

RESEARCH ARTICLE

Somatic Mutations Drive Specific, but Reversible, Epigenetic Heterogeneity States in AML



Sheng Li^{1,2,3,4}, Xiaowen Chen¹, Jiahui Wang¹, Cem Meydan⁵, Jacob L. Glass⁶, Alan H. Shih⁷, Ruud Delwel⁸, Ross L. Levine⁹, Christopher E. Mason⁵, and Ari M. Melnick¹⁰

ABSTRACT

Epigenetic allele diversity is linked to inferior prognosis in acute myeloid leukemia (AML). However, the source of epiallele heterogeneity in AML is unknown. Herein we analyzed epiallele diversity in a genetically and clinically annotated AML cohort. Notably, AML driver mutations linked to transcription factors and favorable outcome are associated with epigenetic destabilization in a defined set of susceptible loci. In contrast, AML subtypes linked to inferior prognosis manifest greater abundance and highly stochastic epiallele patterning. We report an epiallele outcome classifier supporting the link between epigenetic diversity and treatment failure. Mouse models with *TET2* or *IDH2* mutations show that epiallele diversity is especially strongly induced by *IDH* mutations, precedes transformation to AML, and is enhanced by cooperation between somatic mutations. Furthermore, epiallele complexity was partially reversed by epigenetic therapies in AML driven by *TET2/IDH2*, suggesting that epigenetic therapy might function in part by reducing population complexity and fitness of AMLs.

SIGNIFICANCE: We show for the first time that epigenetic clonality is directly linked to specific mutations and that epigenetic allele diversity precedes and potentially contributes to malignant transformation. Furthermore, epigenetic clonality is reversible with epigenetic therapy agents.

INTRODUCTION

Acute myeloid leukemia (AML) is a highly aggressive cancer that arises from hematopoietic stem and progenitor cells. AML remains difficult to treat mainly due to failure to eradicate residual leukemia stem cells, eventually leading to relapse and disease progression (1). Genetic heterogeneity and clonal diversity have been suggested to contribute, at least in part, to treatment failure, as they provide alternative trajectories for cells to escape therapy (2). Genome sequencing has revealed generally low somatic mutation burden in AML compared with most other cancers (2–4). In contrast, aberrant epigenetic patterning is common and has emerged as a hallmark of AML (5–10). Cytosine methylation profiling

studies show that AMLs can be classified into epigenetically defined subtypes with distinct biological features and clinical outcomes, only some of which are defined by particular somatic mutations (6, 7, 10). Even though there are relatively few somatic mutations associated with AML, those mutations can have a major impact when they affect epigenetic programming or when they interact synergistically (3, 11, 12). Along these lines, a genetic theme in AML is recurrent somatic mutation of transcription factors or genomic rearrangement of transcription factors and epigenetic regulators (3, 4). For example, loss of function of *TET2* dioxygenase or of *DNMT3A*, and gain-of-function mutations of the gene encoding *IDH1/2* directly mediate profound perturbation of cytosine-methylation patterning and gene expression in AML (5, 7, 13). Synergistic disruption of epigenetic and transcriptional programming can arise from cooperative effects between mutations such as those affecting *TET2* and *FLT3* (11, 12).

In addition to these effects on global DNA methylation patterning, it has more recently been appreciated that cytosine methylation patterning can differ among AML cells within a given patient, resulting in an epigenetically heterogeneous population of leukemia cells (8, 9). Epigenetic diversity appears to provide tumors with an additional layer of population fitness conceptually similar to the case of genetic diversity, and was accordingly noted to associate with unfavorable outcomes in diffuse large B-cell lymphomas (14) and later in other tumor types including AML (9, 15, 16). Recent efforts to refine this concept have given rise to the concept of “epigenetic allele (epiallele),” referring to the pattern of variation in methylation status among CpGs present in discrete sets that can be tracked as a unit by virtue of the constituent CpGs being located adjacent to each other (9, 15–17).

Epialleles, though not identical in concept to genetic clones, provide a basis for measuring population diversity among individual cells in a tumor in a clinically relevant manner (9, 15, 17). Several methods for identifying epiallele

¹The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut.

²The Jackson Laboratory Cancer Center, Bar Harbor, Maine. ³The Department of Genetics and Genomic Sciences, The University of Connecticut Health Center, Farmington, Connecticut. ⁴Department of Computer Science and Engineering, University of Connecticut, Storrs, Connecticut.

⁵Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York.

⁶Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. ⁷Center for Hematologic Malignancies, Memorial Sloan Kettering Cancer Center, New York, New York. ⁸Department of Hematology, Erasmus University Medical Center and Onco Institute, Rotterdam, the Netherlands. ⁹Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York. ¹⁰Division of Hematology/Oncology, Weill Cornell Medicine, New York, New York.

Note: Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

S. Li and X. Chen contributed equally to this article.

Corresponding Authors: Ari M. Melnick, Weill Cornell Medicine, 413 E 69th Street, BB-1430, New York, NY 10021. Phone: 212-746-7643; Fax: 212-746-8866; E-mail: amm2014@med.cornell.edu; and Sheng Li, The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032. Phone: 860-837-2119; E-mail: sheng.li@jax.org

Cancer Discov 2020;10:1934–49

doi: 10.1158/2159-8290.CD-19-0897

©2020 American Association for Cancer Research.

diversity have been developed that inform on either the epigenetic-allele state of any given population of cells (15, 17) or the degree to which epiallele patterns shift at different time points during tumor evolution (8). The latter approach has been used to demonstrate that genetic clonality and epiallele diversity are independent of each other in AML, indicating that neither one is predictive of the other (9). Importantly, epialleles appear to have functional consequences, because those located in gene promoters have been shown to associate with transcriptional diversity, ranging from full silencing to moderately high-level transcription (15), presumably allowing individual cells within a tumor to sample multiple different transcriptional combinatorial states in a way that fosters population fitness.

Collectively, these studies point to epigenetic heterogeneity as a fundamental property of tumors, which leads naturally to the question of what mechanism drives this phenomenon. Epigenetic heterogeneity varies widely among patients, indicating that this is not a general property of proliferating tumors but rather must be facilitated by particular pathogenic influences. As a related concept, it is reasonable to question whether epigenetic heterogeneity is a stochastic feature that is secondary to the transformed state or whether it is a consequence of the effects of specific genetic lesions. Herein, we explored the epigenetic profiles of a cohort of clinically and genetically annotated patients with AML and mouse models to investigate whether any of the canonical somatic mutations occurring in AML might be linked to development of epiallele diversity.

RESULTS

Genetically and Epigenetically Defined AML Subtypes Associate with Specific Levels of Epigenetic Heterogeneity

To test whether genetic mutations are associated with epiallele diversity, we analyzed cytosine methylation profiles obtained by performing enhanced reduced representation bisulfite sequencing (ERRBS) in a cohort of 119 patients with primary AML curated to reflect many of the common AML genetic lesions (18). Specifically, patients were selected on the basis of presence of *t(8;21)*, *t(15;17)*, *t(v;11q23)*, *inv(16)*, *Del(5/7q)*, *TET2*, *EVII*, *CEBPA* double mutations (*CEBPA-dm*), or mutations in *DNMT3A*, *IDH1/2*, or *NPM1*, or concurrence of mutation in *DNMT3A* and *IDH1/2*, resulting in a total of 12 genetically defined subtypes of patients. Also included were a set of patients previously defined by a hypermethylated signature, silencing of the *CEBPA* gene (*CEBPA-sil*), and a highly unfavorably clinical outcome that is not explained by the presence of any particular somatic mutation (6, 19). We used 14 normal bone marrow (NBM) CD34⁺ hematopoietic stem and progenitor cell (HSPC) samples as controls (8, 9). Epiallele diversity was determined using three orthogonal computational epigenomics approaches: the proportion of disordered reads (PDR; ref. 15), epipolymorphism (17), and Shannon entropy (20). We also measured the number of loci with epiallele shifting per million loci sequenced (EPM; ref. 8) as well as EPM excluding uniformly differentially methylated regions (DMR). These approaches were used to calculate DNA methylation allele diversity among

the twelve genetically defined and one epigenetically defined subtypes of patients mentioned above.

Notably, AMLs with particular genetic lesions manifested characteristic and robust genomic distributions and abundance of epialleles (PDR: $P = 2.26 \times 10^{-9}$, Fig. 1A; EPM: $P = 5.137 \times 10^{-10}$, Fig. 1B; epipolymorphism, Supplementary Fig. S1A; Shannon entropy, Supplementary Fig. S1B; no DMR EPM, Supplementary Fig. S1C: $P < 0.001$, Kruskal-Wallis rank sum test). Of note, we observed that some of the favorable-risk genetic lesions (ref. 4; *t(15;17)* and *CEBPA-dm*) were linked to high levels of epiallele diversity, whereas others, including *inv(16)* and *t(8;21)*, manifested low diversity levels. Overall, global epiallele diversity is linked at least in part to underlying genetic lesions and appears to not always correlate with degree of clinical risk. Nonetheless, those patients with higher epiallele diversity had a lower chance of entering complete remission ($P = 0.02$ measured by PDR, Fig. 1C; $P = 0.023$ measured by epipolymorphism, Supplementary Fig. S2A; $P = 0.045$ measured by Shannon entropy, Supplementary Fig. S2B). This association remained even after adjustment for other relevant features available in our cohort including age, gender, *t(15;17)*, *inv(16)*, *Del(5/7q)*, *t(v;11q23)*, *t(8;21)*, *EVII*, *CEBPA-dm*, *CEBPA-sil*, *NPM1*, *IDH1/2*, *IDH1/2+DNMT3A*, *DNMT3A*, *FLT3^{ITD}*, and *TET2* (PDR in Supplementary Table S1; epipolymorphism and Shannon entropy in Supplementary Fig. S2C and S2D).

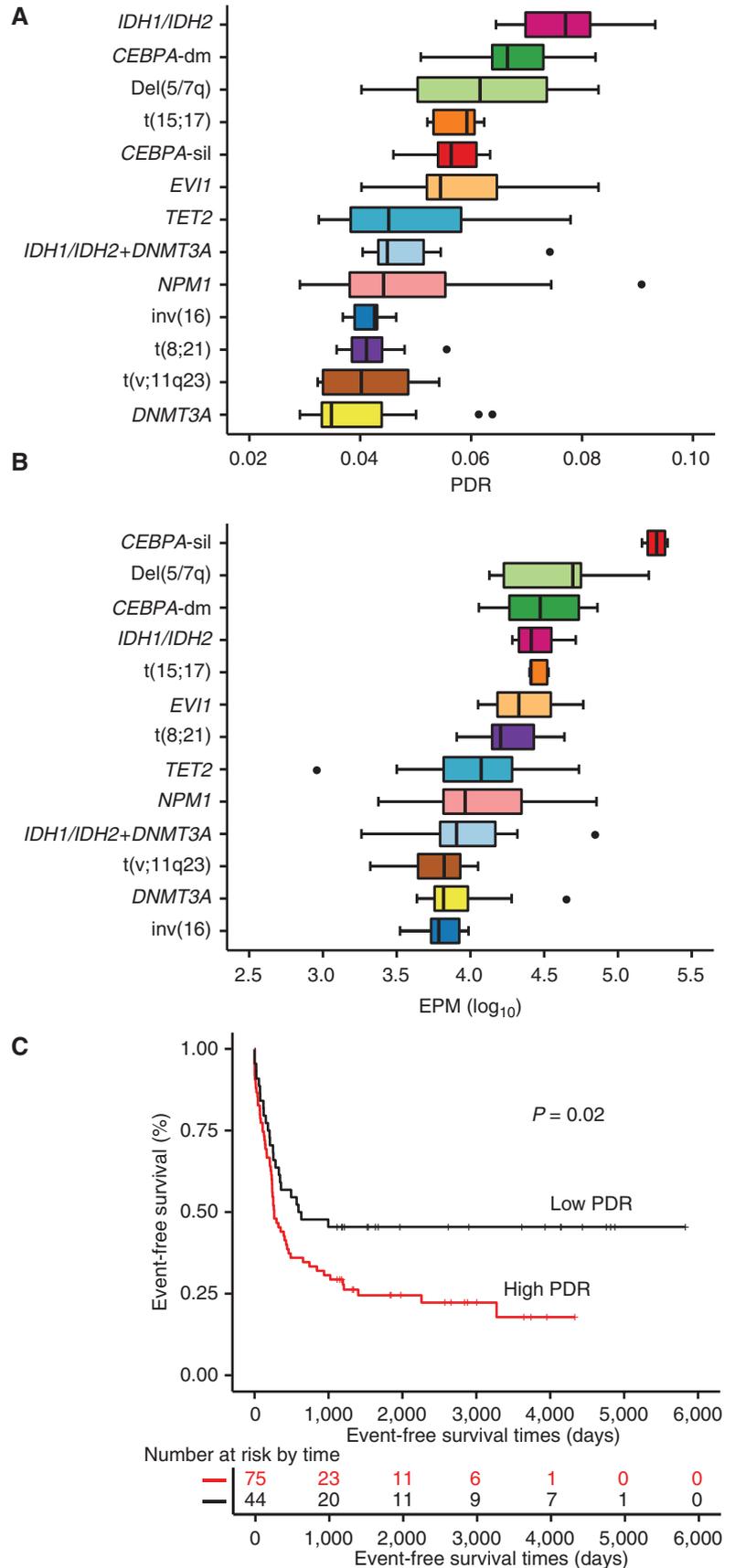
Epialleles Segregate into Specific Patterns Linked to Genetic Background

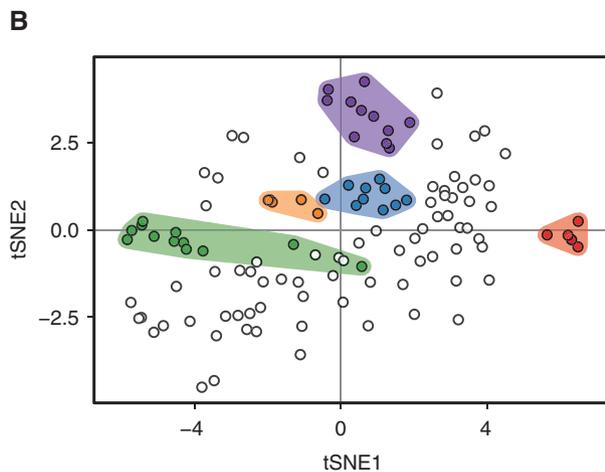
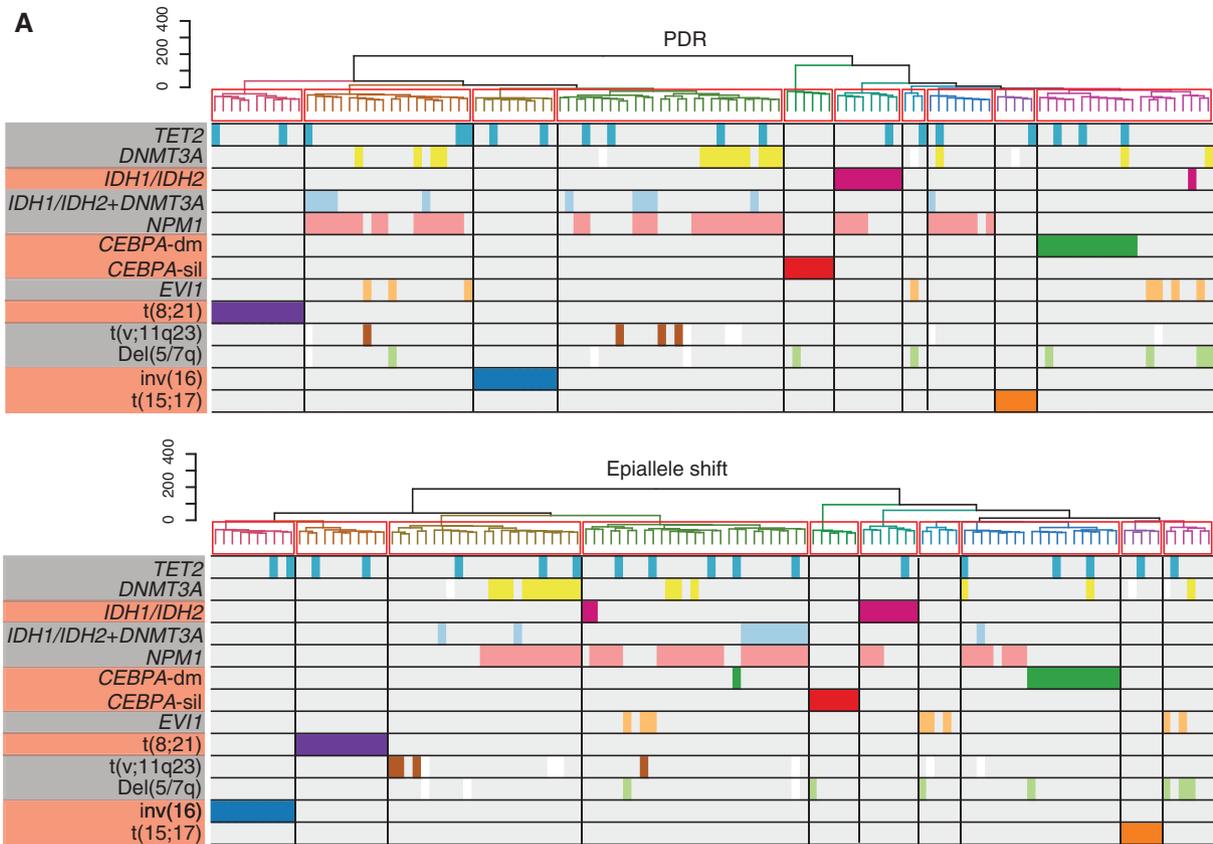
To further understand the association between epiallele complexity and AML subtypes, we used hierarchical clustering and t-distributed stochastic neighbor embedding (tSNE) of the top 50% variable loci (measured by PDR, EPM, epipolymorphism, and Shannon entropy) to determine the relative similarity or differences among epiallele variance and epiallele location in the genome among AML subtypes. Results indicated striking specificity for epiallele genomic distribution in patients with *CEBPA-sil*, *CEBPA-dm*, *inv(16)*, *t(15;17)*, or *t(8;21)*, as these patients formed unique and distinct clusters (Fig. 2A and B; Supplementary Fig. S3A and S3B). In patients with *IDH1/2* mutations, *DNMT3A* mutations, or *IDH1/2+DNMT3A* mutations, epialleles were more broadly distributed and less tightly linked, with partially overlapping distributions (Fig. 2C), suggestive of less tightly defined epiallele disposition. In contrast, epialleles in patients with *TET2*, *Del(5/7q)*, *t(v;11q23)*, *EVII*, or *NPM1* did not form discrete clusters, suggesting that these lesions do not specify where and how epialleles form, but rather that epialleles in these patients occur more stochastically (Supplementary Fig. S3C).

AML Epialleles Are Linked to Particular Transcription Factor Cistromes and Transcriptional Programs

To generate a more precise map of epiallele distribution, we identified the genomic location with gain or loss of the epiallele heterogeneity associated with each AML subtype (Supplementary Fig. S4). Specifically, a total of 5,567 epigenetic loci (eloci) were identified among the various

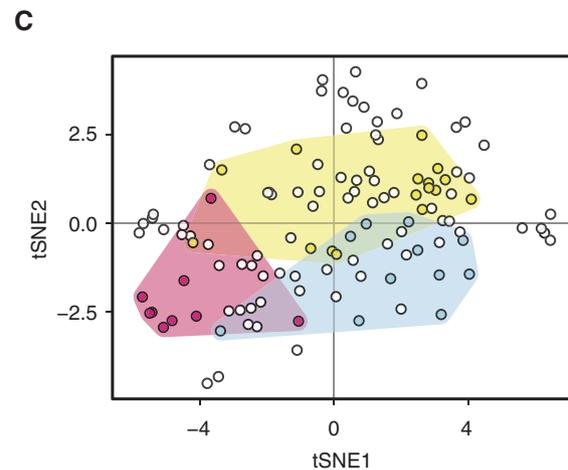
Figure 1. Epigenetic allele diversity is linked to somatic mutations and clinical outcome. **A**, PDR of 13 AML subtypes ranked by median PDR. **B**, EPM of 13 AML subtypes ranked by median EPM. PDR and EPM are significantly different among AML subtypes: $P = 2.26 \times 10^{-9}$ (**A**), and $P = 5.137 \times 10^{-10}$ (**B**), Kruskal-Wallis rank sum test. **C**, Kaplan-Meier curves of patients with primary AML for achievement of complete remission based on PDR.





Subtype

- CEBPA-dm
- CEBPA-sil
- inv(16)
- t(15;17)
- t(8;21)
- Others



Subtype

- IDH1/IDH2
- DNMT3A
- IDH1/IDH2+DNMT3A
- Others

Figure 2. Epigenetic allele diversity of genetic lesions forms discrete clusters. **A**, Unsupervised analysis of epigenetic allele diversity by hierarchical clustering using Euclidean distance and based on the most variable loci across all the patients (PDR, 20,948 loci, top; epiallele shift, 20,764 loci, bottom). Subtypes of patients are represented in the left-most column, and subtypes forming unique clusters are labeled in orange, with other subtypes labeled in gray. **B** and **C**, tSNE analysis of epigenetic allele diversity measured by PDR from all the patients highlighted with five AML subtypes: CEBPA-dm, CEBPA-sil, inv(16), t(15;17), and t(8;21) (**B**) and three AML subtypes: IDH1/IDH2, DNMT3A, and cooccurring IDH1/IDH2 and DNMT3A mutations (**C**).

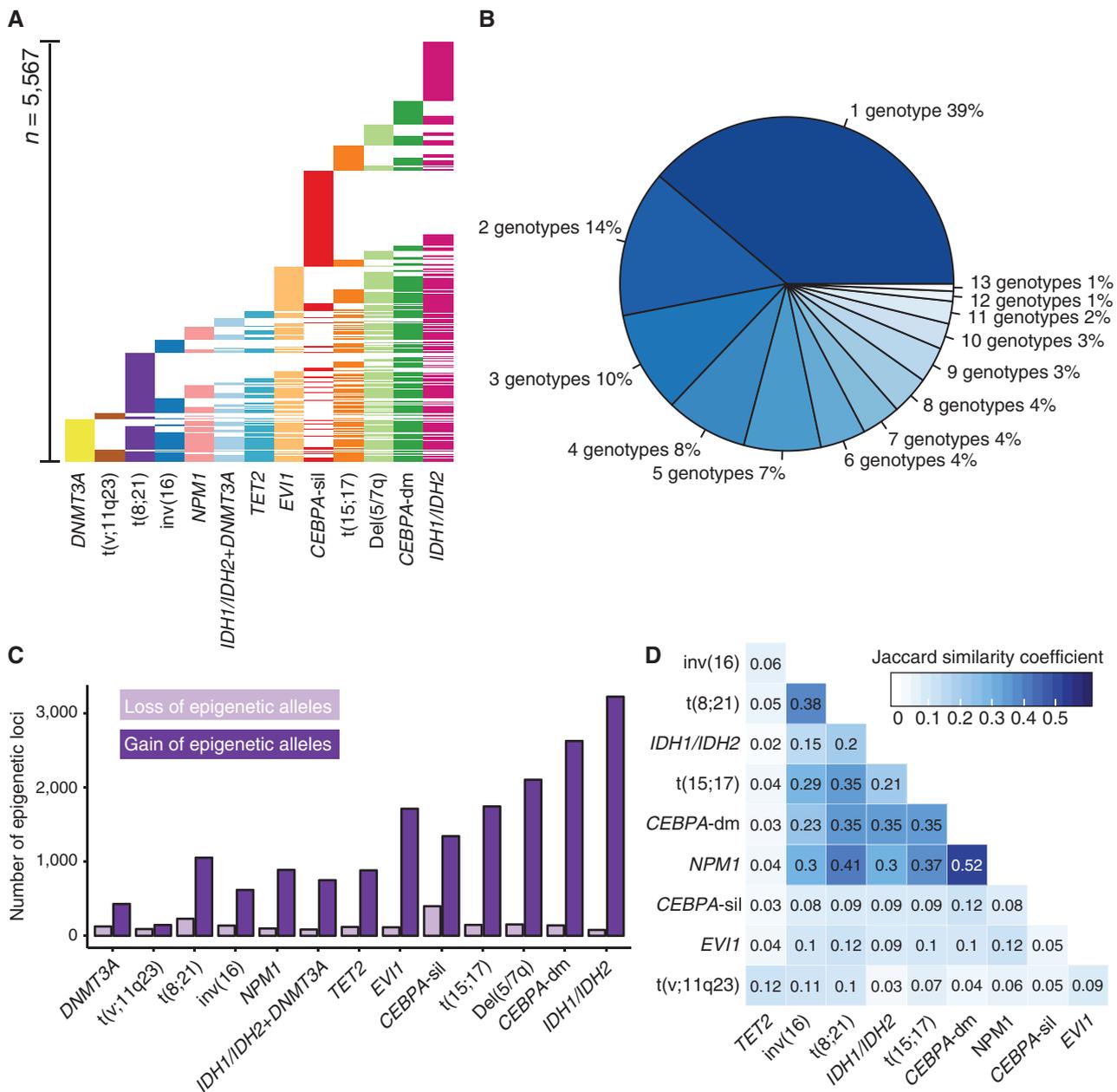


Figure 3. Comparison of AML epigenetic allele diversity measured by PDR with NBM controls. **A**, The number of loci that are shared and unique among 13 AML subtypes. **B**, Pie chart of the shared and unique loci among 13 AML subtypes. **C**, Absolute numbers of loci in each AML subtype, where AML subtypes are ranked in ascending order from left to right based on the median PDR. **D**, Jaccard similarity coefficient of one-on-one comparison among 10 AML subtypes. Only patients with a unique AML subtype from the 13 subtypes are included.

AML subtypes by comparing the AML subtype epiallele diversity values with the epiallele diversity of NBM controls (Fig. 3A). To understand how these loci might be functionally distinct from each other, we determined the transcription factor (TF) binding sites for which the particular pattern of loci specific to each subtype were enriched using motif gene sets from MSigDB v6.1 (see Supplementary Fig. S5: showing enriched TFs known to be expressed in hematopoietic stem cells (HSC) or AML based on BloodSpot; ref. 21).

Several findings are particularly noteworthy. The loci in five subtypes including *CEBPA-dm*, *IDH1/IDH2*, *EVI1*, *Del(5/7q)*, and *TET2* were enriched for AP2 (TFAP2A), which can repress *CEBPA* and *MYC* (22). *MYC* is frequently activated in AML and has a key role in the induction of leukemogenesis (23). The loci in the *IDH* group and *TET2* were enriched for *STAT5A*, a key member of the JAK-STAT pathway regulating HSC functions (24) and leukemia cell maintenance and survival (25). The loci in *CEBPA-sil* were enriched for NF κ B (NFKB1), *STAT3*, and *AP1* (*JUN*). The aberrant activity

of NF κ B in AML fosters cell proliferation and survival (26). STAT3 is involved with hematopoietic growth factor signal transduction (27). Constitutively active STAT3 in AML is related to poor prognosis (28). AP1 is a transcription factor composed of a heterodimer composed of proteins belonging to the FOS, JUN, ATF, and MAF families and controls cell proliferation, differentiation, and apoptosis (29). It was reported JUN family members are overexpressed and activate in AML (30) and have an key role in facilitating AML cell survival and progression (31).

As a complementary approach, we examined whether AML-associated epialleles were enriched for TF chromatin immunoprecipitation sequencing (ChIP-seq) binding profiles derived from human CD34⁺ HSPCs (bloodChIP; ref. 32; Supplementary Fig. S6). Strikingly, the epialleles in 10 AML subtypes (*CEBPA*-dm, *CEBPA*-sil, *Del(5/7q)*, *DNMT3A*, *IDH1/IDH2* + *DNMT3A*, *inv(16)*, *NPM1*, *t(15;17)*, *t(8;21)*, and *TET2*) were enriched for GATA2, which is intriguing given the prominent role of GATA2 dysregulation downstream of many somatic mutations in leukemia (11, 12). In addition, eloci in the *CEBPA*-dm and *CEBPA*-sil subtypes both enriched for SCL (*SCLY*), *LYL1*, and *LMO2* binding sites, whereas the eloci in the *CEBPA*-sil subtype were enriched for ERG. *RUNX1* binding peaks were enriched in the eloci of the *inv(16)* and *TET2* subtypes; *RUNX1* is a binding partner for the CFBF-MYH11 fusion protein (33). However, when we examined the binding profiles of the PML-RARA fusion protein that is generated by the *t(15;17)* translocation (34), there was no enrichment for epialleles ($P > 0.1$, data not shown). Finally, an analysis of curated gene sets in the MSigDB v6.1 linked to epialleles using the hypergeometric test yielded enrichment for genes up-regulated in HSCs in *IDH1/IDH2* [FDR-adjusted P (P_{FDR}) = 0.008], *EVI1* (P_{FDR} = 0.047), *Del(5/7q)* (P_{FDR} = 0.009), and *CEBPA*-dm (P_{FDR} = 0.005) subtypes (Supplementary Tables S2 and S3). These data raise the possibility that epialleles could be linked to gene dysregulation upstream or downstream of key HSC transcription factors.

Notably, epialleles in *CEBPA*-silenced patients were particularly well defined and yielded striking enrichment for gene sets regulated by PRC2 polycomb complexes and genes containing H3K27me3 or bivalent chromatin in embryonic stem cells, as well as genes involved in various signaling pathways ($P_{\text{FDR}} \leq 0.05$; Supplementary Fig. S7A and S7B), including NF κ B, STAT3, and AP1. There was significant enrichment for leukemia stem cell (LSC)-associated gene signatures (e.g., CD34⁺CD38⁻ leukemia repopulating cells compared with more mature CD34⁺CD38⁺ stem cells) and inflammatory response gene signature (Supplementary Table S4). These results are consistent with our prior works (19, 35) showing that these leukemias manifest an immature stem-cell phenotype and dismal clinical outcomes, and suggest possible involvement of deregulated PRC2 function in these cases.

Epigenetic Alleles Are Generally Increased and Highly Diversified across AML Subtypes

Next, we investigated the occurrence of AML-specific eloci in each AML subtype compared with NBM controls. This analysis revealed that, of the 5,567 eloci that we identified

among the various AML subtypes based on PDR (Supplementary Fig. 3A), 39% were detected in only one AML subtype (Fig. 3B). Only 1% of eloci were shared across all 13 subtypes, with the percentages of shared eloci increasing slightly as the number of subtypes sharing a group of eloci decreased (Fig. 3B). When compared with NBM controls, patients with AML across all subtypes manifested a net absolute gain of epiallele heterogeneity (Fig. 3C; epipolymorphism in Supplementary Fig. S8A and B; Shannon entropy in Supplementary Fig. S8C and S8D). This finding prompted us to determine whether there was particular agreement among the sets of eloci of the various AML subtypes. Thus, we measured the Jaccard similarity coefficient for every pair of AML subtype (Fig. 3D). In general, this analysis revealed that the closest degree of similarity occurs between AML subtypes that (i) drive specific epiallele patterns and (ii) are linked to a favorable clinical outcome, which include *inv(16)*, *t(8;21)*, *IDH*, *t(15;17)*, and *CEBPA*-dm, and *NPM1*. AMLs with poor outcome manifested the least agreement of eloci, which may signify that they arise in a more stochastic manner that is perhaps associated with natural selection, whereas the favorable-outcome subtypes harboring a higher agreement of eloci might be a direct or indirect by-product of driver-oncogene effects that are not linked to evolutionary pressures.

Of note, *DNMT3A*-mutant AMLs had fewer eloci compared with other subtypes, perhaps due to allele diversity induced by *DNMT3A* mutations lying outside of the CpG dense regions captured by ERRBS (18). Therefore, we investigated epigenetic heterogeneity in these patients using a different cytosine methylation heterogeneity approach (M-score) suitable to analyze CpGs regardless of density captured through whole-genome bisulfite sequencing profiles generated by Spencer and colleagues (13) from five normal karyotype AML samples with *DNMT3A*^{R882H/C} mutations and five normal karyotype AML samples with no *DNMT3A* mutations. DNA methylation heterogeneity was significantly higher in the *DNMT3A*^{R882H/C} cases ($P = 0.040$ in Supplementary Fig. S9, t test). But it is not further explored herein because these cannot be integrated with our ERRBS epiallele data.

Somatic Mutations Cooperate to Induce Epigenetic Heterogeneity during Leukemogenesis

We wondered whether individual leukemia mutations are sufficient to destabilize the epigenome to yield epigenetic diversity, or whether (like leukemic transformation itself) destabilization would require cooperation between disease alleles. We addressed this question by examining the methylomes of LSK (Lin⁻Sca⁺cKit⁺) cells from healthy (nonleukemic) *Tet2*^{-/-}, *Idh2*^{R140Q} knock-in, *Flt3*^{ITD}, *Tet2*^{-/-};*Flt3*^{ITD}, and *Idh2*^{R140Q};*Flt3*^{ITD} mice. Epiallele diversity measured by PDR, epipolymorphism, and Shannon entropy was first analyzed in an unsupervised fashion using hierarchical clustering and tSNE (PDR: Fig. 4A-D; epipolymorphism: Supplementary Fig. S10A-S10D; Shannon entropy: Supplementary Fig. S10E-S10H). This analysis showed that double-mutant, but not single-mutant, LSK cells segregated to distinct nodes, and that they formed more defined clusters than the single-mutant mice. In addition, and notably, the epiallele

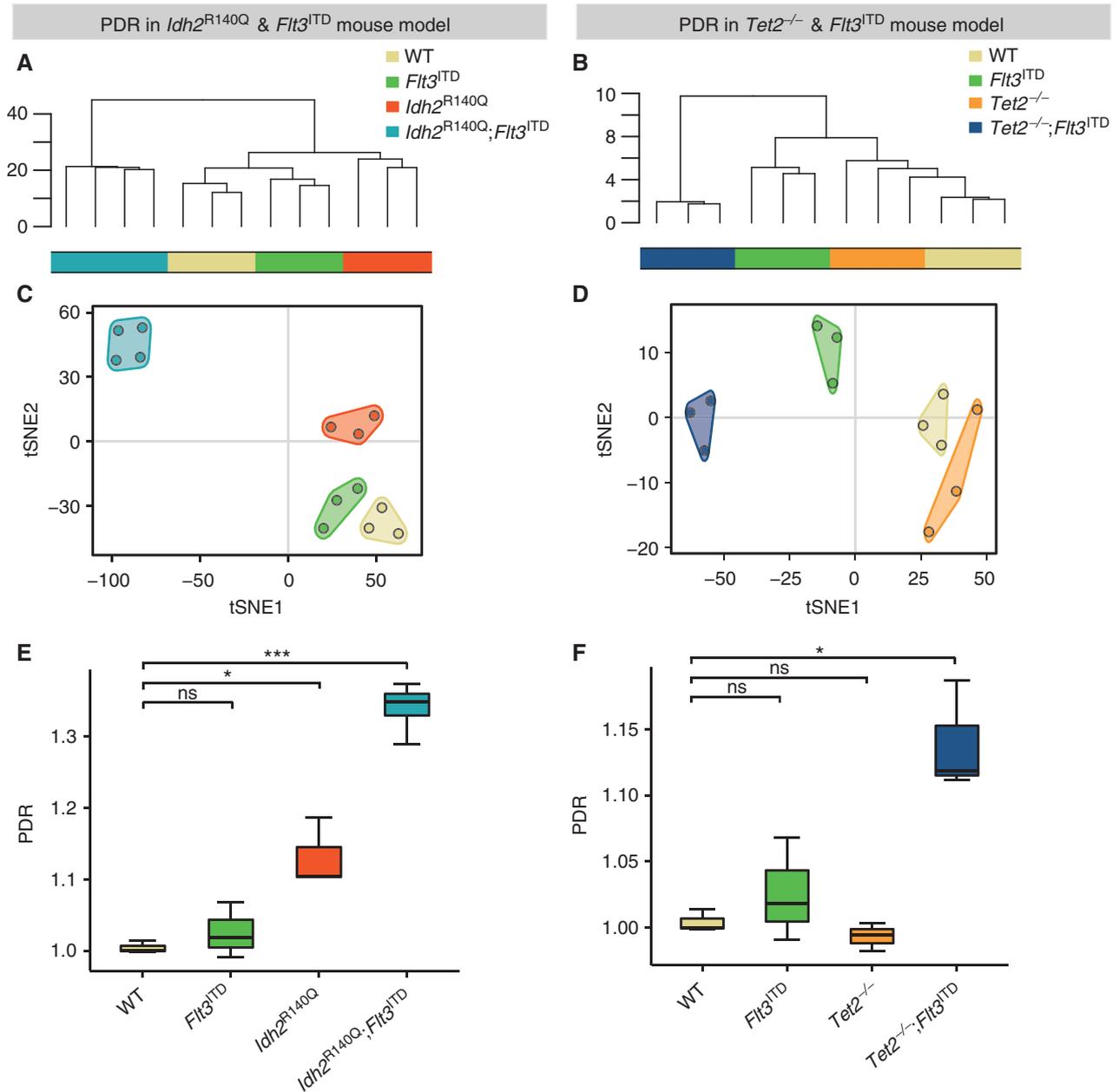


Figure 4. Somatic mutations cooperatively enhance epigenetic allele diversity. **A** and **B**, Unsupervised analysis by hierarchical clustering, **C** and **D**, tSNE analysis, and **E** and **F**, box plot of epigenetic allele diversity measured by PDR in wild-type (WT) mice (3 samples), *Flt3*^{ITD} mice (3 samples), *Idh2*^{R140Q} mice (3 samples), and *Idh2*^{R140Q};*Flt3*^{ITD} mice (4 samples; **A**, **C**, **E**) and in wild-type mice (3 samples), *Flt3*^{ITD} mice (3 samples), *Tet2*^{-/-} mice (3 samples), and *Tet2*^{-/-};*Flt3*^{ITD} mice (3 samples; **B**, **D**, **F**). Here, *, $P \leq 0.05$; ***, $P \leq 0.001$ by *t* test.

profiles of *Idh2*^{R140Q} mice were more severely perturbed than those of *Tet2*^{-/-} mice.

We then performed a supervised analysis of the three epiallele diversity metrics. In both the *Idh2*^{R140Q} and *Tet2*^{-/-} settings, the cross with *Flt3*^{ITD} yielded the greatest and most significant degree of epiallele diversity (PDR: $P = 1.709 \times 10^{-4}$ and $P = 0.027$ by *t* test, compared with wild-type in Fig. 4E and F; epipolymorphism: $P = 2.26 \times 10^{-5}$ and $P = 0.038$ by *t* test in Supplementary Fig. S11A and S11B; Shannon entropy: $P = 5.54 \times 10^{-3}$ and $P = 0.033$ by *t* test in Supplementary Fig. S11C and S11D).

Notably, the *Tet2* mutation alone had little effect, whereas the *Idh2*^{R140Q} mutation alone manifested significantly greater epiallele diversity compared with wild-type mice ($P = 0.041$ by *t* test, Fig. 4E). The *Flt3*^{ITD} mutation alone generated a small degree of epiallele diversity (Fig. 4E and F). Hence, there were major differences between the epigenetic effects of the *Tet2* loss-of-function and *Idh2*^{R140Q} mutation, consistent with biological differences between *IDH* and *TET2* mutations in patients and mouse models (5, 12, 36). Importantly, these results show that increased epiallele diversity can precede transformation to

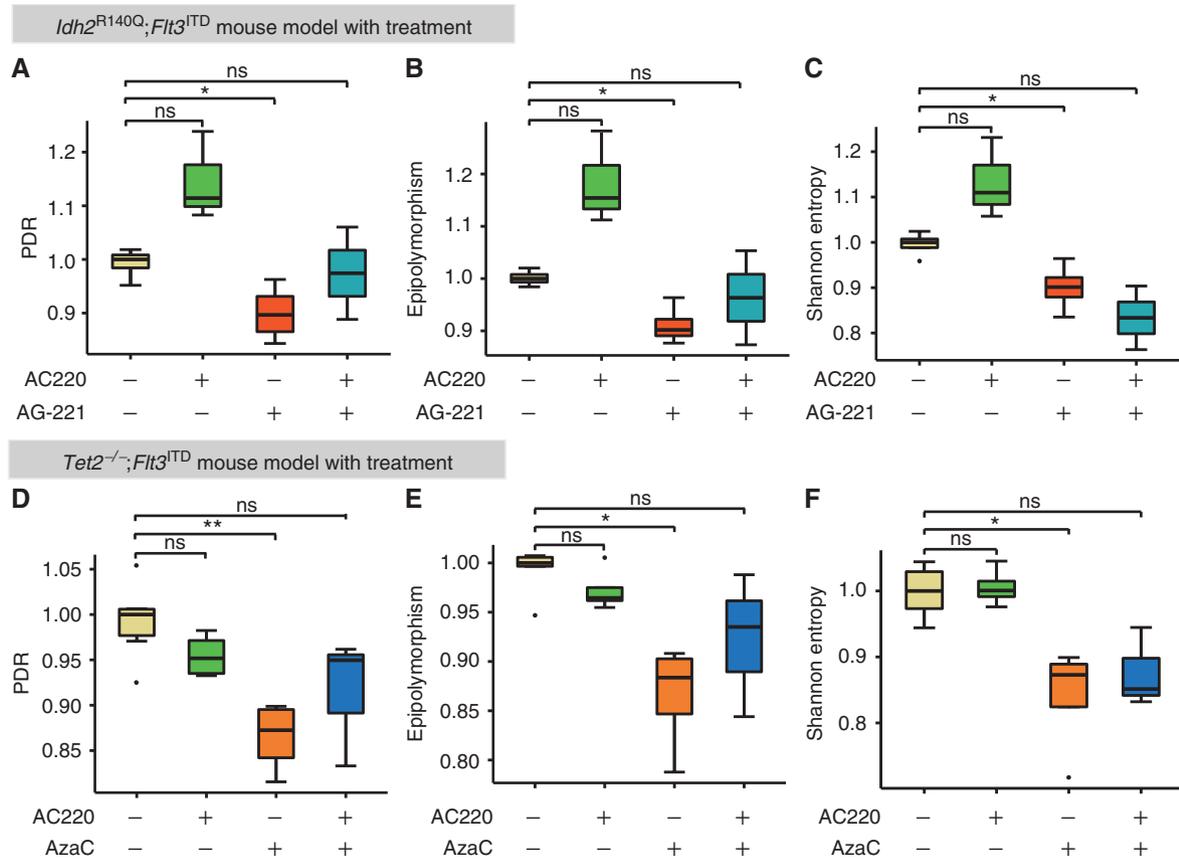


Figure 5. Epigenetic targeted therapy reverses the epigenetic heterogeneity. **A–C**, Boxplot of PDR, epipolymorphism, and Shannon entropy of *Idh2^{R140Q};Flt3^{ITD}* mice treated separately with vehicle (4 samples), AC200 (FLT3 inhibitor; 3 samples), AG-221 (IDH2 inhibitor; 4 samples), or combined AG-221 and AC200 (2 samples). **D–F**, Box plot of PDR, epipolymorphism, and Shannon entropy of *Tet2^{-/-};Flt3^{ITD}* mice treated separately with vehicle (6 samples), AzaC (DNMT inhibitor; 4 samples), AC200 (4 samples), or combined AzaC and AC220 (3 samples). Here, *, $P \leq 0.05$; **, $P \leq 0.01$ by *t* test.

AML and is enhanced by cooperation between somatic mutations, especially those involving *IDH* mutations.

To assess how these epialleles might be linked to murine TF cistromes, we analyzed TF-binding sites associated with *eloci* in the *Idh2^{R140Q};Flt3^{ITD}* double-mutant mouse model. We detected enrichment for 38 TFs ($P_{\text{FDR}} \leq 0.05$; Supplementary Table S5), nine of which were also enriched in patients with *IDH2;FLT3^{ITD}* AML (Supplementary Fig. S12). These include PBX1, a regulator of HSC transcriptional programming (37); GFI1, a transcriptional repressor that restricts HSC proliferation (38); LEF1, which regulates the regenerative fitness of HSCs and self-renewal of LSCs (39); and MAZ, which is a regulator of *MYC* (40). NFATC are a family of TFs, the targeting of which caused antileukemic effects in *FLT3^{ITD}* AML (41) and drove transcriptional programs induced by *FLT3^{ITD}* (42). CDX2 was reported to contribute to AML leukemogenesis (43). Hence, epialleles arising in the *IDH2* and *FLT3^{ITD}* context in humans and mice may affect similar TF programs with relevance to AML biology.

DNMTi and Mutant IDH2 Inhibitors Can Suppress Epigenetic Allele Diversity

Given the association of epiallele diversity with unfavorable clinical outcomes, we examined whether it could be

reversed using epigenetic therapy drugs. To address this question we analyzed ERRBS profiles of LSK cells from syngeneic mice that had been transplanted with CD45.2⁺ *Idh2^{R140Q};Flt3^{ITD}* or CD45.2⁺ *Tet2^{-/-};Flt3^{ITD}* cells (12), and then mice in each group were administered three different treatments: *Idh2^{R140Q};Flt3^{ITD}* mice were treated via (i) twice-daily administration of the IDH2 inhibitor AG-221, for 6 weeks, (ii) daily administration of the FLT3 inhibitor AC220, (iii) combined AG-221 and AC220 therapy, or (iv) vehicle administration; and *Tet2^{-/-};Flt3^{ITD}* mice were treated via (i) daily administration of the DNMTi (DNMTi) 5'-azacytidine (AzaC) for 5 days every 21 days, for four cycles (similar to the clinical context), (ii) daily administration of the FLT3 inhibitor AC220, (iii) combined AzaC and AC220 therapy, or (iv) vehicle administration (12).

We examined the impact of these treatments on the extent of epiallele diversity as well as the absolute numbers of epialleles in each case. Comparing with vehicle, we observed that the most profound and significant reduction in epiallele diversity in either strain was mediated by AG-221 in *Idh2^{R140Q};Flt3^{ITD}* mice ($P = 0.03$ for PDR; $P = 0.011$ for epipolymorphism; $P = 0.028$ for Shannon entropy by *t* test in Fig. 5A–C), whereas in *Tet2^{-/-};Flt3^{ITD}* mice the most profound reduction was observed after AzaC treatment

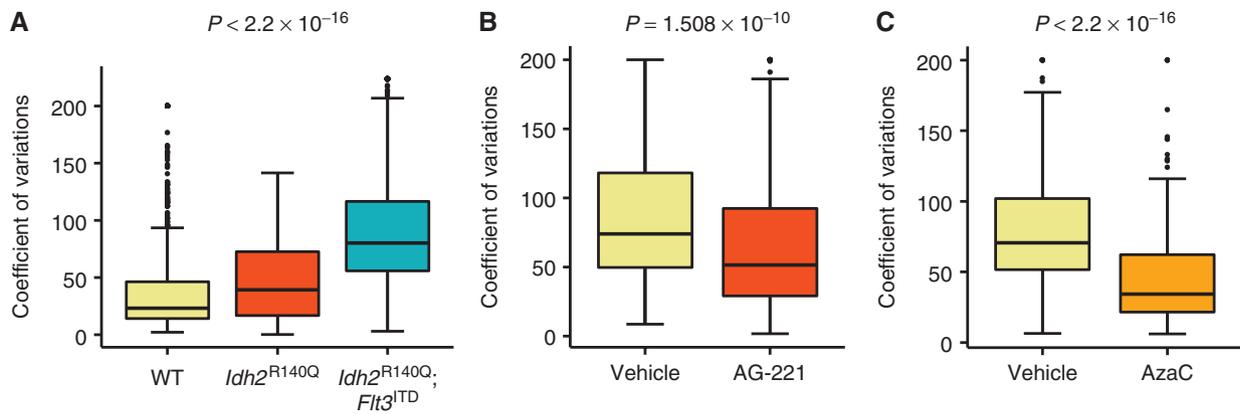


Figure 6. Transcriptional variance is associated with epigenetic loci at genomic regions (neighborhood and gene body). **A**, Box plot comparing variance in transcript expression levels of genes with loci in their genomic regions ($P < 2.2 \times 10^{-16}$, Kruskal-Wallis rank sum test) in wild-type (WT) mice and in *Idh2*^{R140Q} and *Idh2*^{R140Q};*Flt3*^{ITD} mice. **B**, Box plot comparing variance in transcript expression levels of genes with loci in their genomic regions ($P = 1.508 \times 10^{-10}$, Wilcoxon rank sum test) in *Idh2*^{R140Q};*Flt3*^{ITD} mice treated with vehicle and AG-221. **C**, Box plot comparing variance in transcript expression levels of genes with loci in their genomic regions ($P < 2.2 \times 10^{-16}$, Wilcoxon rank sum test) in *Tet2*^{-/-};*Flt3*^{ITD} mice treated with vehicle and AzaC.

($P = 0.002$ for PDR; $P = 0.014$ for epipolymorphism; $P = 0.025$ for Shannon entropy by t test in Fig. 5D–F). AC220 had very little effect as a single agent, and fundamental differences in epiallele diversity between *Tet2*^{-/-};*Flt3*^{ITD} and *Idh2*^{R140Q};*Flt3*^{ITD} mice remained following AC220 treatment. Combined AG-221 and AC220 therapy resulted in a slightly lower degree of epiallele diversity compared with AC220 monotherapy, in both strains. Similar effects were observed for combined treatment of *Tet2*^{-/-};*Flt3*^{ITD} mice with AzaC and AC220 compared with treatment with vehicle. These effects were consistent across PDR, epipolymorphism, and Shannon entropy analysis.

Because we did not have available ERRBS data to measure epialleles in humans treated with DNMTi, we explored DNA methylation heterogeneity scores from six patients profiled using HELP microarrays at diagnosis or at days 15 or 29 post-treatment (14, 44). These results cannot be exactly compared with our epiallele measurements because both the DNA methylation platforms and analyses are completely different. However, it was notable that 3 of 6 patients manifested reduction in DNA methylation heterogeneity at post-treatment (Supplementary Fig. S13). Collectively, our data suggest that epigenetic heterogeneity is potentially reversible by epigenetic therapy.

Transcriptional Hypervariability Is Linked to Epigenetic Allele Diversity

It was previously shown that epiallele diversity is linked to variable expression of the respective genes, suggesting that affected genes can experience multiple transcriptional states (9). To determine whether similar effects could be observed in our primary AML cohort, we investigated the link between epiallele diversity and transcriptome variance. Patients with higher epiallele diversity had significantly higher transcriptome variance than patients with low epiallele diversity (t test, PDR: $P = 2.6334 \times 10^{-11}$; epipolymorphism: $P = 3.2092 \times 10^{-7}$; Shannon entropy: $P = 2.1893 \times 10^{-11}$ in Supplementary Fig. S14A). Using the European LeukemiaNet (ELN)

risk-stratification scheme (45), we then observed that the unfavorable outcome group had significantly higher transcription variance and epiallele diversity compared with the good/intermediate prognosis group (transcription variance: Kruskal-Wallis rank sum test, $P < 2.2 \times 10^{-16}$ in Supplementary Fig. S14B; Kruskal-Wallis rank sum test, PDR: $P = 0.0621$; epipolymorphism: $P = 0.0035$; Shannon entropy: $P = 0.0058$ in Supplementary Fig. S15A–S15C). Dissecting transcriptional heterogeneity across AML subsets indicated significant association with promoter epialleles ($P < 0.05$, binomial test) in all cases except for t(11q23) and inv(16) patients (Supplementary Fig. S16A). AML subtypes with more coordinated epialleles also featured more correlation between their respective sets of differentially expressed genes (vs. normal CD34⁺ cells; Supplementary Fig. S16B). Hence epialleles tend to associate both with differentially expressed as well as differentially variable gene transcripts, suggesting that genes with promoter loci are prone to experience transcriptional deregulation.

Next, to determine the association between transcription variance downstream of leukemia mutations, we compared gene expression heterogeneity among wild-type, *Idh2*^{R140Q}, and *Idh2*^{R140Q};*Flt3*^{ITD} mice. We found that genes containing epialleles manifested significantly greater levels of transcriptional heterogeneity in *Idh2*^{R140Q};*Flt3*^{ITD} mice compared with wild-type and single-mutation mice as measured by the intragroup coefficient of variation (Kruskal-Wallis rank sum test, $P < 2.2 \times 10^{-16}$, Fig. 6A). Notably, in *Idh2*^{R140Q};*Flt3*^{ITD} mice, transcriptional heterogeneity associated with epialleles decreased significantly after treatment with AG-221 (Wilcoxon rank sum test, $P = 1.508 \times 10^{-10}$; Fig. 6B). Similarly, in *Tet2*^{-/-};*Flt3*^{ITD} mice, the transcriptional heterogeneity of genes containing epialleles decreased significantly after treatment with AzaC (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$; Fig. 6C). Together, these results provide one possible explanation of how epigenetic therapies might benefit patients with AML: by reducing the number of different transcriptional states among leukemia cells and hence population fitness.

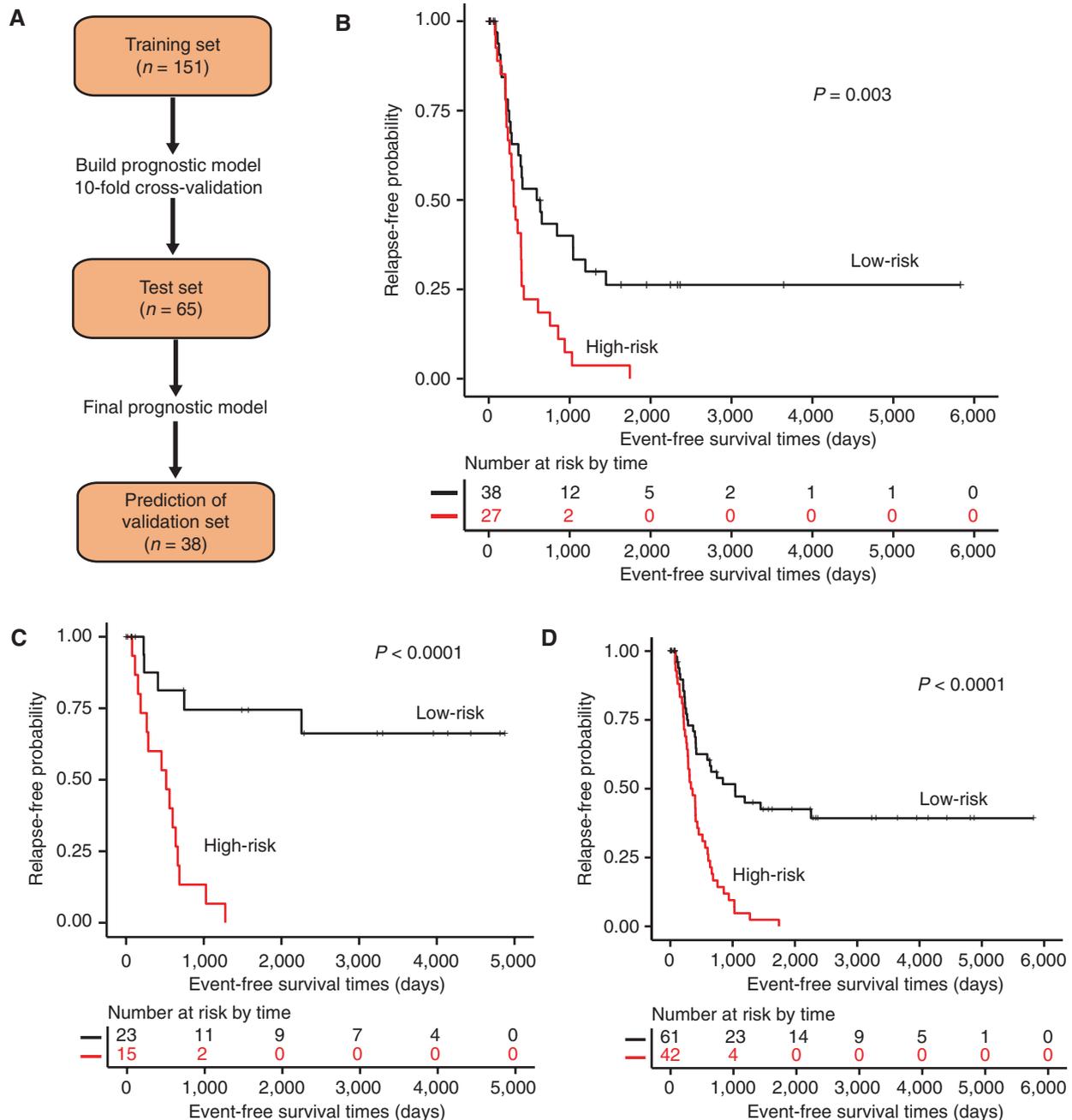


Figure 7. An epiallele prognostic classifier predicts clinical outcome in patients with AML. **A**, Outline describing the steps for building the prognostic classifier. In the first step, 151 randomly selected patients were used to identify loci that best predicted survival. The model was tested on a different cohort of 65 patients (test set). Once the final model was selected, its performance in predicting survival was tested in a validation set consisting of 38 randomly selected cases. **B**, Kaplan-Meier curves for event-free survival for the predicted groups in the test set. Event-free survival was compared between patients in the test set that were predicted either high-risk or low-risk by our classifier. **C**, Kaplan-Meier curves for event-free survival for the predicted groups in the validation set. Event-free survival was compared between patients in the validation set that were predicted either high-risk or low-risk by our classifier. **D**, Kaplan-Meier curves for event-free survival for the predicted groups in the combined test and validation set. Event-free survival was compared between patients in the combined test and validation set that were predicted either high-risk or low-risk by our classifier.

An Epiallele Prognostic Classifier Predicts Clinical Outcome in AML

Differentially methylated gene sets can serve as useful clinical biomarkers for AML and other tumors (10, 46, 47). However, epialleles have not been investigated as potential disease

outcome classifiers. To determine whether epiallele-based biomarkers could yield prognostic value in AML, we followed a three-step approach toward building a putative epiallele prognostic model (Fig. 7A). To increase statistical power, we combined our 119 AML patient ERRBS profiles with those from another set of 137 patients with clinically annotated

AML, all of whom had subsequently relapsed after therapy (9). Two patients were removed because they did not have information on time to relapse. We then calculated the PDR values of the 635 loci shared by all 254 patients to build an epiallele diversity prognostic classifier. The complete cohort was then randomly divided into a training set ($n = 151$), a test set ($n = 65$), and a validation set ($n = 38$).

We used the supervised principal components (SuperPC) method (48) to train a Cox proportional hazards regression model for event-free survival in the training set (event denotes relapse), using a 10-fold cross-validation procedure. The highest-scoring model contained 26 features corresponding to 8 genes (Supplementary Table S6) and successfully classified the test set into high-risk and low-risk groups (log-rank test $P = 0.003$, Cox regression model: HR = 2.36, 95% CI = 1.32–4.21, $P = 0.00385$; Fig. 7B). The prognostic model also predicted event-free survival on the 38-patient validation cohort (log-rank test $P < 1 \times 10^{-4}$, Cox regression model: HR = 8.87; 95% CI = 2.80–28.1, $P = 2.09 \times 10^{-4}$; Fig. 7C), as well as the combined test and validation cohort (log-rank test $P < 1 \times 10^{-4}$, Cox regression model: HR = 3.32, 95% CI = 2.00–5.54, $P = 3.97 \times 10^{-6}$, Fig. 7D). Finally, to determine the significance of this epiallele prognostic classifier as a putative biomarker, we performed a multivariate analysis considering 26-loci SuperPC score, age, gender, ELN risk stratification (45), *FLT3*^{ITD}, *CEBPA*-dm, *t(8;21)*, *NPM1*, and *inv(16)* as covariates. We tested these in a Cox multivariate regression model, and observed that the epiallele classifier retained its statistical significance as a putative clinical outcome biomarker (HR = 3.0238, $P = 4.11 \times 10^{-6}$, 95% CI = 1.8883–4.842; Supplementary Table S7). To confirm the robustness of parameters in the epiallele prognostic model, we performed 1,000 additional random splits of the dataset into training sets of 151 patients and test sets of 103 patients. All 1,000 runs were validated with a $P < 0.05$ in a Cox proportional hazards regression model. Therefore, epiallele classifiers can be considered as potential biomarkers for prediction of event-free survival in patients with AML.

DISCUSSION

Epigenetic heterogeneity is increasingly recognized as a critical feature of tumors that endows them with biological variability and options for subsets of cells to manifest selective growth and survival advantages. However, the source of epigenetic heterogeneity and its potential link to cancer-associated mutations remains unknown. Herein, we addressed this question in AML, a disease that is usually fatal but that nonetheless is characterized by a relative paucity of somatic mutations. We explored whether there is any relationship between canonical leukemia genetic lesions and epiallele diversity that is the result of variability in cytosine methylation patterning. We considered two primary possible scenarios. First, it is possible that epigenetic diversification is a purely stochastic phenomenon that is unrelated to somatic mutations and is a by-product of AML disease progression. Indeed, this mechanism seems to explain in large part the epigenetic heterogeneity that occurs in diffuse large B-cell lymphoma (16), and is strongly linked to the demethylating actions of AICDA (49). Second, it is also possible that epige-

netic heterogeneity is induced by particular somatic mutations, which implies that such mutations might in some way disrupt regulatory states that would normally be strictly controlled by epigenetic regulation. In this latter case, epigenetic heterogeneity might be predicted to precede and perhaps even contribute to malignant transformation.

Our data suggest that both scenarios may be correct and perhaps not mutually exclusive, and that the particular characteristics and degree to which they occur is linked to the mutational profiles of individual patients. Consistent with the notion that mutations can be a source of epigenetic heterogeneity, we observed that canonical transcription factor translocations and mutations, for example, *t(8;21)*, *inv(16)*, *t(15;17)*, and *CEBPA* double mutations, are associated with well-defined patterns of epialleles. Notably, the patterns of epialleles affected in each of these translocations or mutations were more correlated than among other somatic mutations. It seems unlikely that these are direct effects, because in general neither *CEBPA* nor the respective fusion proteins are enriched for DNA consensus motifs, as determined by ChIP-seq binding to these sites. The one exception to this was the enrichment for *RUNX1* motifs among epialleles present in *inv(16)* AMLs. The similarities between these forms of AML must emanate from some other source, perhaps linked to their more differentiated state or some other undiscovered transcriptional effect. A second peculiarity of these AMLs, especially *t(15;17)* and *CEBPA* double mutations, is their relatively high burden of epialleles despite their relatively favorable prognosis. Hence, apparently not all epialleles are created equal. We propose that the well-defined epialleles linked to these mutations do not necessarily confer population fitness, whereas in most other AMLs, the bulk of epiallele diversity is more stochastic and hence more likely to lead to natural selection of favorable epigenetic states. Along these lines, the set of patients with hypermethylation and silencing of *CEBPA* and a chemotherapy-resistant phenotype manifest among the highest burdens of epialleles and may reflect the consequence of epigenetic heterogeneity induced by unknown sources. Moreover, epiallele diversity cannot be blindly assumed to be an unfavorable finding in AML without considering the mutational context, arguing for the need for integrative biomarkers.

On the other hand, patients with aberrant *IDH* alleles, which are known to disrupt the epigenome at many levels (5–7), manifest a relatively heavy burden of allele diversity. Engineering the expression of the *Idh2*^{R140Q} allele alone in mice induced epiallele diversity among hematopoietic stem cells prior to manifestation of overt transformation by these cells, an effect that was significantly enhanced by coexpression of a *Flt3*^{ITD} allele. *Tet2* deficiency did not induce epiallele formation in LSK cells alone, although *Tet2* deficiency in combination with *Flt3*^{ITD} appears to synergize to yield a significant increase in epiallele diversity. Collectively, these data show that epigenetic diversity can occur prior to transformation as a consequence of somatic mutations, perhaps enabling premalignant cells to sample many epigenetic states in a way that facilitates eventual leukemogenesis. These data are in line with findings showing almost universal hypermethylation and silencing of a set of 45 genes (10) that are otherwise expressed in normal hematopoietic cells. It is

conceivable that epigenetic heterogeneity induced by somatic mutations in normal HSCs will eventually lead to silencing of these genes, and that this event plays a key role in transformation. Our work is suggestive of a link between epigenetic heterogeneity, somatic mutations, and clinical outcome. We believe these are important conceptual advances that provide the basis for new mechanistic hypotheses, which we expect will emanate from these findings. Understanding whether and how mutant transcription factors and epigenetic modifiers can destabilize the epigenome could provide important insights into population fitness and the resilience and relapsing nature of AML.

Consistent with this notion, using orthogonal approaches, we show that epigenetic allelic diversity is an indicator of disease severity in patients with newly diagnosed AML (9). That is, the greater the epiallele diversification, the greater the likelihood that subsets of cells will manifest regulatory states that enable leukemia cells to survive exposure to chemotherapy or other drugs and repopulate the disease. The significance of epiallele diversity is further reflected by the finding that genes linked to epialleles at the same time manifest heterogeneous transcription states (Fig. 6). Epigenetic allelic complexity is thus of clinical importance because such heterogeneity provides AML with more avenues of escape when targeted by chemotherapy drugs. This concept is supported by the finding that specimens from patients with relapsed AML manifest epiallele selection (9), a finding that has also been observed in relapsed lymphomas (16). Finally, given the significance of population fitness for disease outcome, it is intriguing to consider the concept of clonal reduction as a novel therapeutic target for AML. Whereas there is no obvious way to reduce disease clonality from a genetic perspective, it is compelling to hypothesize that DNMTi could mediate such effects at the cytosine methylation level. By causing a reduction in cytosine methylation across the genome, it would stand to reason that epialleles would also become reduced in complexity. Indeed, we observe a significant reduction in epiallele diversity in *Tet2*^{-/-};*Flt3*^{ITD}-mutant AMLs treated with AzaC, suggesting that a benefit of this class of drugs in AML could be linked to a reduction in population fitness. This idea is further supported by our finding that AzaC also reduced transcriptional heterogeneity in AMLs. Perhaps this effect might help to explain the potential benefit of priming patients for chemotherapy by first exposing them to a DNMTi, thus minimizing the options that the tumor has to adapt to chemotherapy exposure and reducing the chance that particular cancer cells could escape therapy. Notably, we observed a similar effect on epigenetic and transcriptional diversity in the case of the mutant IDH2 inhibitor AG-221. Hence a similar beneficial effect on epigenetic diversity may be achieved by specific reversal of the effects of epigenetic driver lesions, as opposed to the more global and nonspecific actions of DNMTi. This suggests that a combination of IDH inhibitors and chemotherapy holds promise as an effective approach for AML. Together, these findings support the rationale for future studies to rigorously track epigenetic heterogeneity over time in patients with AML at relevant time points, and to determine whether epiallele reduction associates with improved therapeutic response.

METHODS

Patient Characteristics

The AML patient samples analyzed in this study were obtained from publicly available data (18). Briefly, there were 119 adult primary AML patients, including 67 males and 52 females. The median age of this cohort was 44. Here, 106 patients with AML were annotated by 13 genetically and epigenetically defined AML subtypes (Supplementary Table S8): *CEBPA*-dm (12 patients), *CEBPA*-sil (6 patients), *Del(5/7q)* (7 patients), *inv(16)* (10 patients), *IDH1/IDH2* (9 patients with an *IDH* mutation and without the *DNMT3A* mutation), *DNMT3A* (16 patients with the *DNMT3A* mutation and without an *IDH* mutation), *IDH1/IDH2* + *DNMT3A* (11 patients with co-occurring *IDH* and *DNMT3A* mutations), *NPM1* (42 patients), *t(8;21)* (11 patients), *t(15;17)* (5 patients), *EVII* (8 patients), *t(v;11q23)* (4 patients), and *TET2* (18 patients). In addition, 14 NBM controls (CD34⁺ hematopoietic stem and progenitor cells; 7 males and 7 females) were obtained from our prior study (9), 5 of which were purchased from AllCells and 9 of which were isolated using magnetic bead positive selection for CD34⁺ (Miltenyi Biotec) from freshly collected bone marrow samples from individuals without known hematologic malignancies.

Mouse Models

The mouse samples that were analyzed in this study were obtained from prior works (11, 12). Briefly, the conditional *Vav-cre*⁺*Tet2*^{-/-} (*VTet2*^{-/-}) mice were described previously (36), and *Flt3*^{ITD} mice were kindly provided by Gary Gilliland (University of Pennsylvania, Philadelphia, PA). *Vav-cre*⁺*Tet2*^{-/-};*Flt3*^{ITD} mice were generated by crossing *VTet2*^{-/-} mice to the constitutive knock-in *Flt3*^{ITD} murine model, as reported in Shih and colleagues' work (11). In addition, the *Idh2*^{R140Q} mutation was targeted by a codon change from CGA to CAA in exon 4. A conditional mouse model that expresses the *Idh2*^{R140Q} AML disease allele from the endogenous locus was crossed to mice with the inducible *Mx1-Cre* allele and the *Flt3*^{ITD} knock-in allele to generate *Mx1-Cre Idh2*^{R140Q};*Flt3*^{ITD} mice, as reported by Shih and colleagues (12).

For leukemia therapeutic questions (12), CD45.2⁺ *Flt3*^{ITD};*Tet2*-mutant AML cells and CD45.1⁺ support marrow were engrafted into CD45.1⁺ congenic recipients, and recipient mice were allowed to develop AML with engraftment of 80% to 90% CD45.2⁺ *Flt3*^{ITD};*Tet2*-mutant cells and expansion of leukemic blasts in the peripheral blood, bone marrow, and spleen. Then, mice were treated with vehicle or AC220 daily at 10 mg/kg, and AzaC was administered at 5 mg/kg daily for 5 days every 21 days, for four cycles. Mice also were treated with combination therapy with AzaC and AC220. In addition, CD45.2⁺ *Idh2*^{R140Q};*Flt3*^{ITD}-mutant AML cells and CD45.1⁺ support marrow were engrafted into CD45.1⁺ recipient mice. Then, mice were treated with vehicle or AC220 daily at 10 mg/kg, or AG-221 at 100 mg/kg twice daily for 6 weeks, or combined AC220 + AG-221 therapy where AG-221 was administered at 40 mg/kg (12).

ERRBS

ERRBS was performed using a protocol as described previously (9, 11, 12, 18). Briefly, genomic DNA was digested with *MspI*. DNA fragments were end-repaired. Library fragments were treated with bisulfite and PCR-amplified. Libraries were sequenced on Illumina HiSeq. Reads were aligned to the reference genome using Bismark. The ERRBS data for 119 AML patient samples was downloaded from the Gene Expression Omnibus (GEO; accession number GSE86952; ref. 18). The ERRBS data for 14 NBM controls and an additional 137 clinically annotated AML patient samples (for survival analysis) were downloaded from the Database of Genotypes and Phenotypes (dbGaP), via accession number phs001027.v1.p1(9). The mouse strains were profiled via ERRBS. Briefly, the genome-wide DNA

methylation status of the Lineage⁺Sca⁺cKit⁺ (LSK) cell populations from wild-type (3 samples), *Idh2*^{R140Q} (3 samples), *Flt3*^{ITD} (3 samples), *Tet2*^{-/-} (3 samples), *Idh2*^{R140Q};*Flt3*^{ITD} (4 samples), and *Tet2*^{-/-};*Flt3*^{ITD} mice (3 samples) were profiled (11). The ERRBS datasets for *Idh2*^{R140Q} and *Idh2*^{R140Q};*Flt3*^{ITD} were from the author upon request; others were obtained from GEO (accession number GSE57114). In addition, the DNA methylation status of the LSK populations from *Idh2*^{R140Q};*Flt3*^{ITD} mice treated with vehicle (4 samples), AC220 (3 samples), AG-221 (4 samples), or AC220/AG-211 combination therapy (2 samples) were profiled (GEO, accession number GSE78690). The DNA methylation status of the LSK populations from *Tet2*^{-/-};*Flt3*^{ITD} mice treated with vehicle (6 samples), AC220 (4 samples), AzaC (4 samples), or AC220/AzaC combination therapy (3 samples) were also profiled (GEO, accession number GSE78690; Supplementary Table S9; ref. 12).

Calculation of Epiallele Diversity

In aligned ERRBS data, a read at one given locus with four adjacent CpG sites was considered as a discordant read if it showed different methylated and unmethylated states at a given locus. PDR (15) at each locus was defined as $\frac{\text{Discordant read number}}{\text{Total number of reads}}$. Epipolymorphism of a given locus was calculated as $1 - \sum_{i=1}^{16} p_i^2$, where p_i is the fraction of each epiallele in the cell population (17). Shannon entropy of a given locus was calculated as $-\sum_{i=1}^{16} p_i \log p_i$, where p_i is the fraction of each epiallele in the cell population (20).

Gain and Loss of Epialleles

To increase the statistical power, we first filtered loci by a criterion independent of the test statistic. For human, the loci with an absolute value of the mean difference of epiallele diversity between patients in one AML subtype and NBM controls larger than 0.2 were used for further analysis. For mouse, the loci with an absolute value of the mean difference of epiallele diversity between samples with one mutation and wild-type samples larger than 0.05 were used for further analysis. Then, significantly differential distributions of epialleles between AML subtypes and NBM controls were assessed (PDR and epipolymorphism: *t* test; Shannon entropy: permutation testing in R package EntropyExplorer; ref. 50). Here, eloci were defined as loci with FDR-adjusted *P* values smaller than or equal to 0.1 (human and mouse) and an absolute value of the mean difference of epiallele diversity larger than 0.2 (human) or 0.05 (mouse). Gain of epialleles was defined as eloci with a mean difference of epiallele diversity larger than 0.2 (human) or 0.05 (mouse). Loss of epialleles was defined as eloci with a mean difference of epiallele diversity less than -0.2 (human) or -0.05 (mouse).

Disclosure of Potential Conflicts of Interest

S. Li is an associate editor of Science Advances. C. Meydan reports grants from NIH and grants from LLS during the conduct of the study; and personal fees from Onegevity Health outside the submitted work. J.L. Glass reports personal fees from Gerson Lehrman Group outside the submitted work. R.L. Levine reports other from Qiagen (board of directors), C4 (SAB), Auron (founder), Ajax (founder), Zentalis (SAB); personal fees from Incyte (consultant), Gilead (grant review), Lilly/Loxo (SAB), Morphosys (SAB), Celgene/BMS (consultant), Novartis (consultant); and grants from Prelude (SRA) and Constellation (SRA) outside the submitted work. C.E. Mason reports personal fees from Tempus Labs (advisor to Tempus Labs) during the conduct of the study. A.M. Melnick reports grants from NCI and Leukemia and Lymphoma Society during the conduct of the study, Janssen, and Daiichi Sankyo; personal fees from Epizyme, Constellation, Jubilant; and other from KDAC (scientific Board) outside the submitted work. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

S. Li: Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, investigation, visualization, methodology, writing-original draft, project administration, writing-review and editing. **X. Chen:** Data curation, formal analysis, investigation, visualization, methodology, writing-original draft, writing-review and editing. **J. Wang:** Investigation. **C. Meydan:** Resources. **J.L. Glass:** Resources. **A.H. Shih:** Resources. **R. Delwel:** Resources, data curation, writing-review and editing. **R.L. Levine:** Resources. **C.E. Mason:** Resources. **A.M. Melnick:** Conceptualization, resources, data curation, supervision, funding acquisition, investigation, writing-original draft, project administration, writing-review and editing.

Acknowledgments

We thank Stephen Sampson from The Jackson Laboratory for editing this manuscript. A.M. Melnick is funded by NCI 1UG CA233332, NCI R01 CA198089, LLS SCOR 7013-17, and The Chemotherapy Foundation. S. Li is supported by the National Institute of General Medical Sciences of the NIH under award number R35 GM133562, Leukemia Research Foundation New Investigator Grant, The Jackson Laboratory Director's Innovation Fund 19000-17-31 and 19000-20-05, and The Jackson Laboratory Cancer Center New Investigator Award. Research reported in this publication was partially supported by the NCI of the NIH under award number P30 CA034196. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. R. Delwel is supported by OncoCode Institute. The authors thank Francine Garrett-Bakelman and Peter J.M. Valk for providing the ELN information of the AML cohorts. The authors thank members of Li laboratory and Melnick laboratory for discussion, and the technique support of The Jackson Laboratory Computer Sciences team and Research Informatics Technology team.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 1, 2019; revised July 9, 2020; accepted September 11, 2020; published first September 16, 2020.

REFERENCES

- Ho TC, LaMere M, Stevens BM, Ashton JM, Myers JR, O'Dwyer KM, et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. *Blood* 2016;128:1671-8.
- Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012;481:506-10.
- Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013;368:2059-74.
- Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med* 2016;374:2209-21.
- Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep* 2014;9:1841-55.
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 2010;18:553-67.

7. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttill J, Zhang L, Khrebttukova I, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* 2012;8:e1002781.
8. Li S, Garrett-Bakelman F, Perl AE, Luger SM, Zhang C, To BL, et al. Dynamic evolution of clonal epialleles revealed by methclone. *Genome Biol* 2014;15:472.
9. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med* 2016;22:792–9.
10. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* 2010;17:13–27.
11. Shih AH, Jiang Y, Meydan C, Shank K, Pandey S, Barreyro L, et al. Mutational cooperativity linked to combinatorial epigenetic gain of function in acute myeloid leukemia. *Cancer Cell* 2015;27:502–15.
12. Shih AH, Meydan C, Shank K, Garrett-Bakelman FE, Ward PS, Intlekofer AM, et al. Combination targeted therapy to disrupt aberrant oncogenic signaling and reverse epigenetic dysfunction in IDH2- and TET2-mutant acute myeloid leukemia. *Cancer Discov* 2017;7:494–505.
13. Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, et al. CpG island hypermethylation mediated by DNMT3A is a consequence of AML progression. *Cell* 2017;168:801–16 e13.
14. De S, Shaknovich R, Riester M, Elemento O, Geng H, Kormaksson M, et al. Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet* 2013;9:e1003137.
15. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 2014;26:813–25.
16. Pan H, Jiang Y, Boi M, Tabbo F, Redmond D, Nie K, et al. Epigenomic evolution in diffuse large B-cell lymphomas. *Nat Commun* 2015;6:6921.
17. Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet* 2012;44:1207–14.
18. Glass JL, Hassane D, Wouters BJ, Kunimoto H, Avellino R, Garrett-Bakelman FE, et al. Epigenetic identity in AML depends on disruption of nonpromoter regulatory elements and is affected by antagonistic effects of mutations in epigenetic modifiers. *Cancer Discov* 2017;7:868–83.
19. Figueroa ME, Wouters BJ, Skrabanek L, Glass J, Li Y, Erpelinck-Verschueren CA, et al. Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood* 2009;113:2795–804.
20. Sherwin WB. Entropy and information approaches to genetic diversity and its expression: genomic geography. *entropy* 2010;12:1765–98.
21. Bagger FO, Sasivarevic D, Sohi SH, Laursen LG, Pundhir S, Sonderby CK, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res* 2016;44:D917–24.
22. Hilger-Eversheim K, Moser M, Schorle H, Buettner R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene* 2000;260:1–12.
23. Hoffman B, Amanullah A, Shafarenko M, Liebermann DA. The proto-oncogene c-myc in hematopoietic development and leukemogenesis. *Oncogene* 2002;21:3414–21.
24. Wang Z, Bunting KD. STAT5 in hematopoietic stem cell biology and transplantation. *JAKSTAT* 2013;2:e27159.
25. Hoelbl A, Schuster C, Kovacic B, Zhu B, Wickre M, Hoelzl MA, et al. Stat5 is indispensable for the maintenance of bcr/abl-positive leukemias. *EMBO Mol Med* 2010;2:98–110.
26. Zhou J, Ching YQ, Chng WJ. Aberrant nuclear factor-kappa B activity in acute myeloid leukemia: from molecular pathogenesis to therapeutic target. *Oncotarget* 2015;6:5490–500.
27. Ward AC, Touw I, Yoshimura A. The jak-stat pathway in normal and perturbed hematopoiesis. *Blood* 2000;95:19–29.
28. Benekli M, Xia Z, Donohue KA, Ford LA, Pixley LA, Baer MR, et al. Constitutive activity of signal transducer and activator of transcription 3 protein in acute myeloid leukemia blasts is associated with short disease-free survival. *Blood* 2002;99:252–7.
29. Mehta-Grigoriou F, Gerald D, Yaniv M. The mammalian Jun proteins: redundancy and specificity. *Oncogene* 2001;20:2378–89.
30. Rangatia J, Vangala RK, Singh SM, Peer Zada AA, Elsasser A, Kohlmann A, et al. Elevated c-Jun expression in acute myeloid leukemias inhibits C/EBPalpha DNA binding via leucine zipper domain interaction. *Oncogene* 2003;22:4760–4.
31. Zhou C, Martinez E, Di Marcantonio D, Solanki-Patel N, Aghayev T, Peri S, et al. JUN is a key transcriptional regulator of the unfolded protein response in acute myeloid leukemia. *Leukemia* 2017;31:1196–205.
32. Chacon D, Beck D, Perera D, Wong JW, Pimanda JE. BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Res* 2014;42:D172–7.
33. Mandoli A, Singh AA, Jansen PW, Wierenga AT, Riahi H, Franci G, et al. CBFb-MYH11/RUNX1 together with a compendium of hematopoietic regulators, chromatin modifiers and basal transcription factors occupies self-renewal genes in inv(16) acute myeloid leukemia. *Leukemia* 2014;28:770–8.
34. Martens JH, Brinkman AB, Simmer F, Francois KJ, Nebbioso A, Ferrara F, et al. PML-RARalpha/RXR alters the epigenetic landscape in acute promyelocytic leukemia. *Cancer Cell* 2010;17:173–85.
35. Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA, van Putten WL, Valk PJ, Delwel R. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* 2009;113:3088–91.
36. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011;20:11–24.
37. Ficara F, Murphy MJ, Lin M, Cleary ML. Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell* 2008;2:484–96.
38. Honey K. Keeping HSCs under control. *Nat Rev Immunol* 2004;4:834.
39. Yu S, Li F, Xing S, Zhao T, Peng W, Xue HH. Hematopoietic and leukemic stem cells have distinct dependence on Tcf1 and Lef1 transcription factors. *J Biol Chem* 2016;291:11148–60.
40. Bossone SA, Asselin C, Patel AJ, Marcu KB. MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination. *Proc Natl Acad Sci U S A* 1992;89:7452–6.
41. Metzelder SK, Michel C, von Bonin M, Rehberger M, Hessmann E, Inselmann S, et al. NFATc1 as a therapeutic target in FLT3-ITD-positive AML. *Leukemia* 2015;29:1470–7.
42. Solovey M, Wang Y, Michel C, Metzelder KH, Herold T, Gothert JR, et al. Nuclear factor of activated T-cells, NFATc1, governs FLT3(ITD)-driven hematopoietic stem cell transformation and a poor prognosis in AML. *J Hematol Oncol* 2019;12:72.
43. Faber K, Bullinger L, Ragu C, Garding A, Mertens D, Miller C, et al. CDX2-driven leukemogenesis involves KLF4 repression and deregulated PPARGgamma signaling. *J Clin Invest* 2013;123:299–314.
44. Figueroa ME, Skrabanek L, Li Y, Jiemjit A, Fandy TE, Paietta E, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood* 2009;114:3448–58.
45. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;129:424–47.

46. Luskin MR, Gimotty PA, Smith C, Loren AW, Figueroa ME, Harrison J, et al. A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia. *JCI Insight* 2016;16:9.
47. Meldi K, Qin T, Buchi F, Droin N, Sotzen J, Micol JB, et al. Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. *J Clin Invest* 2015;125:1857-72.
48. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2004;2:E108.
49. Teater M, Dominguez PM, Redmond D, Chen Z, Ennishi D, Scott DW, et al. AICDA drives epigenetic heterogeneity and accelerates germinal center-derived lymphomagenesis. *Nat Commun* 2018; 9:222.
50. Wang K, Phillips CA, Saxton AM, Langston MA. EntropyExplorer: an R package for computing and comparing differential Shannon entropy, differential coefficient of variation and differential expression. *BMC Res Notes* 2015;8:832.

CANCER DISCOVERY

Somatic Mutations Drive Specific, but Reversible, Epigenetic Heterogeneity States in AML

Sheng Li, Xiaowen Chen, Jiahui Wang, et al.

Cancer Discov 2020;10:1934-1949. Published OnlineFirst September 16, 2020.

Updated version	Access the most recent version of this article at: doi: 10.1158/2159-8290.CD-19-0897
Supplementary Material	Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2020/09/16/2159-8290.CD-19-0897.DC1

Cited articles	This article cites 49 articles, 12 of which you can access for free at: http://cancerdiscovery.aacrjournals.org/content/10/12/1934.full#ref-list-1
Citing articles	This article has been cited by 2 HighWire-hosted articles. Access the articles at: http://cancerdiscovery.aacrjournals.org/content/10/12/1934.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/10/12/1934 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.