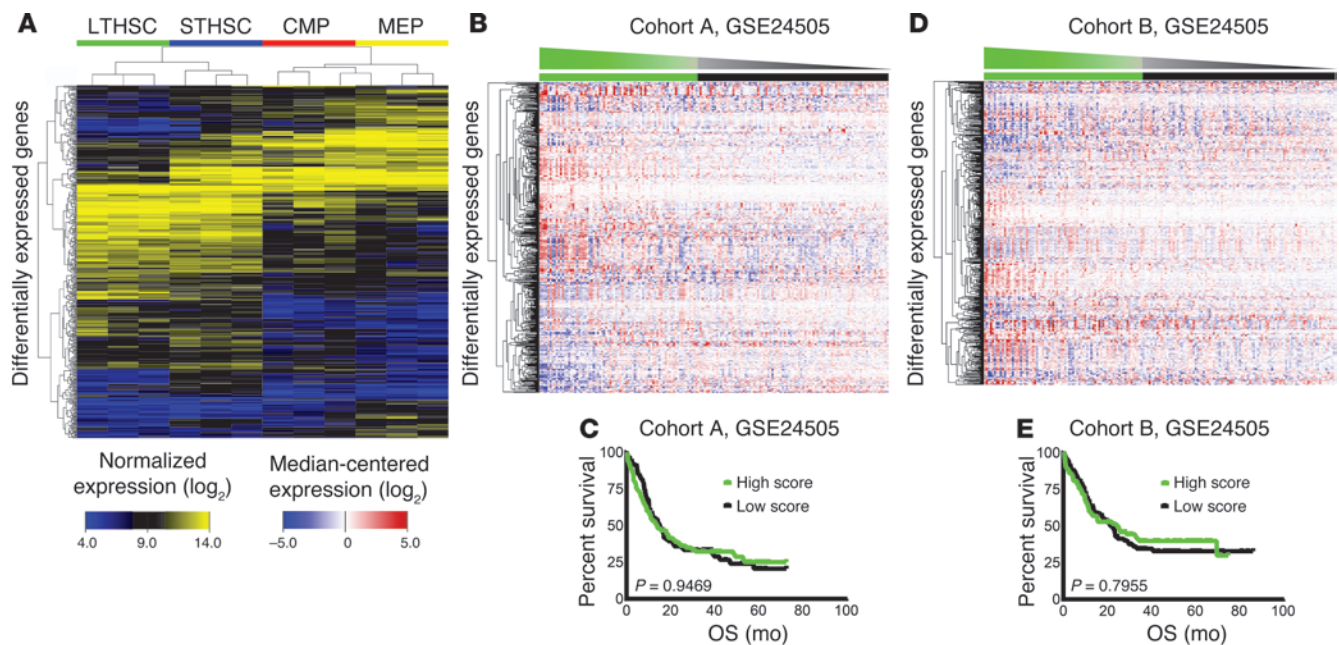


Figure 2

Stem cell commitment-associated epigenetic signature is prognostic in AML. Application of the epigenetic signature to 3 independent published sets of patients with AML (4, 39–42). (A and B) Analysis of patients with AML who received standard chemotherapy. (C and D) Analysis of patients with AML who received chemotherapy with a higher dose of daunorubicin (DNR). (E) Analysis of the combined cohort of AML patients receiving standard or higher doses of daunorubicin (41). (G and H) Analysis of a third independent cohort of AML patients (39, 40). (A, C, and G) Heat maps of the respective patients (horizontal order) and the 561 loci (vertical order). Patients are ranked in descending order based on the signature score. Patients with high signature score are indicated by a green bar; patients with a low signature score by a black bar above the median-centered methylation heat map. (B, D, and E) Kaplan-Meier survival curves of OS of patients with AML are plotted. Green solid lines represent OS of patients with a high signature score; black solid lines represent OS of patients with a low signature score. (F) Overlay of survival curves from B and D. Black/red lines: patients with a low epigenetic stem cell commitment-associated signature score receiving standard or high dose daunorubicin treatment. Green/blue lines: patients with a high epigenetic stem cell commitment-associated signature score receiving standard or high dose daunorubicin treatment.

multivariate Cox proportional hazard regression analysis (44) that included the epigenetic score in addition to the well-established factors cytogenetic and molecular risk stratification (3) and age, the epigenetic score remained independently and significantly associat-

ed with OS (Table 1). As depicted in the overlay of the survival curves from Figure 2, B and D, patients with a low epigenetic signature score receiving high levels of daunorubicin had a significantly better OS than patients from the other groups (Figure 2F, $P = 0.0005$).

**Figure 3**

Lower prognostic power of gene expression signature. (A) Generation of a gene expression signature based on 455 gene expression probes that undergo significant changes between the 4 measured differentiation stages using SAM. Heat map of \log_2 -transformed expression values of this signature is shown. Genomic coordinates and gene names are listed in Supplemental Table 6. (B–E) Application of the stem cell commitment-associated gene expression signature to patients with AML. (B and D) Heat maps of median-centered expression of the probes corresponding to the commitment-associated gene expression signature in patients with AML who received standard chemotherapy (B), or chemotherapy with a higher dose of daunorubicin (D), ranked by score. (C and E) Kaplan-Meier curves of OS of patients with AML are plotted. Green lines represent OS of patients with a high expression signature score; black lines represent OS of patients with a low expression signature score.

Additionally, we validated the power of the epigenetic score in a third, independent cohort of patients with AML. For this, we analyzed published clinical and methylation data from patients from 4 clinical trials included in a study from the Dutch-Belgian Cooperative Trial Group for Hematology Oncology (HOVON) group (4, 39, 40). In this study, patients with a low epigenetic score again had a significantly better OS than those with a high score (median survival 28.1 months versus 14.9 months; HR = 1.390, 95% CI = 1.069–1.838, $P = 0.0150$) (Figure 2, G and H). *Globaltest* analysis (43) of this cohort independently confirmed significant association of the signature score with survival ($P = 0.000335$).

Taken together, the methylation status of the commitment-associated loci identified in human HSPC from healthy individuals showed independent prognostic power in human AML in a total of 688 patients.

Low correlation of commitment-associated gene expression signature with AML patient outcome. Previous studies have defined gene expression signatures prognostic for OS of patients with AML (45–47). Therefore, we sought to determine whether a gene expression signature constructed in analogy to the epigenetic signature had comparable prognostic potential in the AML cohorts studied. We first determined differentiation-specific gene expression changes independent of the variation between biological replicates by SAM. Expression of the identified transcripts distinguished between the 4 investigated stages of human HSPC development (Figure 3A). We repeated the approach chosen to associate the epigenetic signature with OS and applied this gene expression signature to the AML patient cohorts. The signature consisted of 530 genes that

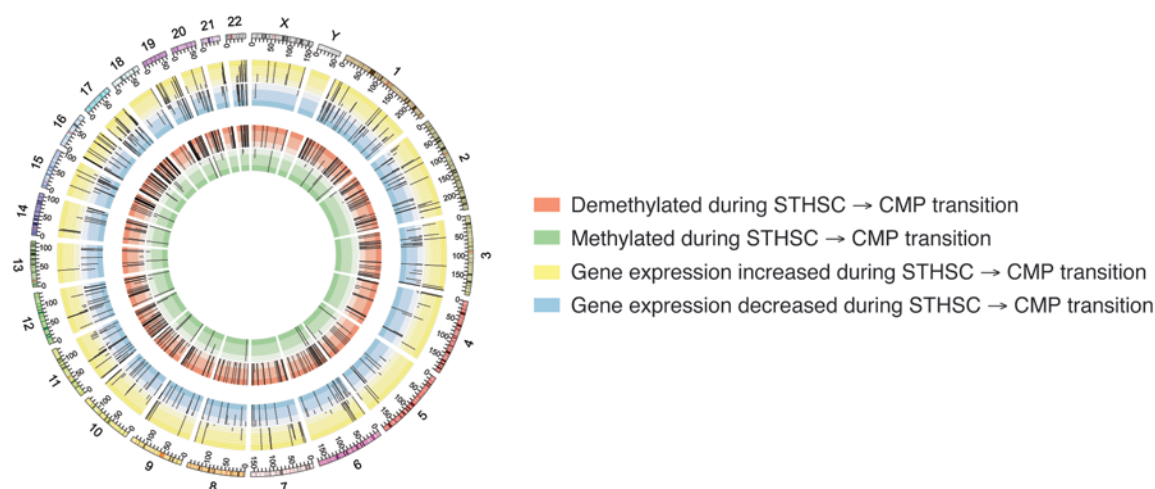
were differentially expressed in the analyzed stem and progenitor cells from healthy human individuals (Figure 3, B and D, and Supplemental Table 6). No significant correlation of the stem cell commitment-associated gene expression signature with OS was observed in either AML treatment group (Figure 3, C and E). Association of gene expression signatures with outcome using the R/Bioconductor *globaltest* package as an alternative algorithm revealed a significant association of these genes with OS only in the combined Eastern Cooperative Oncology Group (ECOG) cohort ($P = 0.00168$) but not in the HOVON cohort ($P = 0.363$). While a published HSC gene expression signature (46) was asso-

Table 1

Multivariate analysis using the epigenetic stem cell commitment signature score, cytogenetic and molecular risk stratification, age, and treatment of the patients from GSE24505 as covariates

	HR (95 %CI)	P value
Epigenetic score (high = 1)	0.6856 (0.5247–0.8957)	0.0056
Intermediate risk	2.0328 (1.3415–3.0803)	0.0008
Unfavorable risk	4.2794 (2.8662–6.3893)	1.17×10^{-12}
Age ($\geq 46 = 1$)	0.6905 (0.5267–0.9052)	0.0073
Treatment arm (A = 1)	0.7682 (0.5889–1.0021)	0.0518

Score, risk and treatment were considered as categorical values. Survival analysis was performed using a Cox proportional hazards model.

**Figure 4**

Correlation of epigenetic signature's constituents with expression of closest mappable gene product. Changes of the 561 constituents of the epigenetic signature during transition from STHSC to CMP are aligned with significant changes in gene expression at mappable loci nearby. Red represents demethylated loci, green methylated loci during STHSC to CMP transition; yellow represents increased and blue decreased gene expression at associated loci.

ciated with OS in the ECOG cohort ($P = 0.00202$, *globaltest*), the association of a leukemia stem cell gene expression signature (46) with OS missed significance ($P = 0.0821$, *globaltest*) as did an additional leukemic stem cell gene expression signature (45) ($P = 0.257$, *globaltest*). These findings suggest that the stem cell commitment-associated epigenetic signature is a more robust indicator of OS than a stem cell commitment-associated gene expression signature obtained in an identical, unbiased fashion.

Correlation of methylation and gene expression changes between stages of human HSC commitment. DNA cytosine methylation has been associated with regulation of transcription. Promoters of developmental genes as well as promoters of housekeeping genes can be silenced by hypermethylation (48), while gene bodies have been reported to be methylated following increased transcription of the respective gene (49). We correlated methylation and gene expression during the respective HSPC transitions. Besides locus-specific inverse correlation between decreasing methylation and increasing gene expression (Figure 4 and Figure 5A, upper right quadrants) and increasing methylation and decreasing gene expression (Figure 5A, lower left quadrants), we also found loci with a positive correlation between decreasing methylation and decreasing gene expression (Figure 5A, upper left quadrants) and increasing methylation and increasing gene expression (Figure 5A, lower right quadrants). A significant correlation between decrease of cytosine methylation and increase in gene expression at the STHSC-to-CMP transition appeared when correlating the commitment-associated gene expression signature with nearby CpG loci (Figure 5B). Changes in methylation at an earlier transition did not significantly associate with changes in gene expression at a later transition (e.g., methylation during transition from LTHSC to STHSC compared with gene expression during transition from STHSC to CMP and data not shown). Taken together, the epigenetic signature is not universally correlated with gene expression, although there are certain loci that show correlation or inverse correlation. Yet, at the STHSC-to-CMP transition, an inverse correla-

tion between gene expression and associated methylation changes can be observed. Changes in expression of the genes associated with the epigenetic stem cell commitment-associated signature were not prognostic for outcome in AML patients ($P = 0.133$, ECOG cohort, *globaltest*). Of note, mutations of genes known to directly affect DNA methylation, such as *IDH1*, *IDH2*, *TET2*, and *DNMT3A*, were not enriched in either the high- or low-scoring group (Supplemental Table 7). Finally, we investigated whether specific DNA motifs were enriched around the constituents of the epigenetic signature, which could provide mechanistic insights into the regulation of these loci. Using HOMER transcription binding-site analysis (50), we observed a significant enrichment of consensus binding sites for several essential transcription factors involved in hematopoietic differentiation (most notably GATA transcription factors, Maf family members, KLF4, and Smad2; refs. 51–53) in the epigenetic signature (Supplemental Table 8).

Discussion

Perturbed epigenetic regulation of differentiation from HSC to mature blood cells can result in a block in cellular differentiation, clinically apparent in hematopoietic malignancies such as AML (1). To study epigenetic regulation during earliest human hematopoiesis, we analyzed the status of and changes in DNA cytosine methylation during in vivo differentiation of human HSC. To this end, we developed a technique that enabled us to characterize DNA cytosine methylation from prospectively isolated highly enriched human HSC from single individuals in small numbers. We coupled prospective isolation of human HSPC with a modified HELP assay, the so-called nanoHELP (22–26). We found that most DNA cytosines in human LTHSC, STHSC, CMP, and MEP were methylated, in agreement with findings in other vertebrate somatic stem cells and differentiated tissues (5–7, 54). Our findings show that, while mean methylation levels are comparable to those found in murine HSC (7), in human HSC, demethylation particularly occurs at the commit-

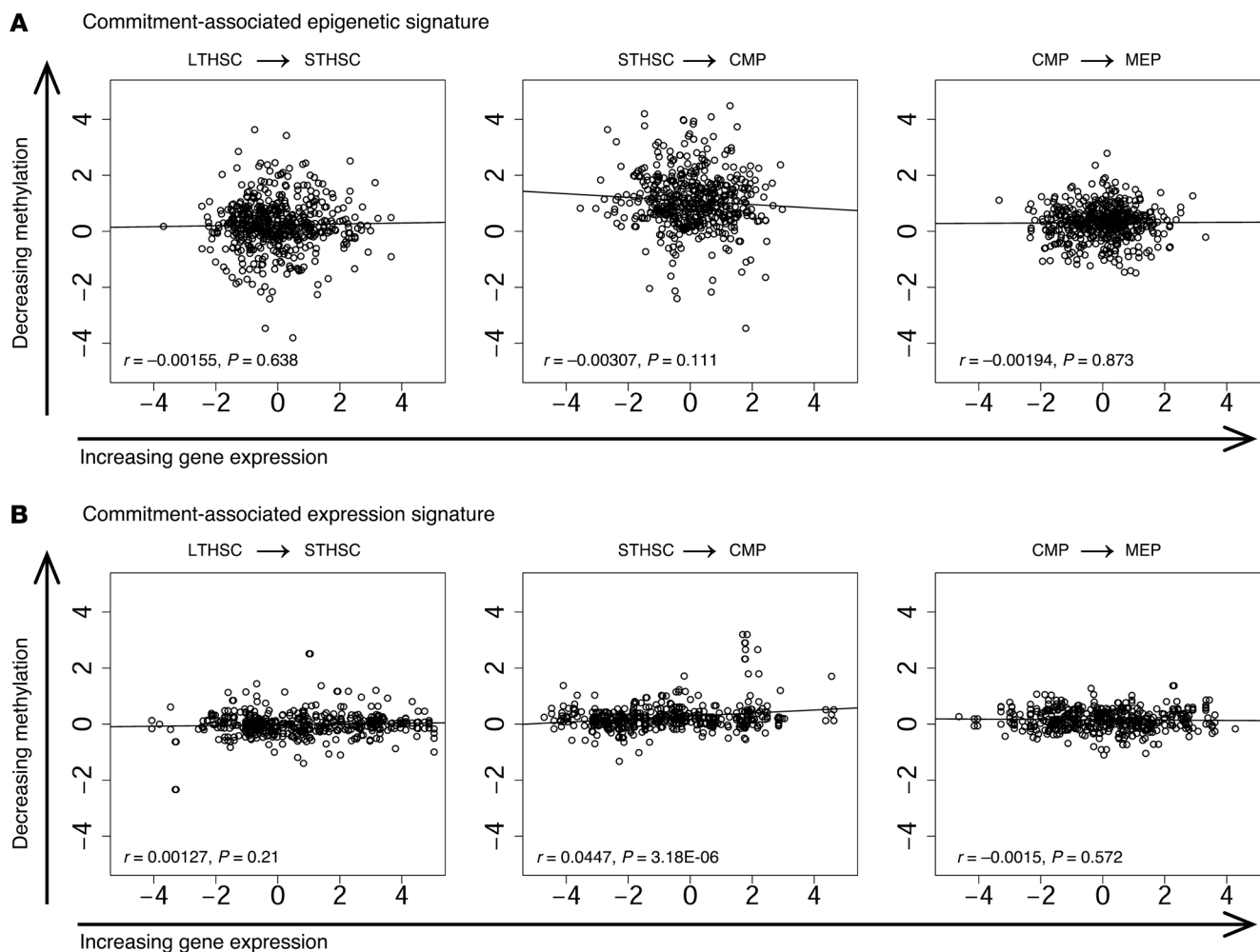


Figure 5

Correlation between methylation and expression of genes in the commitment-associated signatures. Changes in DNA cytosine methylation and gene expression between the indicated differentiation stages were calculated (mean methylation in later vs. earlier compartment, and mean expression in later vs. earlier compartment) and plotted in the graphs. The lines represent the linear model of these data points, and the P value for correlation is indicated. **(A)** Using the stem cell commitment-associated 561-probe epigenetic signature, 530 pairs of methylation probe with adjacent transcript were mapped and their correlation is shown. **(B)** Similarly, the commitment-associated 455 probe expression signature was used to derive 283 pairs of transcripts associated with nearby or overlapping methylation probes, which are plotted in the figure.

ment step from STHSC to CMP (Figure 1A and Supplemental Table 2). This has not, to our knowledge, been described so far. Furthermore, our data define specific loci with dynamic changes in methylation during human HSPC differentiation. These loci represent a stem cell commitment-associated epigenetic signature that clusters the subsequent stages of HSC differentiation (Figure 1C) and is enriched in genes associated with hematopoietic development and also leukemogenesis, particularly AML (Figure 1D). Therefore, we assessed whether the methylation status at these loci would have clinical implications in human AML. Indeed, we found that this signature was able to classify 3 independent cohorts of patients with AML from prospective clinical trials into groups with superior or significantly inferior OS. Patients treated with standard chemotherapy with a low stem cell commitment-associated epigenetic signature score reached significantly longer OS than patients with a high score. We assessed the power of this score using data from a second

cohort of AML patients treated with an experimental approach (41) and found an even stronger distinction between the groups (Figure 2). This is in contrast to some currently used mutational markers (3) and suggests a high degree of robustness of the prognostic value of the stem cell commitment-associated epigenetic signature. Multivariate analysis demonstrated an independent association of the epigenetic score with OS, and no enrichment of mutations of known modifiers of DNA methylation was detected in either the high- or low-scoring group (Supplemental Table 7). The overlay of survival curves from the different clinical cohorts (Figure 2F) suggests that the epigenetic signature might serve as a predictor for OS particularly in AML patients receiving higher doses of daunorubicin. It may be of interest to prospectively test whether the superiority of higher dose daunorubicin over standard dose daunorubicin is restricted to patients with a low epigenetic signature score. A third independent cohort of patients with AML studied by the HOVON group (39, 40) also



segregated into better and worse prognosis on the basis of the epigenetic stem cell commitment–associated score, further demonstrating the robustness and prognostic potential of this score. Taken together, the epigenetic stem cell commitment signature was validated in 3 independent cohorts of AML patients, with a total of 688 patients. Of note, in each of these cohorts, median survival was approximately doubled in patients with low signature score, even in the cohort that was treated with higher dose daunorubicin, indicating the robustness of the prognostic value of this signature. Similarly derived gene expression signatures were not able to achieve the robustness that we observed using the epigenetic signature.

Recent studies have linked changes in methylation to the regulation of microRNAs, and we identified 1 microRNA transcript, MIRLET7, in the signature; in addition, several other microRNA genes were located adjacent to the differentially methylated region (DMR) (Supplemental Table 4). It will be interesting to examine the functional association between regulation of methylation and activation of these microRNAs in future studies.

Sequence analysis of the DMR regions revealed a significant enrichment of motifs for transcription factors that were previously shown to be implicated in hematopoietic differentiation and leukemogenesis, such as GATA factors, MAFF, and KLF4. For instance, it was recently shown that erythroid differentiation is accompanied by functional demethylation of essential erythropoietic genes, including *GATA1* (6, 55). In addition, maintenance of HSC programs and prevention of activation of differentiation programs are controlled by DNA methylation (8). It will be of interest in future studies to investigate whether any of the master regulatory transcription factors have an active role in the modulation of DNA methylation at their target loci, aside from being subject to regulation by promoter methylation.

Our analyses were performed on DNA from highly enriched HSPC, thus avoiding the measurement of DNA cytosine methylation and gene expression from heterogeneous cell populations. In addition, analyzing cells from single donors, as opposed to pooling cells from multiple donors, allowed us to derive changes propagated through various differentiation stages in individuals in addition to changes that occurred in a stage-specific manner across all individuals studied. Furthermore, we were able to access an exhaustive high quality data set that included both data on DNA cytosine methylation in leukemic blasts and clinical data, including a detailed description of risk groups and OS from a prospective randomized clinical trial (41). These data have been the basis for numerous analyses (3, 42, 56). The HELP assay has a bias toward CpG-rich sites, in effect concentrating on promoter regions. The performance of the HELP assay in CpG-poor regions is reduced compared with bisulfite conversion–based methods. Further studies will show whether similar observations can be made using CpG-poor regions. It remains to be seen whether relevant changes in fact have a CpG bias and are sufficiently covered by our approach. The clinical relevance of our findings will need to be confirmed in well-powered prospective randomized clinical trials.

In summary, the findings presented here identify a large fraction of CpG dinucleotides in human HSC as methylated, show a human-specific methylation decrease specifically during STHSC-to-CMP commitment, and reveal a stem cell commitment–associated epigenetic signature as robustly and independently prognostically significant for OS of AML patients.

Methods

BM samples. BM samples from healthy individuals were obtained from All-Cells LLC.

High-speed multi-parameter FACS. FACS of human HSPC populations was performed as described before (15–17, 19–21, 25). Mononuclear cells from BM aspirates were isolated by density gradient centrifugation. CD34⁺ cells were enriched by immunomagnetic beads (Miltenyi Biotec). The resulting cells were lineage depleted (Lin[−]) using PE-Cy5–conjugated (Tricolor) monoclonal antibodies against CD2, CD3, CD4, CD7, CD10, CD11b, CD14, CD15, CD19, CD20, CD56, and Glycophorin A (all BD Biosciences). Further distinction into HSPC subsets was performed using fluorochrome-conjugated antibodies against CD34, CD38, CD90, CD45RA, and CD123 (all eBioscience). LTHSC (Lin[−], CD34⁺, CD38[−], CD90⁺), STHSC (Lin[−], CD34⁺, CD38[−], CD90[−]), CMP (Lin[−], CD34⁺, CD38⁺, CD123⁺, CD45RA[−]), and MEP (Lin[−], CD34⁺, CD38⁺, CD123[−], CD45RA[−]) were sorted into RLT extraction buffer (QIAGEN). Flow cytometric analysis and cell separation were performed on a FACSAriaII special order system (BD) equipped with 4 lasers (407 nm, 488 nm, 561/568 nm, 633/647 nm).

Preparation of nucleic acids. After sorting into RLT buffer (QIAGEN), homogenization of the cells was achieved by passing the cells 5 times through a needle. Simultaneous harvest of RNA and genomic DNA was achieved with the AllPrep Kit (QIAGEN) following the instructions of the manufacturer. Total RNA was linearly amplified and transcribed with the MessageAmp Kit AM1751 (Ambion/ Life Technologies) prior to microarray gene expression analysis following the NimbleGen Arrays user's guide (NimbleGen). Integrity of RNA and cDNA was verified at each step of amplification using the Agilent Bioanalyzer 2100 (Agilent).

DNA methylation analysis by nanoHELP. Methylation analysis by the HELP assay (22, 57–59) and a modified protocol to successfully work with low genomic DNA yield from low numbers of sorted stem and progenitor cells have been described (24, 25). Integrity of genomic DNA of high molecular weight was assured by electrophoresis for all samples used. HpaII or MspI (NEB) digestions of genomic DNA were performed overnight prior to overnight ligation of the HpaII adapter with T4 ligase. PCR-amplified adapter-ligated HpaII or MspI fragments were submitted to Roche-NimbleGen. Labeling and DNA hybridization onto a human hg17 custom-designed oligonucleotide array (50 mers) was carried out. The 2005-07-20_HG17_HELP_Promoter array covers 25,626 HpaII amplifiable fragments (HAF) at gene promoters, defined as regions 2 kb upstream and downstream of transcriptional start sites (TSS). EpiTyper by MassArray (Sequenom) was used to confirm methylation of selected loci as described (23, 60).

Microarray quality control. Uniformity of hybridization was evaluated by adapting a published algorithm (61) for the NimbleGen platform. Hybridizations with strong regional artifacts were discarded and repeated. Normalized signal intensities from each array were compared with a 20% trimmed mean of signal intensities across all arrays in that experiment. Arrays with significant intensity bias that could not be explained by the biology of the sample were excluded.

HELP data processing. Signal intensities at each HAF were calculated as 25% trimmed mean of their component probe-level signal intensities. Any fragments found within the level of background MspI signal intensity (equaling 2.5 mean-absolute-difference, MAD) above the median of random probe signals were regarded “failed”. These “failed” loci represent the population of fragments that did not amplify by PCR. Loci were designated “methylated” when the level of HpaII signal intensity was indistinguishable from background as described for MspI. Fragments successfully amplified by PCR, i.e., distinguishable above background, were subjected to normalization. For this, an intraarray quantile approach was used: HpaII/MspI ratios are aligned across density-dependent sliding windows of fragment size–sorted data. The $\log_2(\text{HpaII/MspI})$ was used as a representative



for methylation and analyzed as a continuous variable. If the centered $\log_2(\text{HpaII}/\text{MspI})$ ratio was less than 0, the corresponding fragment was considered methylated. It was considered hypomethylated in cases where $\log_2(\text{HpaII}/\text{MspI})$ was greater than 0.

Gene expression profiling. Gene expression profiling was performed on NimbleGen HG18 arrays (design name 2006-08-03_HG18_60mer_expr, Roche-NimbleGen). Profiling was performed by the Epigenomics Shared Facility, Albert Einstein College of Medicine.

Meta-analysis of the GSE24505 and GSE18700 AML data sets. Previously published data for gene expression (Nimblegen 2005-04-20_Human_60mer_1in2 arrays) and DNA methylation (2005-07-20_HG17_HELP_Promoter arrays) were retrieved from the GEO server (GSE24505). Additional annotations were extracted from these files. The methylation status of respective loci could be directly compared between the data describing human HSPC that we analyzed and the published GSE24505 AML data due to identical platforms.

Statistics. HELP loci were annotated using UCSC annotations for hg17. Means of locus-specific methylation between consecutive HSPC subsets were compared using Student's 2-sided t test for unpaired samples. Significance was assumed at $P < 0.05$. SAM was performed using Multiple Experiment Viewer, as was supervised clustering using Euclidean distance correlation with complete linkage. SAM ($q < 0.015$) was performed on the values of the 4 cell populations that remained significant after an initial SAM had filtered probes in which the difference between replicates was more significant than the difference between stages of differentiation. A similar approach to account for variability in analyses of DNA cytosine methylation has recently been published (7). Survival data and corresponding methylation values have previously been published (41, 42). An epigenetic score was calculated by summing absolute values of the median-centered methylation values ($\log_2[\text{HpaII}/\text{MspI}]$) of the 561 signature loci for each patient sample. Samples from ECOG (GEO GSE24505) and HOVON (GEO GSE18700, GSE 14468) studies (4, 42) were ranked and uniformly dichotomized according to the 55th percentile into patients with a low and those with a high signature score. An association of this score with Kaplan-Meier survival estimates (62) was probed by the log-rank test and assumed to be statistically significant at $P < 0.05$. The association of individual methylation loci and genes in this set of patient samples was probed by *globaltest* (43) after linear transformation to obtain positive values, similarly to a recently published analysis (63). Gene expression analysis was performed in an identical fashion, with $q < 0.2$. Ingenuity (Ingenuity Systems) was used for pathway analysis. After Benjamini-Hochberg correction, "Top bio functions" that were significantly ($P < 0.05$) associated with the 561 constituents of the epigenetic commitment-associated stem cell signature were entered into a pathway generator. The top 5 "Canonical pathways" and the top 3 characteristics in "Function and disease" were chosen for display. Circos plots were prepared following website instructions (<http://circos.ca>). To perform correlation analyses between methylation probes and gene expression changes, as well as *globaltest* analyses of gene expression signatures from various microarray platforms, all probes were remapped to hg19 using lift-Over (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>), and remapped probes were associated with overlapping hg19 RefSeq genes (UCSC table browser; <http://genome.ucsc.edu/cgi-bin/hgTables>, retrieved 9/18/12) using bedtools intersect, and closest nonoverlapping genes were associated using bedtools closest. Additional identifiers of these genes were retrieved from

ENSEMBL BioMart using biomaRt in R/Bioconductor to match probe identifiers across various microarray platforms (Nimblegen HG18 for the ECOG data set, ref. (42); Nimblegen HG17 for healthy human HSPC, Affymetrix U133plus2.0 for the signatures published by Eppert et al., ref. 46; Entrez IDs for those published by Gentles et al., ref. 45). Collapsing of multiple probes, where necessary, was performed using the collapseRows function in the R/Bioconductor WGCNA package. Genomic coordinates of pre-microRNA in the hg19 genome were retrieved from miRBase (<ftp://mirbase.org/pub/mirbase/20/genomes/hsa.gff2>; miRBase v20, date: 2013-05-24, genome build: GRCh37.p5, NCBI_Assembly:GCA_000001405.6).

Data were compared by 2-sided t test for unpaired samples, or by SAM using Multiple Experiment Viewer (version 4.8) and q -value thresholds as indicated. To determine the association of DNA methylation or RNA expression signatures with OS, Kaplan-Meier survival analysis was performed and survival differences between groups were assessed with the log-rank test. Alternatively, *globaltest* analysis was performed. Univariate and multivariate analyses of hazard ratios were performed using the Cox proportional hazards model. Survival analyses were performed with R/Bioconductor software and the packages *globaltest*, *survival*, *eha*, and *MASS*, or with GraphPad Prism software (version 6). P -values < 0.05 were considered significant.

Study approval. All studies were approved by the Albert Einstein College of Medicine IRB.

Data and materials availability. Data describing DNA cytosine methylation and gene expression of healthy human HSPC have been deposited in GEO (GSE52152).

Acknowledgments

We thank the Einstein Human Stem Cell FACS and Xenotransplantation Facility (supported by NYSTEM grant C024172) for expert technical assistance. This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant CH945/1-1 (to M. Christopeit), a Howard Temin Award of the National Cancer Institute (R00CA131503, to U. Steidl), and a NYSTEM research grant (C024306, to U. Steidl). U. Steidl is the Diane and Arthur B. Belfer Faculty Scholar in Cancer Research of the Albert Einstein College of Medicine.

Received for publication September 25, 2013, and accepted in revised form November 14, 2013.

Address correspondence to: Amit Verma, Division of Hematologic Malignancies, Department of Medicine (Oncology), Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Chanin Bldg., Room 302B, 1300 Morris Park Avenue, Bronx, New York 10461, USA. Phone: 718.430.8761; Fax: 718.430.8702; E-mail: amit.verma@einstein.yu.edu. Or to: Ulrich Steidl, Department of Cell Biology, Division of Hematologic Malignancies, Department of Medicine (Oncology), Albert Einstein College of Medicine, Chanin Bldg., Room 606, 1300 Morris Park Avenue, Bronx, New York 10461, USA. Phone: 718.430.3437; Fax: 718.430.8574; E-mail: ulrich.steidl@einstein.yu.edu.

- Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer*. 2012; 12(9):599–612.
- Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013; 368(22):2059–2074.
- Patel JP, et al. Prognostic relevance of integrated

- genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1079–1089.
- Figuerola ME, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. 2010;17(1):13–27.
- Ji H, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*. 2010;467(7313):338–342.
- Shearstone JR, Pop R, Bock C, Boyle P, Meissner

- A, Socolovsky M. Global DNA demethylation during mouse erythropoiesis in vivo. *Science*. 2011; 334(6057):799–802.
- Bock C, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell*. 2012;47(4):633–647.
- Broske AM, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet*. 2009;41(11):1207–1215.



9. Trowbridge JJ, Snow JW, Kim J, Orkin SH. DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell*. 2009;5(4):442–449.
10. Challen GA, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*. 2012;44(1):23–31.
11. Sasaki M, et al. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature*. 2012;488(7413):656–659.
12. Moran-Crusio K, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011;20(1):11–24.
13. Quivoron C, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*. 2011;20(1):25–38.
14. Li Z, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*. 2011;118(17):4509–4518.
15. Baum CM, Weissman IL, Tsukamoto AS, Buckle AM, Peault B. Isolation of a candidate human hematopoietic stem-cell population. *Proc Natl Acad Sci U S A*. 1992;89(7):2804–2808.
16. Craig W, Kay R, Cutler RL, Lansdorp PM. Expression of Thy-1 on human hematopoietic progenitor cells. *J Exp Med*. 1993;177(5):1331–1342.
17. Manz MG, Miyamoto T, Akashi K, Weissman IL. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A*. 2002;99(18):11872–11877.
18. Steidl U, et al. Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells. *Nat Genet*. 2006;38(11):1269–1277.
19. Steidl U, et al. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J Clin Invest*. 2007;117(9):2611–2620.
20. Will B, Steidl U. Multi-parameter fluorescence-activated cell sorting and analysis of stem and progenitor cells in myeloid malignancies. *Best Pract Res Clin Haematol*. 2010;23(3):391–401.
21. BarreYRO L, et al. Overexpression of IL-1 receptor accessory protein in stem and progenitor cells and outcome correlation in AML and MDS. *Blood*. 2012;120(6):1290–1298.
22. Khulan B, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res*. 2006;16(8):1046–1055.
23. Figueroa ME, et al. An integrative genomic and epigenomic approach for the study of transcriptional regulation. *PLoS ONE*. 2008;3(3):e1882.
24. Oda M, et al. High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res*. 2009;37(12):3829–3839.
25. Will B, et al. Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. *Blood*. 2012;120(10):2076–2086.
26. Heuck CJ, et al. Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis. *J Immunol*. 2013;190(6):2966–2975.
27. Behre G, et al. Ras signaling enhances the activity of C/EBP alpha to induce granulocytic differentiation by phosphorylation of serine 248. *J Biol Chem*. 2002;277(29):26293–26299.
28. Geletu M, et al. Target proteins of C/EBPalpha30 in AML: C/EBPalpha30 enhances sumoylation of C/EBPalpha42 via up-regulation of Ubc9. *Blood*. 2007;110(9):3301–3309.
29. Pulikkan JA, et al. C/EBPalpha regulated microRNA-34a targets E2F3 during granulopoiesis and is down-regulated in AML with CEBPA mutations. *Blood*. 2010;116(25):5638–5649.
30. Pulikkan JA, et al. Cell-cycle regulator E2F1 and microRNA-223 comprise an autoregulatory negative feedback loop in acute myeloid leukemia. *Blood*. 2010;115(9):1768–1778.
31. Braun BS, Archard JA, Van Ziffle JA, Tuveson DA, Jacks TE, Shannon K. Somatic activation of a conditional KrasG12D allele causes ineffective erythropoiesis in vivo. *Blood*. 2006;108(6):2041–2044.
32. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 2012;481(7380):157–163.
33. Porter CC, et al. Integrated genomic analyses identify WEE1 as a critical mediator of cell fate and a novel therapeutic target in acute myeloid leukemia. *Leukemia*. 2012;26(6):1266–1276.
34. Sun SM, et al. Transition of highly specific microRNA expression patterns in association with discrete maturation stages of human granulopoiesis. *Br J Haematol*. 2011;155(3):395–398.
35. Garzon R, Croce CM. MicroRNAs in normal and malignant hematopoiesis. *Curr Opin Hematol*. 2008;15(4):352–358.
36. Chen J, Odenike O, Rowley JD. Leukaemogenesis: more than mutant genes. *Nat Rev Cancer*. 2010;10(1):23–36.
37. Jiang X, et al. Blockade of miR-150 maturation by MLL-fusion/MYC/LIN-28 is required for MLL-associated leukemia. *Cancer Cell*. 2012;22(4):524–535.
38. Li Z, et al. miR-196b directly targets both HOXA9/MEIS1 oncogenes and FAS tumour suppressor in MLL-rearranged leukaemia. *Nat Commun*. 2012;3:688.
39. Valk PJ, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004;350(16):1617–1628.
40. Verhaak RG, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94(1):131–134.
41. Fernandez HF, et al. Anthracycline dose intensification in acute myeloid leukemia. *N Engl J Med*. 2009;361(13):1249–1259.
42. Figueroa ME, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553–567.
43. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*. 2004;20(1):93–99.
44. Cox DR. Regression models and life tables. *J R Stat Soc B*. 1972;34(2):187–220.
45. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA*. 2010;304(24):2706–2715.
46. Eppert K, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011;17(9):1086–1093.
47. Li Z, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. *J Clin Oncol*. 2013;31(9):1172–1181.
48. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25(10):1010–1022.
49. Suzuki M, et al. Late-replicating heterochromatin is characterized by decreased cytosine methylation in the human genome. *Genome Res*. 2011;21(11):1833–1840.
50. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–589.
51. Bhagat TD, et al. miR-21 mediates hematopoietic suppression in MDS by activating TGF-beta signaling. *Blood*. 2013;121(15):2875–2881.
52. Zhou L, et al. Reduced SMAD7 leads to overactivation of TGF-beta signaling in MDS that can be reversed by a specific inhibitor of TGF-beta receptor I kinase. *Cancer Res*. 2011;71(3):955–963.
53. Zhou L, et al. Inhibition of the TGF-beta receptor I kinase promotes hematopoiesis in MDS. *Blood*. 2008;112(8):3434–3443.
54. Schubeler D. Molecular biology. Epigenetic islands in a genetic ocean. *Science*. 2012;338(6108):756–757.
55. Yu Y, et al. High resolution methylome analysis reveals widespread functional hypomethylation during adult human erythropoiesis. *J Biol Chem*. 2013;288(13):8805–8814.
56. Gonen M, et al. CD25 expression status improves prognostic risk classification in AML independent of established biomarkers: ECOG phase 3 trial, E1900. *Blood*. 2012;120(11):2297–2306.
57. Zhou L, et al. Aberrant epigenetic and genetic marks are seen in myelodysplastic leukocytes and reveal Dock4 as a candidate pathogenic gene on chromosome 7q. *J Biol Chem*. 2011;286(28):25211–25223.
58. Nischal S, et al. Methylome profiling reveals distinct alterations in phenotypic and mutational subgroups of myeloproliferative neoplasms. *Cancer Res*. 2013;73(3):1076–1085.
59. Figueroa ME, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest*. 2013;123(7):3099–3111.
60. Figueroa ME, et al. Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood*. 2009;113(12):2795–2804.
61. Thompson RF, et al. An analytical pipeline for genomic representations used for cytosine methylation studies. *Bioinformatics*. 2008;24(9):1161–1167.
62. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457–481.
63. Kawahara M, et al. H2O-like Homeobox regulates early hematopoiesis and promotes acute myeloid leukemia. *Cancer Cell*. 2012;22(2):194–208.