DNA methylation epitypes highlight underlying developmental and disease pathways in acute myeloid leukemia

Brian Giacopelli, Min Wang, Ada Cleary, Yue-Zhong Wu, Anna Reister Schultz, Maximilian Schmutz, James S. Blachly, Ann-Kathrin Eisfeld, Bethany Mundy-Bosse, Sebastian Vosberg, Philipp A. Greif, Rainer Claus, Lars Bullinger, Ramiro Garzon, Kevin R. Coombes, Clara D. Bloomfield, Brian J. Druker, Jeffrey W. Tyner, John C. Byrd, Christopher C. Oakes

Angaben zur Veröffentlichung / Publication details:

Giacopelli, Brian, Min Wang, Ada Cleary, Yue-Zhong Wu, Anna Reister Schultz, Maximilian Schmutz, James S. Blachly, et al. 2021. "DNA methylation epitypes highlight underlying developmental and disease pathways in acute myeloid leukemia." *Genome Research* 31: 747–61. https://doi.org/10.1101/gr.269233.120.

Research

DNA methylation epitypes highlight underlying developmental and disease pathways in acute myeloid leukemia

Brian Giacopelli,[1,2] Min Wang,[3] Ada Cleary,[1,2] Yue-Zhong Wu,[1,2] Anna Reister Schultz,[4] Maximilian Schmutz,[5] James S. Blachly,[1,2,3] Ann-Kathrin Eisfeld,[1,2] Bethany Mundy-Bosse,[1,2] Sebastian Vosberg,[6,7] Philipp A. Greif,[6,8,9] Rainer Claus,[10] Lars Bullinger,[11] Ramiro Garzon,[1,2] Kevin R. Coombes,[3] Clara D. Bloomfield,[1,2] Brian J. Druker,[4] Jeffrey W. Tyner,[4] John C. Byrd,[1,2] and Christopher C. Oakes[1,2,3]

[1]*Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, Ohio 43210, USA;* [2]*The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA;* [3]*Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio 43210, USA;* [4]*Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon 97239, USA;* [5]*Hematology and Oncology, Medical Faculty, University of Augsburg, 86159 Augsburg, Germany;* [6]*Department of Medicine III, University Hospital, LMU Munich, 80539 Munich, Germany;* [7]*Institute of Computational Biology, Helmholtz Zentrum München–German Research Center for Environmental Health, 85764 Munich, Germany;* [8]*German Cancer Consortium (DKTK), Partner Site Munich, 69120 Heidelberg, Germany;* [9]*German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany;* [10]*Department of Medicine II, Stem Cell Transplantation Unit, Klinikum Augsburg, Ludwig-Maximilians University Munich, 86156 Munich, Germany;* [11]*Department of Hematology, Oncology and Tumorimmunology, Charité–Universitätsmedizin, 13353 Berlin, Germany*

Acute myeloid leukemia (AML) is a molecularly complex disease characterized by heterogeneous tumor genetic profiles and involving numerous pathogenic mechanisms and pathways. Integration of molecular data types across multiple patient cohorts may advance current genetic approaches for improved subclassification and understanding of the biology of the disease. Here, we analyzed genome-wide DNA methylation in 649 AML patients using Illumina arrays and identified a configuration of 13 subtypes (termed "epitypes") using unbiased clustering. Integration of genetic data revealed that most epitypes were associated with a certain recurrent mutation (or combination) in a majority of patients, yet other epitypes were largely independent. Epitypes showed developmental blockage at discrete stages of myeloid differentiation, revealing epitypes that retain arrested hematopoietic stem-cell-like phenotypes. Detailed analyses of DNA methylation patterns identified unique patterns of aberrant hyper- and hypomethylation among epitypes, with variable involvement of transcription factors influencing promoter, enhancer, and repressed regions. Patients in epitypes with stem-cell-like methylation features showed inferior overall survival along with up-regulated stem cell gene expression signatures. We further identified a DNA methylation signature involving STAT motifs associated with *FLT3*-ITD mutations. Finally, DNA methylation signatures were stable at relapse for the large majority of patients, and rare epitype switching accompanied loss of the dominant epitype mutations and reversion to stem-cell-like methylation patterns. These results show that DNA methylation-based classification integrates important molecular features of AML to reveal the diverse pathogenic and biological aspects of the disease.

[Supplemental material is available for this article.]

Acute myeloid leukemia (AML) is a clinically and molecularly heterogeneous disease. Recurrent genetic aberrations, such as chromosomal rearrangements and gene mutations, primarily form the basis of our current understanding of pathogenesis and are used for patient classification (Lowenberg et al. 1999; Döhner et al. 2010, 2017; The Cancer Genome Atlas Research Network 2013). AML has a low level of genetic aberrations relative to other cancers, but several recurrent aberrations are significantly associated with prognosis and to tumor cell phenotypes. However, genetic markers do not completely explain the range of phenotypes observed in tumor cells and disease outcomes.

Efforts have been made to classify AML based on the phenotype rather than the genotype in the form of morphology or gene expression (Bennett et al. 1982; Mrózek et al. 2009; Ng et al. 2016). AML arises from cells developing from hematopoietic stem and progenitor cells (HSPCs) into a wide range of developmental phenotypes within the myeloid lineage, and developmental arrest is a key aspect of AML pathogenesis. Epigenetic mechanisms are central to cellular differentiation by governing the control expression of key developmental gene expression programs. DNA

Corresponding author: Christopher.Oakes@osumc.edu
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.269233.120. Freely available online through the *Genome Research* Open Access option.

methylation, the addition of a methyl group to the 5′ carbon of cytosines, is the most broadly studied epigenetic mark. Differential DNA methylation patterns among AML patients has been used to classify patients with varying results identifying between two and 16 subgroups depending on the study design (Bullinger et al. 2010; Figueroa et al. 2010; Melnick 2010; Glass et al. 2017). Many of these subgroups showed associations with genetic aberrations but others represented novel subgroups. Decoding altered genome-wide DNA methylation patterns can provide insight into novel disease-relevant pathways by association to global chromatin states and enrichment of genomic features. DNA methylation is an attractive biomarker owing to its stability and has been effectively used in multiple cancers to guide therapy (Koch et al. 2018).

Because of the high degree of heterogeneity in AML, we sought to assemble a large cohort of AML samples to uncover the breadth of distinct genome-wide DNA methylation states and to use this classification structure as a basis for a novel investigation of aberrant disease pathways. Here, we analyzed genome-wide DNA methylation profiles from well-characterized AML samples from the Beat AML project combined with published data, compiling the largest number of methylation profiles studied to date. We used an unbiased clustering approach to define distinct subtypes of AML patients and integration with genomic and gene expression data uncovered that each subtype is associated with a unique combination of developmental and disease-specific features. Our findings highlight prevalent, subtype-specific activation of inflammatory pathways as a key mechanism uniting epigenetic, expression and genetic features with poor survival in AML.

Results

Classification of AML patients into distinct epitypes using genome-wide DNA methylation

AML patients were classified in an unbiased fashion by DNA methylation patterns using samples from newly diagnosed patients obtained from the Beat AML Consortium ($n = 226$) (Tyner et al. 2018) and the Ohio State University (OSU) $n = 27$, combined with TCGA ($n = 190$) (The Cancer Genome Atlas Research Network 2013) and five other independent studies comprising $n = 206$ patients (Schmutz et al. 2013; Leonard et al. 2014; Jung et al. 2015; Ferreira et al. 2016; Eisfeld et al. 2017; Qu et al. 2017) for a total of $n = 649$ patients. DNA methylation was interrogated by Illumina methylation arrays, which provide the DNA methylation levels of CpG dinucleotides primarily in promoter and regulatory regions (Bibikova et al. 2011). We reduced the data based on overall variance using the 500 most variable probes for cluster analysis and performed unsupervised *k*-medoids-based clustering. The total number of groups (k) was determined using the Auer-Gervini method (Auer and Gervini 2008; Wang et al. 2018) that uncovered a minimum of 11 informative principal components (Supplemental Fig. S1A). By performing clustering and subsequent silhouette analyses with increasing k from 11, we determined the optimal group number to be 13 (Supplemental Fig. S1B,C). We termed these clusters AML DNA methylation epitypes 1–13 (E1–E13). All epitypes comprised samples from multiple studies (median = 5, range = 3–8). Epitype classification remained largely stable upon varying the number of most variable probes used, with ~90% sample assignments unchanged and variation in epitype assignment primarily restricted within E5,E6 and E11–E13 (Supplemental Fig. S1D). Although sample purity from publicly available sources was not uniformly available, tumor cell content inferred from somatic mutation data revealed similar sample purity levels across epitypes (Supplemental Fig. S1E). Hierarchical clustering revealed three primary clusters (superclusters) each containing 3–5 distinct DNA methylation epitypes (Fig. 1A). Owing to the high degree of complexity of epitype-specific patterns, t-distributed stochastic neighbor embedding (t-SNE) plots were used for subsequent visualization of epitypes and largely agreed with *k*-medoids-based clustering (Fig. 1B). Several CpGs in this signature were proximal to genes implicated in AML pathogenesis, such as *MEIS1* and several within the *HOXB* locus (Supplemental Table S1; Ferreira et al. 2016). However, the majority of the CpGs composing the epityping signature were located in loci with undescribed associations to AML.

Epitypes frequently associate with genetic aberrations

To explore the underlying basis of distinct epitypes, we first considered the relationship to recurrent genetic aberrations. We found associations between epitypes and common genetic aberrations consistent with past studies (Figueroa et al. 2010; Glass et al. 2017); however, we found that this linkage was not universal (Fig. 1C; Supplemental Table S2). Four epitypes were enriched for alterations in key myeloid transcription factors (TFs): E1–E3 were enriched for the TF fusions *PML-RARA*, inv(16)/*CBFB*, *AML-ETO*, respectively, and E4 was enriched for *CEBPA* mutations. These epitypes showed the highest association of genetic aberrations (Fig. 1C). Epitypes E1–E4 together formed a distinct supercluster (Fig. 1A) with dominant, epitype-defining genetic aberrations known to result in arrest of myeloid development and associate with favorable outcomes (Speck and Gilliland 2002; Pabst and Mueller 2007; De Braekeleer et al. 2014). Epitypes 5 and 6 were enriched in a variety of chromosomal rearrangements generating fusions involving *KMT2A* (previously known as *MLL*) on 11q23. Multiple *KMT2A* fusion partners have been described in acute leukemias (Winters and Bernt 2017), and we observed common AML fusion partners in both epitypes. Epitypes E7–E10 were strongly enriched for cytogenetically normal genotypes carrying mutations in the *NPM1* gene. Epitype 8 was enriched for *NPM1* mutations alone, whereas E7, E9, and E10 were enriched for *NPM1* mutations in conjunction with *DNMT3A*, *TET2*, and *IDH1/2* mutations, respectively. E11–E13 formed a patient supercluster with relative epigenetic similarity among patterns (Fig. 1D). E11 was enriched in *IDH1/2* mutations lacking accompanying *NPM1* mutations. Epitypes E12 and E13 lacked a consistent mutation pattern involving a majority of samples, yet retained mutations associated with genomic instability, such as *TP53* mutations and complex karyotype, in a minority of samples. These results show that there is a close association between recurrent genetic aberrations in many AML epitypes, yet others lack a dominant, epitype-defining genotype. In addition, many samples lack the dominant mutation within a particular epitype (Fig. 1D), indicating that other cellular events may converge within epitypes to phenocopy the impact on the epigenome, termed “epiphenocopy” events.

Differences between differentiation states reveal nonmutational features of epitypes

Unlike identifying somatic mutations from germline sequences, all cell types have distinct epigenetic patterns, thus patterns originating from normal counterparts must be accounted for when determining tumor cell–specific epigenetic changes. We have previously shown this to be important for deriving tumor-specific

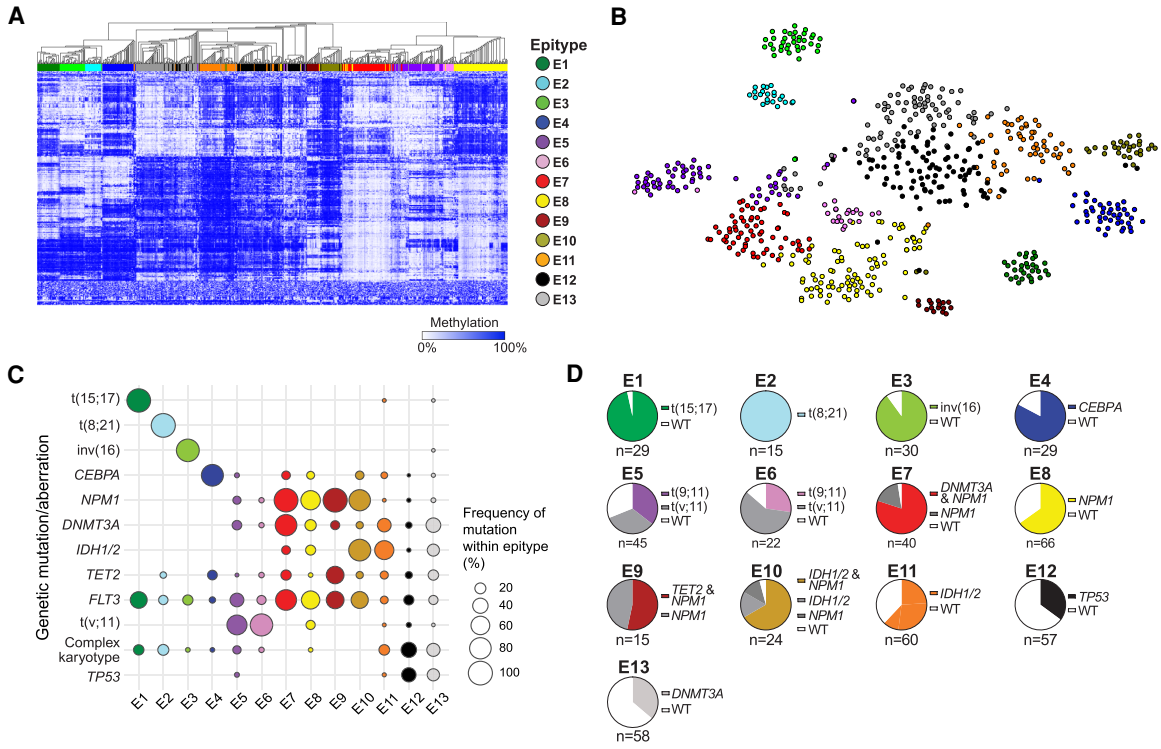

Figure 1. Unsupervised clustering of 649 AML samples using DNA methylation and relationship with genetic mutations. (*A*) Heatmap of the 500 most variable CpGs across all samples organized by hierarchical clustering. Samples are annotated by epitype assignment using PAM clustering (colors). (*B*) The same 500 most variable displayed by t-SNE plot. (*C*) The distribution of the most common recurrent genetic aberrations in AML within the epitypes. Bubble size represents the percentage of patients within the epitype with the corresponding aberration. (*D*) Pie charts displaying the frequency of the most common (dominant mutation/combination) within each epitype.

events in chronic lymphocytic leukemia (CLL) (Oakes et al. 2016). AML is known to display a wide range of hematopoietic differentiation states, from undifferentiated, stem-cell-like phenotypes, to differentiated mature cells of the myeloid lineage (Bennett et al. 1982; Griffin et al. 1983). In addition, some AML cells retain immunophenotypic features of granulocytic, erythrocytic, or lymphocytic lineages (Bradstock et al. 1989; Macedo et al. 1995; Matutes et al. 1997). To clearly identify AML epitype-specific DNA methylation events, we first expanded our analysis to include all CpGs measured across all samples ($n = 426{,}862$). We generated Illumina array DNA methylation profiles of sorted hematopoietic populations and combined with publicly available sources (Reinius et al. 2012; The Cancer Genome Atlas Research Network 2013; Jung et al. 2015; Qu et al. 2017). We generated a DNA methylation signature that encompasses normal hematopoietic development by assembling a probe set of differential methylation between each subpopulation and HSPC. This signature recapitulated the branches of the hematopoietic lineages (Supplemental Fig. S2A). Investigation of this signature together with all AML samples revealed that the largest proportion of the variation (principal component 1 [PC1]) among AML samples occurred between HSPC/myeloid progenitors and mature myeloid cells as expected (Fig. 2A, left). Lymphocytes were positioned on the side of PC1 with the myeloid progenitors, indicating that PC1 relates to myeloid-specific development. PC2 primarily related to a lymphoid-dominant signature distinct from the vast majority of AML samples. PC3 largely separated some myeloid progenitors from HSPCs, as well as granulocytes from monocytes and macrophages, revealing that AMLs were more similar to monocytes/macrophages than granulocytes, and, on the progenitor side, are more similar to HSPC than other progenitors, such as CMP and MEP (Fig. 2A, right). Taken together, this analysis supports that AML DNA methylation states generally occur between HSPCs and monocytes/macrophages. Indeed, GMPs are a known intermediate transitional subtype between progenitors and mature myeloid cells and were located centrally in PC1. We further consolidated the developmental signature to the 5000 most differentially methylated probes between HSPC and monocyte samples. Using this signature, we observed that AML epitypes occupy specific ranges within the HSPC to monocyte developmental spectrum (Fig. 2B). E11, E12, and E13 fell closer to HSPCs, with E11 (*IDH1/2*) generally less differentiated than HSPCs. E5 (*KMT2A*) and E7 (*NPM1 +DNMT3A*) fell closer to monocytes, with some samples showing further differentiation toward macrophages, likely caused by tumor-specific methylation changes at developmentally regulated CpGs. The French-American-British (FAB) classification is a morphological assessment incorporating the differentiation stage of AML cells that has historically been used as a prognostic marker (Bennett et al. 1982). Comparing FAB classifications across epitypes with available annotation ($n = 247$), we observed a discrete pattern across epitypes, with M0 (undifferentiated leukemia) scores occurring almost exclusively in E11–E13 and M5 scores (monocytic leukemia) highly enriched in E5, E7, E8 (Supplemental Fig. S2B). To further control for potential HSPC-monocyte developmental signature in AML samples, we investigated the enrichment of transcription factor recognition sequence motifs in regions displaying altered methylation in monocytes versus HSPCs. Several TF motifs were highly enriched in monocyte-

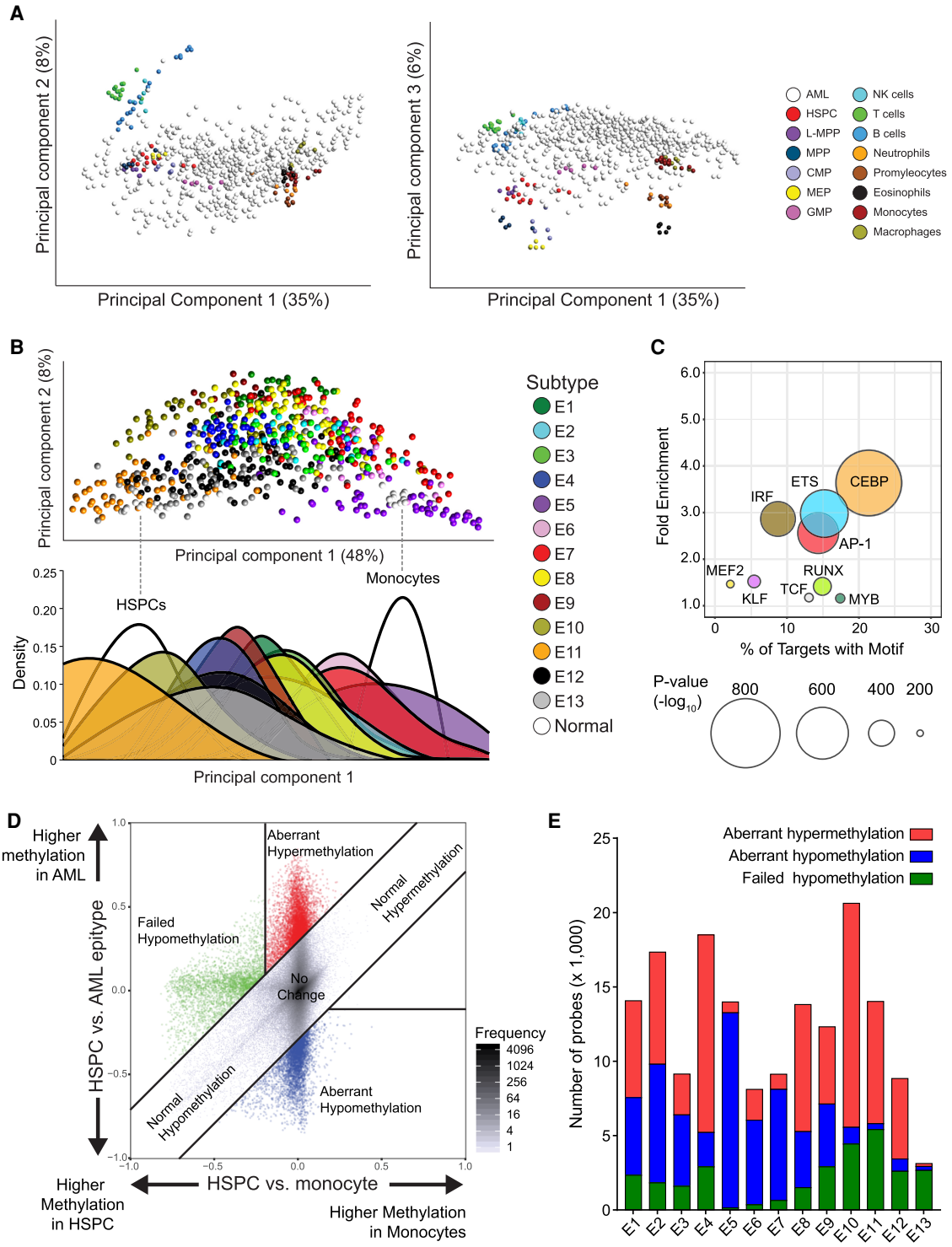

Figure 2. Assessment of DNA methylation associated with normal myeloid development enables identification of tumor-specific methylation. (*A*) Principal component analysis including healthy cell populations (colored) and AML samples (white) using the hematological developmental probe set (*left*, principal component [PC] 1 vs. PC2; *right*, PC1 vs. PC3). (*B*) Principal component analysis using a probe set of differentially methylated CpGs between HSPC and monocytes (white), including AML samples (colored by epitype). (*Below*) Density plot showing the distribution of samples with each epitype across PC1. (*C*) Bubble scatterplot of transcription factor motif enrichment in regions hypomethylated in monocytes compared to HSPC. Bubble size corresponds to the *P*-value, and color corresponds to transcription factor family. (*D*) A representative scatterplot simultaneously visualizing the DNA methylation differences in monocyte development (HSPCs to monocytes, *x*-axis) versus AML development using HSPCs as a reference (*y*-axis). Values represent average levels within HSPCs, monocytes, and AML epitype. Tumor-specific methylation changes are categorized as having aberrant hypermethylation (red) or aberrant hypomethylation (blue), separately from changes occurring in parallel with normal development (gray) or failing to occur as normally observed in monocytes (green). (*E*) Distribution of the tumor-specific methylation changes in each epitype. DNA methylation changes were compared simultaneously between normal and tumors (as shown in *D*) for all 13 epitypes.

specific hypomethylated regions, including CEBP, AP-1, ETS, IRF, and RUNX TF families (Fig. 2C). Disruption of several members of these TF families are associated with a block in AML differentiation (Tenen 2003).

We next used changes associated with normal development to independently identify tumor-specific methylation alterations. We visualized DNA methylation changes in individual AML epitypes versus the change that normally occurs in monocyte differentiation using HSPCs as a fixed reference (Fig. 2D). The probes that change methylation equally in both comparisons represent normal differentiation, whereas those that diverge from this axis represent tumor-specific methylation changes. DNA methylation gains and losses that were not observed to involve normal differentiation were termed aberrant hyper- or hypomethylation, respectively. As we consider AML samples that have not reached the differentiation state of mature myeloid cells to developmentally arrested, hypomethylation events that occurred during normal differentiation but failed to occur developmentally in AML were termed failed (developmental) hypomethylation. Each epitype displayed unique amounts and proportions of these classes of methylation change (Fig. 2E). TF-rearranged epitypes (E1–E4) generally displayed less variation of differentially methylated CpGs among samples, whereas variation was higher in more differentiated epitypes (E5–E7) (Supplemental Fig. S3). AML epitypes displayed variable amounts of failed hypomethylation that closely correlated with the degree of differentiation in the HSPC-monocyte signature (median PC1) (Supplemental Fig. S2C). Motif enrichment analyses revealed that all AML epitypes that show differentiation block (all except E5–E7) involve loss of hypomethylation programming associated with CEBP, SPI1/ETS, RUNX, AP-1, and IRF TFs, suggesting attenuated activity of these TF pathways broadly in AML (Supplemental Table S3).

Aberrant DNA methylation patterns reveal disease features associated with *NPM1* mutations

NPM1 is one of the most commonly mutated genes in AML, occurring in 30% of patients and is usually associated with a favorable outcome except in cases with certain co-occurring mutations (Papaemmanuil et al. 2016; Tyner et al. 2018). The vast majority (91%) of *NPM1* mutations were found in epitypes E7–E10 (Supplemental Table S2) either occurring alone (E8) or frequently in combination with known epigenetic modifier genes *DNMT3A* (E7), *TET2* (E9), or *IDH* (E10) (Fig. 1C,D). Although *NPM1* by itself is not described as an epigenetic modifier or regulator, epitype E8 retained among the most aberrant DNA methylation changes, involving both hyper- and hypomethylation (Fig. 2E). This pattern of aberrant methylation was modulated in combination with other epigenetic modifiers, skewing toward either hyper- or hypomethylation by *IDH1/2*, *TET2*, or *DNMT3A*, respectively (Fig. 3A). Regions of tumor-specific methylation can be used to infer pathway activation by investigation of TF motif enrichment in selectively hypomethylated regions (Hovestadt et al. 2014). Analysis of hypomethylated CpGs among E7–E10 revealed that E10 and E8 were largely subsets of the hypomethylation observed in E7, with E9 demonstrating a subset of uniquely hypomethylated CpGs (Fig. 3B). E7–E10 shared enrichment of RUNX, AP-1, and SPI1 motifs in the aberrant hypomethylated regions, which along with enrichment in failed hypomethylation (Supplemental Table S4), suggests that activity of these TFs are redirected from patterns of binding that occur normally (Fig. 3C). E7–E9 shared enrichment for EGR and TCF sequence motifs in tumor cells only, suggesting aberrant activation of these pathways in the NPM1 supercluster. E7 and E8 showed enrichment for HOX motifs, consistent with known activation of HOX genes in *NPM1*-mutated AML (Spencer et al. 2015). Epitype E9 displayed selective enrichment for FOX motifs, suggesting a novel activation of this TF family coincident with *TET2* mutations. Despite the combination of *DNMT3A* and 262 *NPM1* mutations, E7 displayed significantly more hypomethylation than others in the NPM1 supercluster (Fig. 3B), there was little difference in the TF enrichments, indicating that loss of *DNMT3A* function is not associated with specific pathway activation. These findings suggest that NPM1 loss is a strong modifier of DNA methylation patterns, which amplify methylation changes when combined with the disruption of an epigenetic regulator. Investigation of aberrant hypermethylation within the NPM1 supercluster revealed that E9 (TET2) and E10 (IDH) displayed higher levels of largely overlapping hypermethylation, which differed from those observed in E8 (*NPM1* alone) (Fig. 3D). *TET2* and *IDH1/2* mutations largely act through the same pathway leading to the inhibition of TET2-dependent demethylation in cancer (Scourzic et al. 2015).

To gain insight into the targeting and functional impact of hypermethylation, we partitioned the genome into chromatin states. These states functionally define regions as active, poised, repressed, or quiescent states in combination with enhancer, promoter, transcribed, and heterochromatic function by a combination of histone modifications using HSPCs as a reference (Ernst and Kellis 2010). Hypermethylated regions in E8 (*NPM1* alone) were enriched in regions containing the polycomb repressive histone modification, H3K27me3, and chromatin states containing this mark, such as poised promoters and enhancers as well as polycomb repressed regions (Fig. 3E; Supplemental Table S5). Conversely, hypermethylation in E9 (*TET2*) and E10 (*IDH1/2*) showed depletion in polycomb repressed regions and instead were enriched for active enhancers and regions flanking promoters/transcriptional start sites. Furthermore, hypermethylated enhancers in E9 and E10 were selectively enriched with TF motifs belonging to MEF2 and SPI1/ETS (Fig. 3F; Supplemental Table S6). Aberrant hypermethylation indicates not only the selective loss of the normal activity of these TFs in myeloid differentiation, but a further reversion to a state for these enhancer regions that is more immature than HSPC in *TET2* and *IDH*-mutant AML (Schüler et al. 2008; Will et al. 2015). These findings illustrate commonalities among AML methylation epitypes containing *NPM1* mutations (E7–E10) and highlight the distinct differential impact of mutations in epigenetic modifying enzymes when co-occurring with *NPM1* mutations.

AML epitypes E11–E13 display undifferentiated, HSPC-like features

Epitypes E11–E13 formed a distinct constellation of AML samples separate from clusters with highly prevalent NPM1 mutations, recurrent chromosomal rearrangements, and other genetic abnormalities. Although E11 contained *IDH1/2* mutations, E12 and E13 lacked highly recurrent genetic features (Fig. 1C,D), thus we further endeavored to uncover unique features associated with these enigmatic epitypes. E13 revealed little difference in the DNA methylation pattern to normal cells, with almost all changes representing failed hypomethylation (Fig. 4A). *DNMT3A* was the most commonly mutated gene in E13, but it was not associated with methylation loss in this epitype. CpGs displaying failed hypomethylation in E13 overlap almost entirely with E11 and E12 (Fig. 4B). These three epitypes were among the most

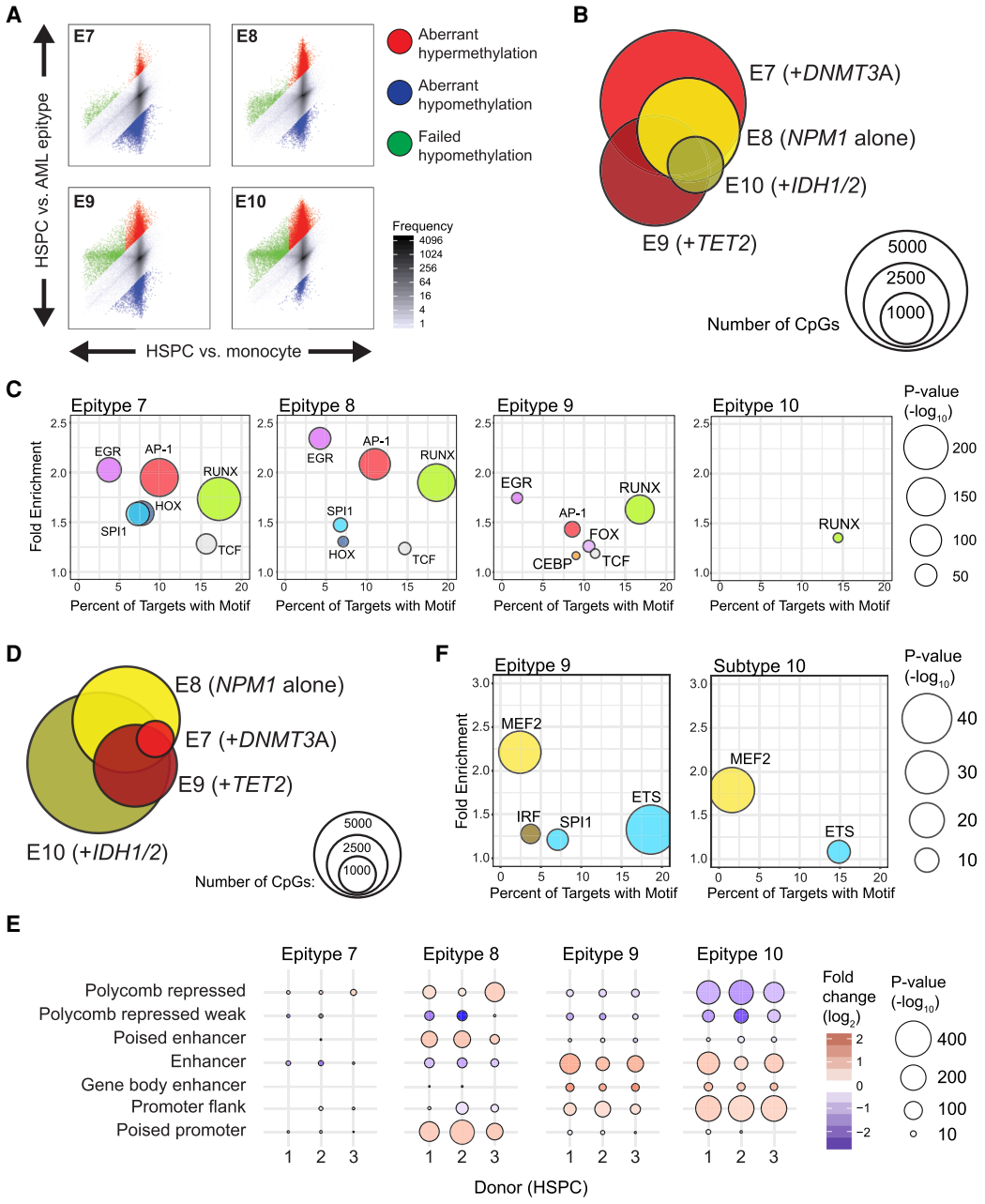

Figure 3. Analysis of tumor-specific methylation in the NPM1 constellation of epitypes (E7–E10). (*A*) Scatterplots comparing normal and tumor developmental methylation changes in E7–E10 highlight differential degrees of failed hypomethylation (green), aberrant hypermethylation (red), or aberrant hypomethylation (blue). (*B*) Venn diagram illustrating the numbers and overlap of aberrantly hypomethylated CpGs in E7–E10, with the dominant mutations within each epitype indicated (*NPM1* alone or *NPM1* plus a modifier mutation). (*C*) Bubble scatterplot of transcription factor motif enrichment in regions aberrantly hypomethylated in E7–E10. Bubble size corresponds to the *P*-value and color corresponds to transcription factor family. (*D*) Venn diagram of the aberrant hypermethylation in epitypes E7–E10. (*E*) Enrichment of aberrantly hypermethylated regions in selected chromatin states defined using the 15-state ChromHMM model in three independent HSPC samples. (*F*) Bubble scatterplot of transcription factor motif enrichment in regions aberrantly hypermethylated in epitypes 9 and 10.

undifferentiated in epigenetic developmental analyses (Fig. 2B) and furthermore contained all samples with undifferentiated (FAB M0) morphology (Supplemental Fig. S2B). Thus, we explored if these epitypes represent samples showing a stem-cell-like phenotype. We first performed t-SNE clustering using the 500-probe subtyping signature of all AML samples combined with normal hematopoietic lineage populations, and we found that normal cell types cluster in the vicinity of E11–E13, with HSPCs clustering within E13 (Supplemental Fig. S4A). We next incorporated gene expression data available in Beat AML and TCGA cohorts to examine the degree that these epitypes show hematopoietic stem cell gene expression signatures, such as the LSC17 signature (Ng et al. 2016). We found that epitypes E11–E13 showed the highest LSC17 scores across both data sets ($P < 0.001$) (Fig. 4C,D), which

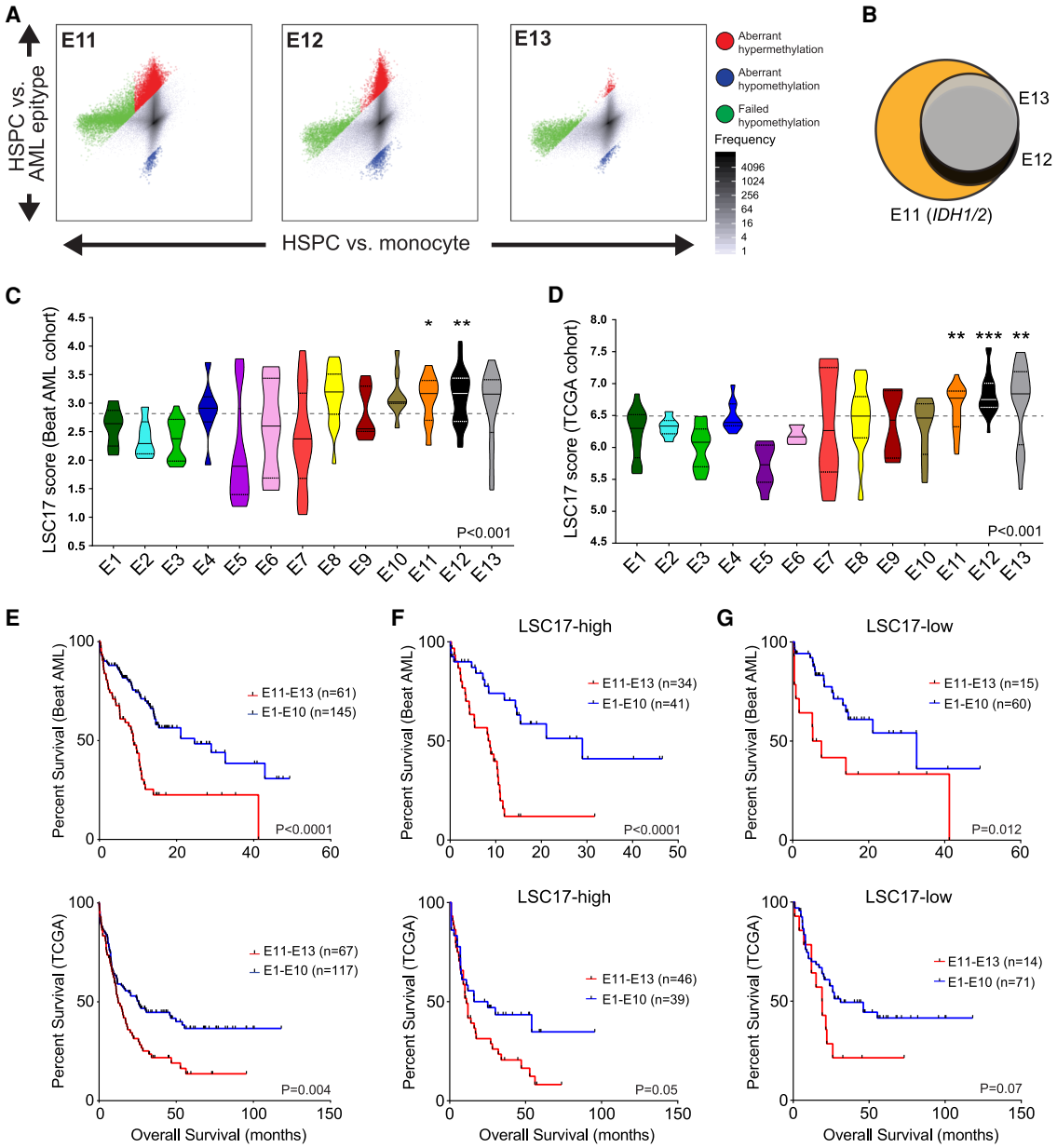

Figure 4. AML epitypes E11–E13 display stem-cell-like features. (*A*) Differential methylation scatterplots of E11–E13 highlight tumor-specific methylation changes. (*B*) Venn diagram showing overlap of failed hypomethylation in E11–E13. (*C*,*D*) LSC17 gene expression scores in the Beat AML (*C*) and the TCGA (*D*) cohort arranged by epitype. Cohort median value is indicated by the dotted line; significance evaluated by ANOVA test followed by comparison of E11–E13 individually versus E1–E10; adjusted *P*-values: (*) $P<0.05$; (**) $P<0.01$; (***) $P<0.001$. (*E*) Kaplan-Meier analysis of overall survival of E11–E13 compared to the other epitypes (E1–E10) in the Beat AML and TCGA cohorts. (*F*,*G*) Kaplan-Meier analysis of overall survival of E11–E13 compared to the other epitypes in the Beat AML and TCGA cohorts following separation into LSC17-high (*F*) and LSC17-low (*G*) groups using median dichotomization indicated above in *C* and *D*, respectively.

were further supported by high overall enrichment in two other independent stem cell signatures (Supplemental Fig. S4B; Gal et al. 2006; Gentles et al. 2010). Because stemness has been associated with poor outcomes (Amadori et al. 1996; Barbaric et al. 2007), we next investigated if patients from E11–E13 showed significantly poorer outcomes than other epitypes. We found that E11–E13 displayed inferior overall survival in both Beat AML and TCGA cohorts (Fig. 4E), which was also generally observed when considering individual epitypes (Supplemental Fig. S5). As LSC17 is a well-described prognostic signature of stemness, we next investigated if stem-cell-like epitypes E11–E13 retained independent prognostic impact relative to LSC17. Following median dichotomization of LSC17 scores, stem-cell-like epitypes further separated the overall survival of AML patients in the Beat AML cohort, with a similar trend observed in TCGA samples (Fig. 4F,G). In multivariate analyses, in the TCGA cohort, E11–E13 retained significance ($P<0.001$) and LSC17 did not (Supplemental Table S7).

Further investigation of stem-cell-like epitypes revealed substantial tumor-specific aberrant hypermethylation in E11 and E12, and no enrichment was found for E13 owing to the paucity

of hypermethylated CpGs (Supplemental Fig. S6A). E11 showed hypermethylation enriched in enhancer regions (Supplemental Fig. S6B), as expected with prevalent *IDH1/2* mutations. However, E12 did not display a dominant mutation or TF enrichment that potentially explained aberrant hypermethylation. E12 hypermethylation was highly enriched in regions marked by polycomb repressed/poised regions. We next examined differential gene expression between E11–E13 and healthy HSPCs. We identified 52, 54, and 107 differentially expressed genes in E11, E12, and E13, respectively (greater than or equal to twofold change, adjusted $P \leq 0.01$), with 68/218 genes showing evidence of differential promoter methylation (Supplemental Table S8). Ingenuity pathway analysis comparing relative activation of upstream regulators revealed that the top results in E13 were enriched for inflammatory pathways, including TNF, IL1B, and IFNG (Supplemental Fig. S6C). We found similar results in E11 and E12, an absence in E7–E10, and variable enrichment in E1–E6 (Supplemental Fig. S6D). Hypermethylation of polycomb-marked regions is commonly observed in tumors, especially in tumors with activating mutations in signaling pathways (Gal-Yam et al. 2008; Sproul and Meehan 2013). Indeed, hematopoietic cells chronically exposed to inflammatory chemokines induces hypermethylation of polycomb regions (Spencer et al. 2017). Combining observations of DNA methylation and gene expression changes, our findings suggest that stem-cell-like epitypes that lack a dominant driver mutation may use pro-inflammatory signaling to drive AML cell proliferation and survival.

FLT3-ITD is linked to a distinct DNA methylation signature targeting STAT sites

Pro-inflammatory signaling is commonly associated with cancer and often generated by mutations in tumor cells (Balkwill and Coussens 2004). In AML, gain-of-function *FLT3*-internal tandem duplication (*FLT3*-ITD) mutations activate the JAK/STAT pathway and are associated with poor outcomes (Meshinchi and Appelbaum 2009). *FLT3*-ITD mutations were spread across several epitypes (Fig. 1C) and were not enriched in stem-cell-like epitypes (E11–E13) consistent with past studies (Figueroa et al. 2010; Döhner et al. 2017; Glass et al. 2017). Thus, we next sought to determine if there was a DNA methylation signature associated with *FLT3*-ITD indicative of pro-inflammatory signaling that was not captured in the most variable methylation signature that defined the AML epitypes. Because *FLT3*-ITD mutations were most frequent in the NPM1 supercluster, we compared *FLT3*-ITD to *FLT3* wild-type samples within E7–E10 only to avoid introducing differences specific to epitypes with less frequent *FLT3* mutations. We identified 253 probes significantly hypomethylated in *FLT3*-ITD samples (20% methylation change, FDR $Q < 0.01$). Motif enrichment analysis revealed hypomethylated regions were highly enriched for STAT family sequence motifs, with STAT5A as the top match (Fig. 5A), consistent with known activation of STAT5A (previously known as STAT5) in *FLT3*-ITD AMLs (Choudhary et al. 2007). We further selected probes in *FLT3*-ITD-associated hypomethylated regions that contained a proximal STAT motif to create a probe set of 101 CpGs that we termed the STAT hypomethylation signature (SHS) (Supplemental Table S9). We next expanded our analysis to investigate this signature across all AML samples. Hierarchical clustering in all samples identified a subset of SHS-enriched samples we designated as SHS+ AMLs (Fig. 5B). SHS positivity was not limited to E7–E10 and was found across epitypes (Fig. 5C). SHS positivity was not restricted to *FLT3*-ITD cases: 73% of SHS+ were *FLT3*-ITD, 6% had a *FLT3* mutation other than ITD (commonly single-nucleotide mutations in the kinase domain),

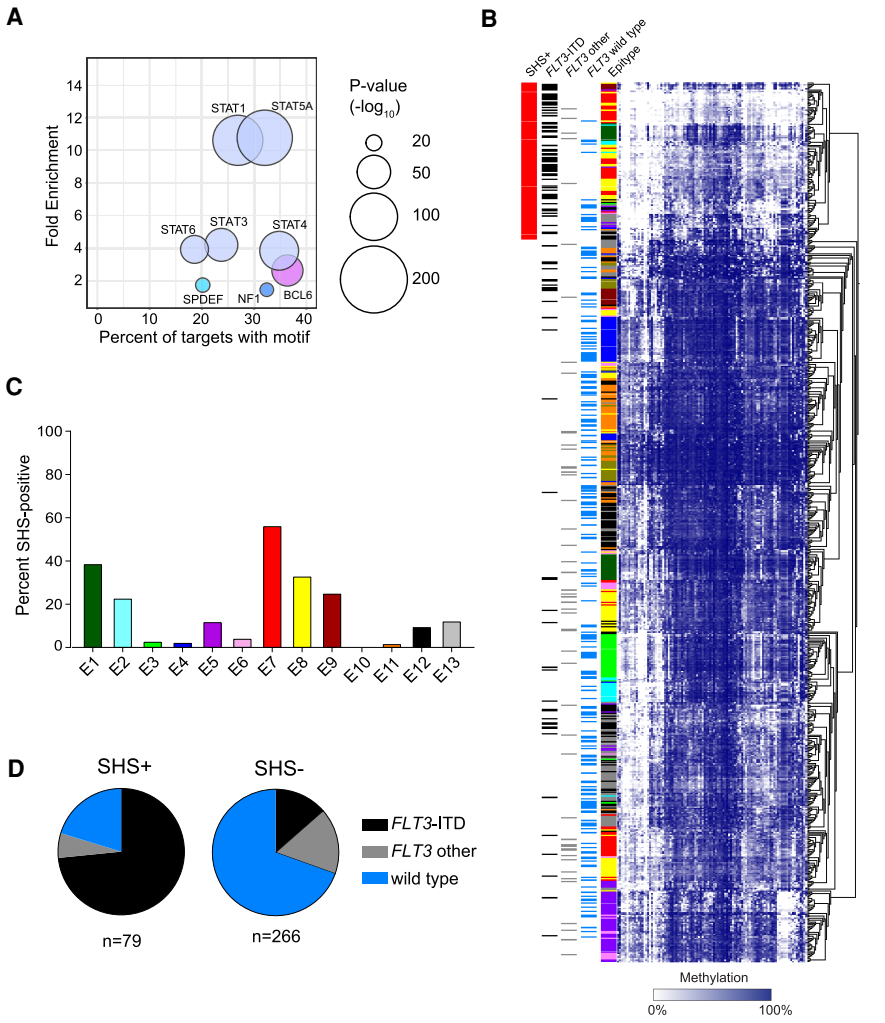

Figure 5. A hypomethylation signature involving STAT is associated with *FLT3*-ITD mutations. (*A*) Bubble scatterplot of transcription factor motif enrichment in hypomethylated regions in *FLT3*-ITD-mutated AMLs. Bubble size corresponds to the *P*-value, and color corresponds to transcription factor family. (*B*) Heatmap of the STAT hypomethylation signature with samples arranged by hierarchical clustering. (*C*) Distribution of STAT hypomethylation signature-positive (SHS+) samples across AML epitypes. (*D*) Breakdown of *FLT3* mutations in SHS+ (*left*) and SHS− (*right*) groups.

and 20% of SHS+ cases were devoid of *FLT3* mutations. We did not observe statistically significant enrichment of other pathway mutations in SHS+/FLT3 mutation-negative samples. Conversely, 70% of SHS negative cases lacked *FLT3* mutations, 14% were *FLT3*-ITD, and 17% had a non-ITD *FLT3* mutation (Fig. 5D). These results indicate that hypomethylation of STAT binding sites may occur by other aberrant biological events converging on activation of the STAT pathway.

DNA methylation patterns at relapse

Most AML patients relapse despite achieving a complete remission. Relapse can involve regrowth of the major clone at diagnosis or outgrowth of a small resistant tumor cell population that exists at baseline (Vosberg and Greif 2019). To examine if reexpanded tumor populations at relapse stably maintain or evolve novel DNA methylation patterns, we analyzed 26 paired diagnosis and relapse samples using Illumina arrays. Patients achieved a complete remission of at least 6 mo before relapse, and samples were sort-purified or had a BLAST percentage >80% to avoid methylation differences caused by impurity. Using the epityping probe set, 22/26 (85%) of patient epitypes remained stable at relapse, often producing nearly identical (overlapping) profiles following clustering (Fig. 6A). However, four of the patients fell into a different epitype at relapse than observed at diagnosis. In each of these cases, the relapse sample migrated to one of the stem-cell-like epitypes (E11–E13), suggesting the relapse tumor cell population retained a more immature differentiation state. To determine if the change in the epigenetic pattern was associated with genetic evolution, samples were sequenced for 80 commonly recurrent genetic mutations in AML (Supplemental Table S10; Eisfeld et al. 2017). All patients showing different epitypes at relapse showed evidence of clonal evolution, with the relapse sample often losing the dominant epitype mutation found in the diagnosis sample (Fig. 6B). Patients showing the same epitypes at relapse showed minimal genetic variation between time points. To validate these findings, we obtained a second cohort of 41 patients with paired diagnosis/relapse samples. We observed the same epitype at diagnosis and relapse in 39/41 (95%) patients (Supplemental Fig. S7; Supplemental Table S11). The two patients that showed a change of epitype evolved to E13 and showed genetic changes between time points, consistent with the preceding findings. We next determined if there were global methylation changes between diagnosis and relapse. We found that many of the samples showed similar epigenetic patterns at diagnosis and relapse. For the cases that did not change epitype, on average only 5% of the probes showed a difference (Δ±20% methylation) compared to 11% for cases that changed epitype (P=0.0099) (Fig. 6C). We did however observe a minority of cases that did not change epitype, yet still displayed a relatively large proportion of altered methylation. These tended to show gain- or loss-of-signaling pathway mutations, such as *RAS* or *FLT3* (Supplemental Table S10), that were not found to be strongly associated with epitypes. Methylation differences between diagnosis and relapse in these patients were generally gains and losses at subclonal (<30%) frequencies, compared to patients that displayed a change in epitype that showed a higher proportion of clonal (>30%) differences (Fig. 6D). These findings indicate that DNA methylation patterns are generally stable through therapy, likely owing to overall high stability and homogeneity of DNA methylation patterns in tumor cells enabling the clonal population that arises at relapse to be phenotypically similar to the population at diagnosis. Of the relatively few patients that showed evolution of DNA methylation patterns, all (4/4) showed reversion to an epigenetic pattern consistent with a more stem-cell-like phenotype.

Discussion

In this study, we used global DNA methylation patterns to gain a better understanding of the molecular heterogeneity observed in AML. Using unbiased clustering on a large cohort of AML samples we identified 13 distinct epitypes. Several epitypes associated with common AML genetic aberrations and different stages of myeloid development. Analysis of tumor-specific methylation changes identified potential mechanisms for tumor development in some of the less well-defined epitypes. Gene expression analysis identified epitypes displaying a stem-cell-like phenotype that was associated with overexpression of inflammatory pathways and not associated with a particular recurrent mutational pattern. We also identified a separate DNA methylation signature associated with *FLT3*-ITD that detects additional patients that use the STAT inflammatory pathway. Finally, we found that epitypes are stable between diagnosis and relapse, with the majority of cases retaining the same epitype and those that change epitype do so with evidence of clonal genetic evolution. Collectively, these findings provide evidence of AML development based upon acquisition of developmental pattern of methylation similar to what we have described in CLL. Notably, mutations identified in AML are not defining of subgroups when classification is approached in an unbiased manner.

Prior studies have used DNA methylation patterns to cluster AML patients using varying technologies and cohort sizes (Bullinger et al. 2010; Figueroa et al. 2010; Melnick 2010; Glass et al. 2017). Studies have found a general relationship of methylation-based clusters and genetic aberrations (Bullinger et al. 2010; Figueroa et al. 2010; Glass et al. 2017), and uncovered methylation gains and losses associated with these aberrations in epigenetic regulators (Glass et al. 2017). In support of these previous studies, we found tight associations with three of the epigenetic subgroups and chromosomal rearrangements t(15;17), t(8;21), and inv(16). Although past studies identified multiple epitypes associated with *CEBPA* mutations (Figueroa et al. 2010), we found *CEBPA* mutations enriched in a single subgroup. Unlike the previous studies, we did not find clear differences between the epigenetic patterns associated with mutations in *IDH1* and *IDH2* (Glass et al. 2017). Instead, we found that differences among *IDH1/2* mutant AMLs were dependent on whether an accompanying *NPM1* mutation was present or absent (E10 vs. E11, respectively). We found the *IDH2* R172 mutation exclusively in E11, consistent with previous studies showing mutual exclusivity with *NPM1* (Patel et al. 2011). Although the previous studies indicated some altered DNA methylation patterns associated with *NPM1* mutations, we found that the *NPM1* mutation has an impact on tumor-specific epigenetic patterns and is a dominant mutation in 4/13 epitypes. We found the impact that mutations in known epigenetic regulators *DNMT3A*, *IDH1/2*, or *TET2* have on DNA methylation patterns was lessened or absent when not co-occurring with *NPM1* mutations, strongly implicating a role for NPM1 in epigenetic regulation.

Because the genetic picture of AML is complex (The Cancer Genome Atlas Research Network 2013), classification of patients using DNA methylation patterns may help to describe a simplified number of phenotypes and also include patients with the same underlying biology yet lack the recurrent marker mutation. We have

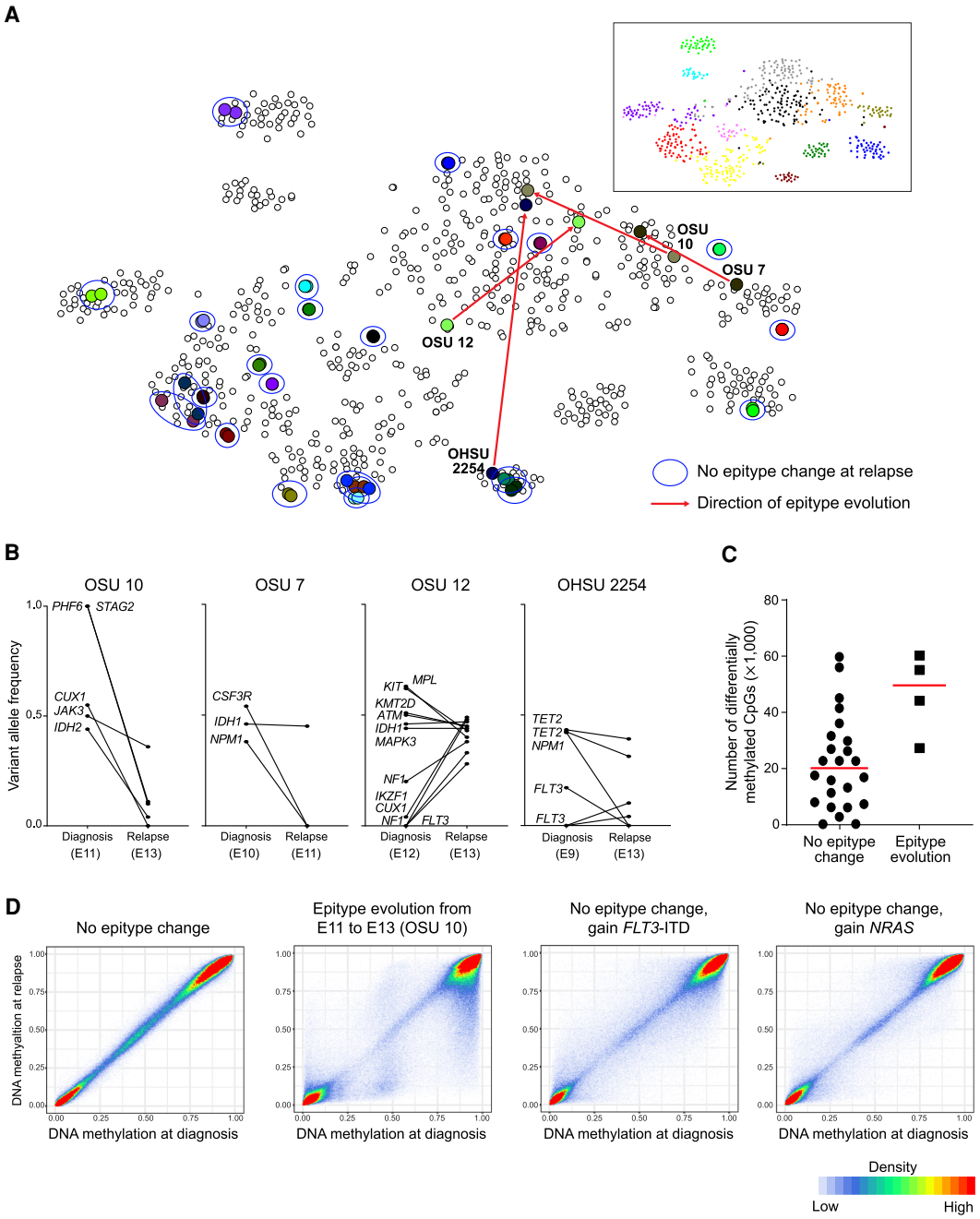

Figure 6. DNA methylation patterns are stable at relapse except in a minority of cases. (*A*) t-SNE plot of the AML epityping probe set including all AML samples along with paired diagnosis/relapse samples. The diagnosis and relapse sample (often completely overlapping) are indicated by the same color within pairs, and those pairs not changing epitype are circled in blue. Red arrows indicate pairs in which the relapse sample changed epitype. Epitypes are illustrated by standard colors in the *inset*. (*B*) Changes in mutant variant allele fraction between diagnosis and relapse in the 4/26 pairs that changed epigenetic epitype. (*C*) The number of probes that change by >20% between diagnosis and relapse; patients that showed change of epigenetic epitype are displayed separately. (*D*) Correlation of methylation values from all 426,862 probes at diagnosis and relapse in a representative sample that displayed a stable pattern, one that changed epitype, and two that remained within the same epitype but gained a signaling pathway mutation at relapse as indicated.

uncovered that epiphenocopying broadly occurs across the spectrum of recurrent mutations in AML. Indeed, phenotypic studies have revealed *CEBPA* wild-type AMLs may mimic the biology of *CEBPA* mutant AML (Wouters et al. 2007). In our study, epiphenocopying was particularly observed in epitypes we explored in more detail (E7–E13), where many patients lacked the dominant epitype mutation. A noteworthy example is E9, which displays an enrichment for *TET2* mutations yet half of the cases within this epitype have wild-type *TET2*. These epiphenocopies have attained the same epigenetic pattern through other means, such as potentially altered α-ketoglutarate metabolism (Morin et al. 2014; Raffel et al. 2017).

Mutations in epigenetic modifying genes are common in AML, but characterization of their role in disease development has been difficult. Mouse models carrying disrupted epigenetic modifiers, such as *DNMT3A*, *TET2*, and *IDH1/2*, do not alone result in frank leukemia and require serial transplantation for an overt AML phenotype to develop (Li et al. 2011; Quivoron et al. 2011; Sasaki et al. 2012; Celik et al. 2015). The presence of these mutations in asymptomatic individuals, described as clonal hematopoiesis of indeterminate potential (Steensma et al. 2015), suggests that these mutations occur early in disease development, and may loosen control of the epigenome, allowing for other events to more readily cause phenotypic reprogramming and disease development (Feinberg et al. 2016). Indeed, we observed methylation losses and gains in E7 (*NPM1+DNMT3A*) and E9,E10 (*NPM1+TET2*, *IDH1/2*) occurred in addition to changes that occurred in *NPM1*-mutated alone (E8). Although *DNMT3A*-associated global methylation loss occurred in conjunction with *NPM1* mutations, *DNMT3A* mutations did not cause hypomethylation in E13, despite being the most common genetic aberration in the epitype. This indicates that DNMT3A loss alone does not induce global DNA hypomethylation, supporting findings in DNMT3A-null mice (Guryanova et al. 2016). Because we have uncovered that E13 uses other disease-specific pathways, DNMT3A in the context of E13 may have primarily played a role in premalignant stages or has a non-DNA methylation-dependent role.

Because DNA methylation is an important factor controlling gene regulation, in addition to epitype classification, it can also convey critical aspects of the biology underling AML epitypes. We found that most AMLs retain developmental DNA methylation signatures restricted within the development of monocytes from HSPCs, consistent with results from chromatin accessibility landscapes performed on a small AML cohort versus a wide variety of hematopoietic cell types (Corces et al. 2016). AML cells with granulocytic morphologies may be depleted from samples as a result of the routine use of ficoll to purify AML cells. We found that the spread across the developmental axis is associated with the discrete epitypes we identified. Accounting for normal development enabled us to identify tumor-specific changes to the epigenome, which in turn inferred disease-specific TF activity and chromatin states. The changes in E13 relative to HSPCs may represent initial changes in AML development common to the majority of AMLs, because only failed hypomethylation was observed without aberrant (tumor-specific) changes. Most epitypes are deficient in normal development similarly involving loss of key developmental TFs activities to varying degrees. This initial step is likely the block in differentiation that can be achieved by a variety of mechanisms and is then followed by some form of proliferation as is suggested by the two-hit hypothesis of AML development (Lagunas-Rangel et al. 2017). Our data indicate that *NPM1*-mutant epitypes universally show activation of components of RUNX, AP-1, EGR, TCF, and HOX TF families implicating a variety of upstream pathway activation. E9,E10 (*NPM1+TET2*, *IDH1/2*) show loss of MEF2, ETS, and IRF sites focused on enhancer regions, potentially further repressing their normal development.

A common feature of tumor methylomes is hypermethylation of CpG islands located in gene promoter regions, referred to as the CpG island hypermethylator phenotype (CIMP) (Toyota et al. 1999). CIMP subtypes have been described within several tumor types, including colorectal, breast, brain, and gastric cancers, as well as AML (Roman-Gomez et al. 2005; Weisenberger et al. 2006; Noushmehr et al. 2010; Zouridis et al. 2012; Mack et al. 2014; Kelly et al. 2017). Recent studies have shown that CIMP frequently targets promoters that are marked by the polycomb repressive mark H3K27me3 in developmental precursors, which commonly mark poised promoters of developmentally regulated genes (Bernstein et al. 2006; Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007). In our studies, the association of aberrant hypermethylation was highly epitype-dependent, being either largely absent (E5, E7, E13), enriched at enhancers (E9–E11), or enriched in polycomb repressed and poised regions, which include CpG islands (E1–E4, E6, E8, E12). CIMP is associated with prolonged inflammation and stress (Jones and Baylin 2007), and prolonged treatment of HSPCs from healthy donors with inflammatory cytokines induces hypermethylation of polycomb repressed and poised regions (Spencer et al. 2017). Work in other cancers suggests CIMP tumors may not respond well to DNA damaging treatments but may respond better to hypomethylating agents (Mack et al. 2014). Some AML patients with high-risk genetic markers, such as complex karyotype and *TP53* mutations, have achieved favorable responses with hypomethylating agents in some patients (Santini and Ossenkoppele 2019). Of interest, E12 was associated with high-risk genetics, and poor overall survival. Elevated methylation of polycomb regions may predict favorable responses to hypomethylating agents in this epitype that responds poorly to standard chemotherapy. In addition, as we have shown activation of inflammatory pathways coincident with polycomb hypermethylation in this epitype, targeting pathways such as JAK/STAT may provide additional benefit to these patients.

Our findings show that DNA methylation is a useful approach for classifying this genetically heterogeneous, complex disease, and significantly adds to our understanding of distinct biological aspects of individual patients. We have shown that epitypes integrate the majority of highly recurrent mutations, developmental states, and other phenotypes. Our approach further incorporated patients lacking recurrent mutations based on epigenetic and phenotypic similarity (epiphenocopies). Epitypes use different pathways, several of which involve activation of inflammatory pathways and are associated with poor survival. The use of epitypes as a biomarker in AML is further supported by the stability of the methylation patterns throughout disease course in the vast majority of patients. Future development of a targeted approach for classification of individual patients will be vital to unlock potential clinical utility of assessing epitypes. Coordinated evolution of epitype and genetic markers may identify patients that may benefit from a change in therapy, including hypomethylating and novel agents.

Methods

AML patients and Illumina DNA methylation array data

AML patient samples were collected as part of the Beat AML study (Tyner et al. 2018), and genomic DNA from 226 bone marrow samples was obtained at diagnosis along with 13 matched relapse samples. DNA was obtained from an additional 27 patients at diagnosis with t(v;11) rearrangements and 22 patients sampled at diagnosis and relapse from the Leukemia Tissue Bank Shared Resource at the Ohio State University Comprehensive Cancer Center. Mononuclear cells from AML samples were isolated using density gradient separation. Samples with low variant allele frequency (VAF) of recurrent mutations in myeloid malignancies suggestive of low tumor cell purity were excluded. Leukemic cells from low (<80%) BLAST count relapse samples were further purified using fluorescence-assisted cell sorting (FACS) (BD Biosciences). HSPC

(CD34$^+$), monocyte (CD14$^+$) populations were isolated using FACS from bone marrow and peripheral blood. Macrophages were obtained by plating sorted monocyte populations for 14 d. All patients and donors provided informed consent according to the Declaration of Helsinki. Genomic DNA was isolated using column-based preparation (Qiagen). DNA (500 ng) was bisulfite converted using the EZ DNA Methylation Gold Kit (Zymo Research). The Infinium methylationEPIC assay was carried out following standard protocol (Illumina) at the Molecular Genetics Laboratory at the Cincinnati Children's Hospital. Illumina 450K Human Methylation Array raw data files for additional AML samples and sorted healthy populations were obtained from previously published studies (Kulis et al. 2012; Reinius et al. 2012; The Cancer Genome Atlas Research Network 2013; Schmutz et al. 2013; Leonard et al. 2014; Jung et al. 2015; Ferreira et al. 2016; Qu et al. 2017). For the diagnosis/relapse validation cohort, gene mutation data for 33 patients at diagnosis, remission, and relapse were obtained from Greif et al. (2018) along with Illumina methylation array beta values composing the epityping signature. Patients were excluded from the study that did not show high frequency tumor-specific mutations (VAF > 0.3) at both diagnosis and relapse. The absence of tumor-specific mutations at remission were also required to indicate clearance of tumor cells following treatment. Data from the Infinium methylationEPIC Array and Illumina HumanMethylation450 Array were normalized by the beta mixture quantile (BMIQ) method (Teschendorff et al. 2013) using the RnBeads analysis software package (Müller et al. 2019). Only intersecting probes on both platforms were included, and probes targeting sex chromosomes, non-CpG sites, and single-nucleotide polymorphisms were removed, resulting in a final probe set of 426,862 probes. Heatmaps, t-SNE plots, and principal component analysis plots of combined data were visualized using the Qlucore Omics Explorer software.

DNA methylation analysis

For clustering analysis, *k*-medoids-based clustering was used because of the uneven levels of similarity within clusters. An Auer-Gervini plot was used to identify the minimum number of dimensions as the first long step, determined by twice the length of the average (Wang et al. 2018). *k* was set using silhouette analysis (Rousseeuw 1987).

Transcription factor sequence motif enrichment for known motifs was performed using HOMER software (Heinz et al. 2010). Windows containing 100 bp of sequence upstream of and downstream from selected probes were searched against a background assembled from the remaining probes that were adjusted for GC and CpG content as well as a similar methylation distribution in HSPCs. Motifs with a high degree of similarity were replaced with a single consensus motif. Chromatin states were defined using the standard 15-state model previously described using the ChromHMM algorithm (Ernst et al. 2011). Chromatin states were defined in three HSPC samples using data available through the Roadmap Epigenomics Project (Roadmap Epigenomics Consortium et al. 2015). Enrichment analysis was performed using the EpiAnnotator R version 4.02 package (Pageaud et al. 2018; R Core Team 2021).

To generate a DNA methylation signature that encompasses normal hematopoietic development we used published methylation array data from sorted healthy cell populations. We determined the probes significantly differentially methylated between each cell population and HSPCs (20% methylation change, FDR $Q < 0.01$). These individual probe lists were then combined to create the hematopoietic development signature ($n = 28{,}361$). To generate a normal myeloid development signature, we used the 5000 most variable probes between HSPCs ($n = 14$) and monocytes ($n = 15$). To identify tumor-specific differences occurring outside of normal development, we compared changes in individual AML epitypes versus the change that normally occurs in monocyte differentiation using HSPCs as a fixed reference across all analyzed probes. The methylation values of all probes were averaged within each AML and normal subtype in scatterplots. The probes that diverged from the expected normal development value greater than a mean of 30% and maintained a false discovery rate (FDR) of $Q < 0.05$ when considering individual tumor samples were retained and classified as tumor specific.

Gene expression analysis

Differential gene expression was performed using DESeq2 (Love et al. 2014) on raw counts were obtained from the Beat AML Consortium (Tyner et al. 2018). Samples within each epitype were treated as biological replicates and compared to HSPCs. Differentially expressed genes were defined as greater than twofold change and FDR $Q < 0.01$ were used. The upstream regulator tool in the Ingenuity Pathway Analysis software was used to interpret the results. LSC17 score was calculated using the 17 genes weighted by regression coefficients as reported in Ng et al. (2016). Beat AML expression value was calculated from RPKM-normalized RNA-seq data (Tyner et al. 2018), and TCGA was calculated from Affymetrix U133 Plus 2 platform (The Cancer Genome Atlas Research Network 2013). The stem cell signature scores of Gal et al. (2006) and Gentles et al. (2010) were calculated by the median expression of the genes in each gene set within each sample using RPKM-normalized RNA-seq data. ANOVA was used to determine significant differences between epitypes, followed by group-specific *t*-tests adjusting for multiple comparisons using the Bonferroni method.

Gene mutation analysis

Annotation of genetic mutations and other aberrations were obtained from each respective study where available ($n = 511$ samples total) (The Cancer Genome Atlas Research Network 2013; Schmutz et al. 2013; Jung et al. 2015; Eisfeld et al. 2017; Tyner et al. 2018). For the analysis of paired diagnosis and relapse samples from Ohio State University, a panel of 80 genes including common recurrent AML mutations was targeted using a capture oligo-based approach followed by sequencing on the MiSeq platform (Illumina) as previously described in Eisfeld et al. (2017). For paired diagnosis and relapse samples from the Beat AML project, mutation data were obtained from published whole exome data (Tyner et al. 2018).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE159907.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank all the patients who contributed to this study. We also thank the Beat AML Consortium and affiliated members for collaboration and support of the project. This work was supported by the

Ohio State University Comprehensive Cancer Center (OSUCCC). We thank Stephanie Monzon, Brenna Hott, and Emily Liston from the Cincinnati Children's Hospital for their expertise with Illumina arrays. We thank Dr. David Lucas and Chris Manring for assistance to obtain samples and data from the OSUCCC Leukemia Tissue Bank Shared Resource, supported by the National Cancer Institute (NCI) P30 CA016058. B.J.D. is supported by Howard Hughes Medical Institute and the National Institutes of Health/NCI U54 CA224019 and U01 CA217862. J.W.T. received grants from the Mark Foundation for Cancer Research, the Silver Family Foundation, and the National Cancer Institute (1R01CA183947, 1U01CA217862, 1U54CA224019). S.V. was supported by the Deutsche José Carreras Leukämie-Stiftung. L.B. and R.C. were supported by the German Cancer Aid, DKH 110530. C.C.O. is supported by the Gabrielle's Angel Foundation for Cancer Research.

Author contributions: B.G., A.C., and Y.-Z.W. performed laboratory experiments; B.G., M.W., A.C., J.S.B., K.R.C., and C.C.O. performed data analysis and interpretation; A.R.S., A.-K.E., M.S., S.V., P.A.G., R.C., L.B., B.M.-B., R.G., C.D.B., B.J.D., and J.W.T. contributed reagents, materials, and/or data; B.G., C.D.B., J.W.T., J.C.B., and C.C.O. wrote the manuscript; all coauthors reviewed and edited the manuscript.

References

Amadori S, Venditti A, Del Poeta G, Stasi R, Buccisano F, Bruno A, Tamburini A, Cox MC, Maffei L, Aronica G, et al. 1996. Minimally differentiated acute myeloid leukemia (AML-MO): a distinct clinico-biologic entity with poor prognosis. *Ann Hematol* **72:** 208–215. doi:10.1007/s002770050162

Auer P, Gervini D. 2008. Choosing principal components: a new graphical method based on Bayesian model selection. *Commun Stat Simul Comput* **37:** 962–977. doi:10.1080/03610910701855005

Balkwill F, Coussens LM. 2004. Cancer: an inflammatory link. *Nature* **431:** 405–406. doi:10.1038/431405a

Barbaric D, Alonzo TA, Gerbing RB, Meshinchi S, Heerema NA, Barnard DR, Lange BJ, Woods WG, Arceci RJ, Smith FO. 2007. Minimally differentiated acute myeloid leukemia (FAB AML-M0) is associated with an adverse outcome in children: a report from the children's oncology group, studies CCG-2891 and CCG-2961. *Blood* **109:** 2314–2321. doi:10.1182/blood-2005-11-025536

Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DAG, Gralnick HR, Sultan C. 1982. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol* **51:** 189–199. doi:10.1111/j.1365-2141.1982.tb08475.x

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326. doi:10.1016/j.cell.2006.02.041

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, et al. 2011. High density DNA methylation array with single CpG site resolution. *Genomics* **98:** 288–295. doi:10.1016/j.ygeno.2011.07.007

Bradstock KF, Kirk J, Grimsley PG, Kabral A, Hughes WG. 1989. Unusual immunophenotypes in acute leukaemias: incidence and clinical correlations. *Br J Haematol* **72:** 512–518. doi:10.1111/j.1365-2141.1989.tb04315.x

Bullinger L, Ehrich M, Döhner K, Schlenk RF, Döhner H, Nelson MR, Van Den Boom D. 2010. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood* **115:** 636–642. doi:10.1182/blood-2009-03-211003

The Cancer Genome Atlas Research Network. 2013. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368:** 2059–2074. doi:10.1056/NEJMoa1301689

Celik H, Mallaney C, Kothari A, Ostrander EL, Eultgen E, Martens A, Miller CA, Hundal J, Klco JM, Challen GA. 2015. Hematopoiesis and stem cells: enforced differentiation of Dnmt3a-null bone marrow leads to failure with c-Kit mutations driving leukemic transformation. *Blood* **125:** 619–628. doi:10.1182/blood-2014-08-594564

Choudhary C, Brandts C, Schwable J, Tickenbrock L, Sargin B, Ueker A, Böhmer FD, Berdel WE, Müller-Tidow C, Serve H. 2007. Activation mechanisms of STAT5 by oncogenic Flt3-ITD. *Blood* **110:** 370–374. doi:10.1182/blood-2006-05-024018

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, et al. 2016. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* **48:** 1193–1203. doi:10.1038/ng.3646

De Braekeleer E, Douet-Guilbert N, De Braekeleer M. 2014. *RARA* fusion genes in acute promyelocytic leukemia: a review. *Expert Rev Hematol* **7:** 347–357. doi:10.1586/17474086.2014.903794

Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, et al. 2010. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115:** 453–474. doi:10.1182/blood-2009-07-235358

Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, Dombret H, Ebert BL, Fenaux P, Larson RA, et al. 2017. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129:** 424–447. doi:10.1182/blood-2016-08-733196

Eisfeld AK, Mrózek K, Kohlschmidt J, Nicolet D, Orwick S, Walker CJ, Kroll KW, Blachly JS, Carroll AJ, Kolitz JE, et al. 2017. The mutational oncoprint of recurrent cytogenetic abnormalities in adult patients with de novo acute myeloid leukemia. *Leukemia* **31:** 2211–2218. doi:10.1038/leu.2017.86

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28:** 817–825. doi:10.1038/nbt.1662

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473:** 43–49. doi:10.1038/nature09906

Feinberg AP, Koldobskiy MA, Göndör A. 2016. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* **17:** 284–299. doi:10.1038/nrg.2016.13

Ferreira HJ, Heyn H, Vizoso M, Moutinho C, Vidal E, Gomez A, Martínez-Cardús A, Simó-Riudalbas L, Moran S, Jost E, et al. 2016. DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia. *Oncogene* **35:** 3079–3082. doi:10.1038/onc.2015.359

Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, Schifano E, Booth J, van Putten W, Skrabanek L, et al. 2010. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17:** 13–27. doi:10.1016/j.ccr.2009.11.020

Gal H, Amariglio N, Trakhtenbrot L, Jacob-Hirsh J, Margalit O, Avigdor A, Nagler A, Tavor S, Ein-Dor L, Lapidot T, et al. 2006. Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. *Leukemia* **20:** 2147–2154. doi:10.1038/sj.leu.2404401

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A. 2008. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci* **105:** 12979–12984. doi:10.1073/pnas.0806437105

Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. 2010. A leukemic stem cell gene expression signature is associated with clinical outcomes in acute myeloid leukemia. *JAMA* **304:** 2706–2715. doi:10.1001/jama.2010.1862

Glass JL, Hassane D, Wouters BJ, Kunimoto H, Avellino R, Garrett-Bakelman FE, Guryanova OA, Bowman R, Redlich S, Intlekofer AM, et al. 2017. Epigenetic identity in AML depends on disruption of nonpromoter regulatory elements and is affected by antagonistic effects of mutations in epigenetic modifiers. *Cancer Discov* **7:** 868–883. doi:10.1158/2159-8290.CD-16-1032

Greif PA, Hartmann L, Vosberg S, Stief SM, Mattes R, Hellmann I, Metzeler KH, Herold T, Bamopoulos SA, Kerbs P, et al. 2018. Evolution of cytogenetically normal acute myeloid leukemia during therapy and relapse: an exome sequencing study of 50 patients. *Clin Cancer Res* **24:** 1716–1726. doi:10.1158/1078-0432.CCR-17-2344

Griffin JD, Mayer RJ, Weinstein HJ, Rosenthal DS, Coral FS, Beveridge RP, Schlossman SF. 1983. Surface marker analysis of acute myeloblastic leukemia: identification of differentiation-associated phenotypes. *Blood* **62:** 557–563. doi:10.1182/blood.V62.3.557.557

Guryanova OA, Shank K, Spitzer B, Luciani L, Koche RP, Garrett-Bakelman FE, Ganzel C, Durham BH, Mohanty A, Hoermann G, et al. 2016. *DNMT3A* mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat Med* **22:** 1488–1495. doi:10.1038/nm.4210

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Hovestadt V, Jones DTW, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz HJ, et al. 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510:** 537–541. doi:10.1038/nature13268

Jones PA, Baylin SB. 2007. The epigenomics of cancer. *Cell* **128:** 683–692. doi:10.1016/j.cell.2007.01.029

Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. 2015. An LSC epigenetic signature is largely mutation independent and implicates the *HOXA* cluster in AML pathogenesis. *Nat Commun* **6:** 8489. doi:10.1038/ncomms9489

Kelly AD, Kroeger H, Yamazaki J, Taby R, Neumann F, Yu S, Lee JT, Patel B, Li Y, He R, et al. 2017. A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia* **31:** 2011–2019. doi:10.1038/leu.2017.12

Koch A, Joosten SC, Feng Z, De Ruijter TC, Draht MX, Melotte V, Smits KM, Veeck J, Herman JG, Neste LV, et al. 2018. Analysis of DNA methylation in cancer: location revisited. *Nat Rev Clin Oncol* **15:** 459–466. doi:10.1038/s41571-018-0004-4

Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, Martínez-Trillos A, Castellano G, Brun-Heath I, Pinyol M, et al. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet* **44:** 1236–1242. doi:10.1038/ng.2443

Lagunas-Rangel FA, Chávez-Valencia V, Gómez-Guijosa MÁ, Cortes-Penagos C. 2017. Acute myeloid leukemia-genetic alterations and their clinical prognosis. *Int J Hematol Stem Cell Res* **11:** 328–339.

Leonard SM, Perry T, Woodman CB, Kearns P. 2014. Sequential treatment with cytarabine and decitabine has an increased anti-leukemia effect compared to cytarabine alone in xenograft models of childhood acute myeloid leukemia. *PLoS One* **9:** e87475. doi:10.1371/journal.pone.0087475

Li Z, Cai X, Cai CL, Wang J, Zhang W, Petersen BE, Yang FC, Xu M. 2011. Deletion of *Tet2* in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118:** 4509–4518. doi:10.1182/blood-2010-12-325241

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Lowenberg B, Downing JR, Burnett A. 1999. Acute myeloid leukemia. *N Engl J Med* **341:** 1051–1062. doi:10.1056/NEJM199909303411407

Macedo A, Orfão A, Vidriales MB, López-Berges MC, Valverde B, González M, Caballero MD, Ramos F, Martínez M, Fernández-Calvo J, et al. 1995. Characterization of aberrant phenotypes in acute myeloblastic leukemia. *Ann Hematol* **70:** 189–194. doi:10.1007/BF01700374

Mack SC, Witt H, Piro RM, Gu L, Zuyderduyn S, Stütz AM, Wang X, Gallo M, Garzia L, Zayne K, et al. 2014. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506:** 445–450. doi:10.1038/nature13108

Matutes E, Morilla R, Farahat N, Carbonell F, Swansbury J, Dyer M, Catovsky D. 1997. Definition of acute biphenotypic leukemia. *Haematologica* **82:** 64–66.

Melnick AM. 2010. Epigenetics in AML. *Best Pract Res Clin Haematol* **23:** 463–468. doi:10.1016/j.beha.2010.09.017

Meshinchi S, Appelbaum FR. 2009. Structural and functional alterations of FLT3 in acute myeloid leukemia. *Clin Cancer Res* **15:** 4263–4269. doi:10.1158/1078-0432.CCR-08-1123

Morin A, Letouzé E, Gimenez-Roqueplo AP, Favier J. 2014. Oncometabolites-driven tumorigenesis: from genetics to targeted therapy. *Int J Cancer* **135:** 2237–2248. doi:10.1002/ijc.29080

Mrózek K, Radmacher MD, Bloomfield CD, Marcucci G. 2009. Molecular signatures in acute myeloid leukemia. *Curr Opin Hematol* **16:** 64–69. doi:10.1097/MOH.0b013e3283257b42

Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, Bock C. 2019. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol* **20:** 55. doi:10.1186/s13059-019-1664-9

Ng K, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, Arruda A, Popescu A, Gupta V, Schimmer AD, et al. 2016. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540:** 433–437. doi:10.1038/nature20598

Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, et al. 2010. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17:** 510–522. doi:10.1016/j.ccr.2010.03.017

Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, Wang Q, Imbusch CD, Serva A, Koser SD, et al. 2016. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* **48:** 253–264. doi:10.1038/ng.3488

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, et al. 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39:** 237–242. doi:10.1038/ng1972

Pabst T, Mueller BU. 2007. Transcriptional dysregulation during myeloid transformation in AML. *Oncogene* **26:** 6829–6837. doi:10.1038/sj.onc.1210765

Pageaud Y, Plass C, Assenov Y. 2018. Genome analysis enrichment analysis with epiAnnotator. *Bioinformatics* **34:** 1781–1783. doi:10.1093/bioinformatics/bty007

Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, et al. 2016. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* **374:** 2209–2221. doi:10.1056/NEJMoa1516192

Patel KP, Ravandi F, Ma D, Paladugu A, Barkoh BA, Medeiros LJ, Luthra R. 2011. Acute myeloid leukemia with *IDH1* or *IDH2* mutation. *Am J Clin Pathol* **135:** 35–45. doi:10.1309/AJCPD7NR2RMNQDVF

Qu Y, Siggens L, Cordeddu L, Gaidzik VI, Karlsson K, Bullinger L, Döhner K, Ekwall K, Lehmann S, Lennartsson A. 2017. Cancer-specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood* **129:** e13–e25. doi:10.1182/blood-2016-07-726877

Quivoron C, Couronné L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, Do Cruzeiro M, Delhommeau F, Arnulf B, Stern MH, et al. 2011. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and Is a recurrent event during human lymphomagenesis. *Cancer Cell* **20:** 25–38. doi:10.1016/j.ccr.2011.06.003

Raffel S, Falcone M, Kneisel N, Hansson J, Wang W, Lutz C, Bullinger L, Poschet G, Nonnenmacher Y, Barnert A, et al. 2017. BCAT1 restricts αkG levels in AML stem cells leading to IDH^{mut}-like DNA hypermethylation. *Nature* **551:** 384–388. doi:10.1038/nature24294

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, Söderhäll C, Scheynius A, Kere J. 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* **7:** e41361. doi:10.1371/journal.pone.0041361

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330. doi:10.1038/nature14248

Roman-Gomez J, Jimenez-Velasco A, Agirre X, Prosper F, Heiniger A, Torres A. 2005. Lack of CpG island methylator phenotype defines a clinical subtype of T-cell acute lymphoblastic leukemia associated with good prognosis. *J Clin Oncol* **23:** 7043–7049. doi:10.1200/JCO.2005.01.4944

Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* **20:** 53–65. doi:10.1016/0377-0427(87)90125-7

Santini V, Ossenkoppele GJ. 2019. Hypomethylating agents in the treatment of acute myeloid leukemia: a guide to optimal use. *Crit Rev Oncol Hematol* **140:** 1–7. doi:10.1016/j.critrevonc.2019.05.013

Sasaki M, Knobbe CB, Munger JC, Lind EF, Brenner D, Brüstle A, Harris IS, Holmes R, Wakeham A, Haight J, et al. 2012. IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488:** 656–659. doi:10.1038/nature11323

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39:** 232–236. doi:10.1038/ng1950

Schmutz M, Zucknick M, Schlenk RF, Döhner K, Döhner H, Plass C, Bullinger L, Claus R. 2013. Differential DNA methylation predicts response to combined treatment regimens with a DNA methyltransferase inhibitor in acute myeloid leukemia (AML). *Blood* **122:** 2539. doi:10.1182/blood.V122.21.2539.2539

Schüler A, Schwieger M, Engelmann A, Weber K, Horn S, Müller U, Arnold MA, Olson EN, Stocking C. 2008. The MADS transcription factor Mef2c is a pivotal modulator of myeloid cell fate. *Blood* **111:** 4532–4541. doi:10.1182/blood-2007-10-116343

Scourzic L, Mouly E, Bernard OA. 2015. TET proteins and the control of cytosine demethylation in cancer. *Genome Med* **7:** 9–16. doi:10.1186/s13073-015-0134-6

Speck NA, Gilliland DG. 2002. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer* **2:** 502–513. doi:10.1038/nrc840

Spencer DH, Young MA, Lamprecht TL, Helton NM, Fulton R, O'Laughlin M, Fronick C, Magrini V, Demeter RT, Miller CA, et al. 2015. Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia* **29:** 1279–1289. doi:10.1038/leu.2015.6

Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, Fronick CC, O'Laughlin M, Heath SE, Shinawi M, et al. 2017. Cpg island hypermethylation mediated by DNMT3A is a

consequence of AML progression. *Cell* **168:** 801–816.e13. doi:10.1016/j.cell.2017.01.021

Sproul D, Meehan RR. 2013. Genomic insights into cancer-associated aberrant CpG island hypermethylation. *Brief Funct Genomics* **12:** 174–190. doi:10.1093/bfgp/els063

Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, Ebert BL. 2015. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126:** 9–16. doi:10.1182/blood-2015-03-631747

Tenen DG. 2003. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3:** 89–101. doi:10.1038/nrc989

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. 2013. A β-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29:** 189–196. doi:10.1093/bioinformatics/bts680

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa J-PJ. 1999. Cpg island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci* **96:** 8681–8686. doi:10.1073/pnas.96.15.8681

Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, Long N, Schultz AR, Traer E, Abel M, et al. 2018. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562:** 526–531. doi:10.1038/s41586-018-0623-z

Vosberg S, Greif PA. 2019. Clonal evolution of acute myeloid leukemia from diagnosis to relapse. *Genes Chromosomes Cancer* **58:** 839–849. doi:10.1002/gcc.22806

Wang M, Kornblau SM, Coombes KR. 2018. Decomposing the apoptosis pathway into biologically interpretable principal components. *Cancer Inform* **17:** 1176935118771082. doi:10.1177/1176935118771082

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, et al. 2006. Cpg island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with *BRAF* mutation in colorectal cancer. *Nat Genet* **38:** 787–793. doi:10.1038/ng1834

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, et al. 2007. Epigenetic stem cell signature in cancer. *Nat Genet* **39:** 157–158. doi:10.1038/ng1941

Will B, Vogler TO, Narayanagari S, Bartholdy B, Todorova TI, da Silva Ferreira M, Chen J, Yu Y, Mayer J, Barreyro L, et al. 2015. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat Med* **21:** 1172–1181. doi:10.1038/nm.3936

Winters AC, Bernt KM. 2017. MLL-rearranged leukemias—an update on science and clinical approaches. *Front Pediatr* **5:** 11–13. doi:10.3389/fped.2017.00004

Wouters BJ, Jordà MA, Keeshan K, Louwers I, Erpelinck-Verschueren CAJ, Tielemans D, Langerak AW, He Y, Yashiro-Ohtani Y, Zhang P, et al. 2007. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*. *Blood* **110:** 3706–3714. doi:10.1182/blood-2007-02-073486

Zouridis H, Deng N, Ivanova T, Zhu Y, Wong B, Huang D, Wu YH, Wu Y, Tan IB, Liem N, et al. 2012. Methylation subtypes and large-scale epigenetic alterations in gastric cancer. *Sci Transl Med* **4:** 156ra140. doi:10.1126/scitranslmed.3004504

Received July 23, 2020; accepted in revised form March 9, 2021.

DNA methylation epitypes highlight underlying developmental and disease pathways in acute myeloid leukemia

Brian Giacopelli, Min Wang, Ada Cleary, et al.

Genome Res. 2021 31: 747-761 originally published online March 11, 2021
Access the most recent version at doi:10.1101/gr.269233.120