Transcriptional Silencing of ALDH2 Confers a Dependency on Fanconi Anemia Proteins in Acute Myeloid Leukemia

Zhaolin Yang1, Xiaoli S. Wu1,2, Yiliang Wei1, Sofya A. Polyanskaya1, Shruti V. Iyer1,2, Moonjung Jung3, Francis P. Lach3, Emmalee R. Adelman4, Olaf Klingbeil1, Joseph P. Milazzo1, Melissa Kramer1, Osama E. Demerdash1, Kenneth Chang1, Sara Goodwin1, Emily Hodges5, W. Richard McCombie1, Maria E. Figueroa4, Agata Smogorzewska3, and Christopher R. Vakoc1

ABSTRACT

Hundreds of genes become aberrantly silenced in acute myeloid leukemia (AML), with most of these epigenetic changes being of unknown functional consequence. Here, we demonstrate how gene silencing can lead to an acquired dependency on the DNA repair machinery in AML. We make this observation by profiling the essentiality of the ubiquitination machinery in cancer cell lines using domain-focused CRISPR screening, which revealed Fanconi anemia (FA) proteins UBE2T and FANCL as unique dependencies in AML. We demonstrate that these dependencies are due to a synthetic lethal interaction between FA proteins and aldehyde dehydrogenase 2 (ALDH2), which function in parallel pathways to counteract the genotoxicity of endogenous aldehydes. We show DNA hypermethylation and silencing of ALDH2 occur in a recurrent manner in human AML, which is sufficient to confer FA pathway dependency. Our study suggests that targeting of the ubiquitination reaction catalyzed by FA proteins can eliminate ALDH2-deficient AML.

SIGNIFICANCE: Aberrant gene silencing is an epigenetic hallmark of human cancer, but the functional consequences of this process are largely unknown. In this study, we show how an epigenetic alteration leads to an actionable dependency on a DNA repair pathway through the disabling of genetic redundancy.

INTRODUCTION

Organisms evolve redundant genetic pathways for carrying out essential cellular processes. This often arises from gene duplication events, which produce paralogs that carry out overlapping functions (1). Alternatively, nonhomologous gene pairs can be redundant if they encode parallel pathways that regulate a shared cellular process (2). While genetic redundancies support robustness during normal processes, cancer cells often lack such redundancies owing to genetic or epigenetic alterations, a process often referred to as synthetic lethality (3, 4). Because of its therapeutic significance, the identification of synthetic lethal genetic interactions remains an important objective in the study of human cancer (5).

1Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 2Genetics Program, Stony Brook University, Stony Brook, New York. 3Laboratory of Genome Maintenance, The Rockefeller University, New York, New York. 4Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, Florida. 5Department of Biochemistry and Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, Tennessee.

Note: Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).
In recent years, there has been a resurgent interest in the ubiquitination machinery as a target for cancer therapy. Ubiquitin is a 8.6-kDa protein that is covalently attached to lysine side chains as a form of posttranslational regulation of protein stability and function (6). The cascade reaction of ubiquitination requires the consecutive action of three enzymes: a ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin ligases (E3; ref. 7). With approximately 40 E2 and more than 600 putative E3 proteins encoded in the human genome, the ubiquitination machinery regulates many aspects of cell biology, including the cell cycle, DNA repair, and transcription (8). In addition, it has been established that the function of E2 and E3 proteins can be modulated with small molecules (9, 10). For example, the E3 ligase protein MDM2 can be inhibited with small molecules to stabilize p53 and promote apoptosis of cancer cells (11). In addition, small molecules have been developed that alter the specificity of E3 ligases to trigger proteasome-mediated degradation of neosubstrates (12, 13). Despite the clear therapeutic potential of the ubiquitination machinery, there have been few efforts to date aimed at identifying E2/E3 dependencies in cancer cells using genetic screens.

The Fanconi anemia (FA) pathway, comprised of more than 20 protein components, repairs DNA interstrand cross-links (ICL; ref. 14). A key step in the activation of the FA repair pathway is the monoubiquitination of FANCD2 and FANCI, which is performed by UBE2T/FANCT (an E2-conjugating enzyme) and FANCL (an E3 ligase; refs. 14, 15). Once ubiquitinated, a FANCD2/FANCI heterodimer encircles the DNA (16, 17), which facilitates lesion processing by nucleases and completion of the repair by homologous recombination (14, 18). While DNA cross-links can be caused by exogenous mutagens, emerging evidence suggests that endogenously produced aldehydes are an important source of this form of DNA damage (19). Loss-of-function mutations in FA genes lead to Fanconi anemia, a genetic disorder characterized by bone marrow failure and a predisposition to leukemia and aerodigestive tract squamous cell carcinoma (15). While a deficiency in FA proteins is cancer-promoting in certain tissue contexts, for the majority of human cancers, the FA pathway remains intact and therefore may perform an essential function in established cancers. To our knowledge, a role of FA proteins as cancer dependencies has yet to be identified.

Here, we discovered an acquired dependency on FA proteins in a subset of acute myeloid leukemia (AML) that lacks...
expression of aldehyde dehydrogenase 2 (ALDH2). We show that epigenetic silencing of ALDH2 occurs in a recurrent manner in human AML and is sufficient to confer FA protein dependency in this disease. Our study suggests that blocking the UBE2T/FANCL-mediated ubiquitination can selectively eliminate ALDH2-deficient AML cells while sparing ALDH2-expressing normal cells present in the majority of human tissues.

RESULTS

A Domain-Focused CRISPR Screen Targeting the Ubiquitination Machinery Identifies UBE2T and FANCL as AML-Biased Dependencies

In this study, we pursued the identification of AML-specific dependencies on the ubiquitination machinery using domain-focused CRISPR single-guide RNA (sgRNA) screening, which is a strategy for profiling the essentiality of protein domain families in cancer cell lines (20). Using an sgRNA design algorithm linked to protein domain annotation, we designed 6,060 sgRNAs targeting exons encoding 564 domains known to be involved in ubiquitin conjugation or ligation, which were cloned in a pooled manner into the LRG2.1T lentiviral vector (Fig. 1A; Supplementary Table S1). Using this sgRNA library, we performed negative selection “dropout” screening in 12 Cas9-expressing human cancer cell lines, which included six AML and six solid tumor cell lines (Fig. 1B; Supplementary Table S2). Many of the dependencies identified in these screens were pan-essential across the 12 lines, such as ANAPC11, CDC16, and RBX1 (Supplementary Fig. S1A). We ranked all ubiquitination-related genes based on their degree of essentiality in AML versus solid tumor contexts, which nominated UBE2T and FANCL as AML-specific dependencies (Fig. 1C and D). The known function of UBE2T and FANCL as E2 and E3 proteins, respectively, within the FA pathway suggested a unique necessity of this DNA repair function in AML (14, 15).

We corroborated the biased essentiality of UBE2T/FANCL in blood cancers relative to other cancer types by analyzing data obtained from Project Achilles (DepMap.org), in which genome-wide CRISPR essentiality screening was performed in 729 cancer cell lines (Fig. 1E; Supplementary Fig. S1B; refs. 21, 22). In addition, these data revealed that several other FA pathway genes, including FANCI, FANCB, and FANCC, are also blood cancer–biased dependencies in a manner that correlated with UBE2T/FANCL essentiality. These findings reinforce that blood cancer cells are hypersensitive to perturbation of FA proteins relative to other cancer types.

FA Proteins Are Dependencies in a Subset of AML Cell Lines under In Vitro and In Vivo Conditions

To further validate the specificity of FA protein dependencies in AML, we performed sgRNA competition assays following inactivation of UBE2T, FANCL, and FANCD2 genes in 27 human cancer cell lines, including 14 leukemia, 5 pancreatic cancer, 4 lung cancer, and 4 sarcoma lines (Fig. 2A). These experiments showed that targeting any of the three FA genes suppressed the fitness of nine human leukemia lines, including three generated by retroviral transduction of oncogenes into human hematopoietic stem and progenitor cells (23). In contrast, the fitness of five other human leukemia lines and all of the solid tumor cell lines was less sensitive to targeting of FA genes (Fig. 2A). As a positive control for this assay, targeting of CDK1 arrested the growth of all cancer cell lines tested. Western blotting confirmed that the variable pattern of growth arrest following UBE2T targeting was not due to differences in genome editing efficiency (Fig. 2B; Supplementary Fig. S2A and S2B).

As additional controls, we verified that the growth arrest caused by UBE2T or FANCL knockout in MOLM-13 cells was due to an on-target effect by rescuing this effect with a cDNA carrying silent mutations that abolish sgRNA recognition (Fig. 2C; Supplementary Fig. S2C–S2F). Using this rescue assay, we investigated whether the catalytic function of UBE2T was essential for AML growth by comparing the wild-type cDNA with the C86A mutation, which abolishes ubiquitin conjugation activity (24, 25). Despite being expressed at a similar level to the wild-type protein, the C86A mutant of UBE2T was unable to support AML growth, suggesting that the ubiquitination cascade involving UBE2T–FANCL is essential in this context (Fig. 2D; Supplementary Fig. S2G and S2H). To ensure that this dependency was not influenced by DNA damage caused by CRISPR/Cas9, we also validated FANCD2 dependence in AML using RNAi-based knockdown (Fig. 2E; Supplementary Fig. S2J). In addition, inactivation of UBE2T, FANCL, or FANCD2 suppressed the growth of MOLM-13 cells and extended animal survival when propagated in vivo in immune-deficient mice (Fig. 2F; Supplementary Fig. S2J; Supplementary Fig. S3A–S3C). The survival benefit observed in these experiments was limited, at least in part, by incomplete genome editing and the preferential outgrowth of leukemia cells lacking FA gene mutations (Supplementary Fig. S3D–S3G).

Targeting of FA Proteins in AML Leads to Cell-Cycle Arrest and Apoptosis in a Partially p53-Dependent Manner

We next performed a deeper characterization of the cellular phenotype following inactivation of FA proteins in AML. A flow cytometry analysis of BrdU incorporation and DNA content revealed an accumulation of UBE2T- and FANCD2-deficient MOLM-13 cells in G2–M phase, with a significant loss of cells in S-phase (Fig. 3A and B). An Annexin V staining analysis of UBE2T- or FANCD2-deficient MOLM-13 cells also revealed evidence of apoptosis (Fig. 3C and D). The levels of γH2AX and of metaphase chromosome breaks, both markers of DNA damage, were also increased following FA gene inactivation (Supplementary Fig. S4A–S4C). As apoptosis and cell-cycle arrest are known downstream consequences of DNA damage–induced p53 activation, we investigated this pathway in FA-deficient AML cells. Although FA dependency is present in both TP53 wild-type and TP53-mutant AML cell lines, we noticed that the growth arrest phenotype tended to be stronger in the TP53 wild-type context (Fig. 2A) and in the p53-competent MOLM-13 (Fig. 3E; Supplementary Fig. S5B). In addition, we observed induction of proapoptotic genes and suppression of DNA replication genes following FA protein inactivation, in accord with the phenotypes described...
Figure 1. Domain-focused CRISPR screening identifies UBE2T and FANCL as AML-biased dependencies. A, Categories of E2 and E3 domains targeted in the ubiquitination domain sgRNA library. B, Schematic of CRISPR/Cas9 screening strategy. C, Ubiquitination machinery dependencies ranked by the average log2 fold-change difference comparing six AML and six non-AML cell lines in the pooled CRISPR screening. D, Average log2 fold-change for sgRNAs targeting UBE2T, FANCL, and ANAPC11 in the 12 screened cell lines. E, Dependency scores of five FA genes extracted from Project Achilles (2Q2; refs. 21, 22). The box plots indicate the distribution of dependency score (CERES, a normalized metric of gene essentiality) of five FA genes across all 769 cell lines or in the subset of 41 blood cell lines. P value was calculated using an unpaired Student t test. All sgRNA experiments were performed in Cas9-expressing cell lines.
above (Fig. 3E). In contrast, these transcriptional changes did not occur in the TP53-mutant AML line SEM, but instead we found that transcription of genes related to metabolism was disproportionate suppressed following FA protein inactivation (Fig. 3E; Supplementary Fig. SSB and SSC). To evaluate the contribution of p53 to FA essentiality in AML, we targeted TP53 or its target gene CDKN1A in several AML lines using CRISPR and found that this alleviated the UBE2T dependency in TP53 wild-type contexts (Fig. 3F and G; Supplementary Fig. S6A–S6G). Taken together, these findings indicate that loss of FA proteins in AML triggers DNA damage and growth arrest through a combination of p53-dependent and p53-independent mechanisms.

**Low Aldefluor Activity and ALDH1A1/ALDH2 Expression in FA-Dependent AML Cell Lines**

We next hypothesized that an endogenous source of DNA damage might drive the elevated demand for FA proteins in AML. For example, excess accumulation of aldehydes can exacerbate the phenotypes of FA-deficient mice and humans (19, 26, 27). This led us to investigate the status of ALDHs, 19 enzymes that oxidize diverse aldehydes into nontoxic acetates in a NAD- or NADP-dependent manner (19). We first made use of the Aldefluor assay, in which cells are treated with BODIPY-aminoc_acetaldehyde (BAAA), a cell-permeable fluorescent aldehyde that becomes trapped in cells following ALDH-mediated conversion to BODIPY-aminocacetate (BAA). After applying BAAA to a diverse collection of human AML cell lines, we observed that most FA-dependent lines exhibited minimal ALDH activity, with fluorescence levels similar to control cells treated with the pan-ALDH inhibitor N,N-diethylaminobenzaldehyde (DEAB). In contrast, ALDH activity was higher in most of the FA-dispensable group of AML lines (Fig. 4A). One exception to this correlation was U937 cells, which are FA dispensable, but lack ALDH activity. However, we note that U937 lacks a functional p53 pathway (28), which would be expected to alleviate the FA dependence in this context.

An RNA-seq analysis revealed that among the 19 ALDH enzymes, only ALDH1A1 and ALDH2 were expressed at reduced levels in the FA-dependent lines when compared with FA-disposable lines (Fig. 4B; Supplementary Fig. S7A). Using Western blotting, we confirmed that ALDH1A1 and ALDH2 proteins are present at higher levels in the FA-disposable versus the FA-dependent AML lines (Fig. 4B). Of note, while ALDH1A1 and ALDH2 are both capable of oxidizing BAAA, each enzyme is known to have distinct cellular functions (29–35). ALDH1A localizes in the cytosol and is known to oxidize retinol aldehydes into retinoic acids (29). In contrast, ALDH2 is localized in the mitochondria and oxidizes acetaldehyde (derived from exogenous ethanol; ref. 32).

**Reexpression of ALDH2, but Not ALDH1A1, Rends FA Proteins Dispensable for AML Growth**

Considering the inverse correlation between ALDH1A1/ALDH2 expression and FA dependence in AML lines, we next performed experiments to explore a synthetic lethal genetic interaction in this context. We lentivirally transduced the FA-dependent MOLM-13 line with ALDH1A1 or ALDH2 cDNA and confirmed protein expression via Western blotting (Fig. 4C). Aldefluor assays revealed that the levinvically expressed proteins were enzymatically active, with both ALDH1A1 and ALDH2 causing oxidation of BAAA (Fig. 4D), consistent with prior findings (26, 36). We next used CRISPR to target UBE2T in the ALDH1A1- or ALDH2-expressing MOLM-13 cells, followed by competition-based assays to track changes in cell fitness. Remarkably, the ALDH2-expressing MOLM-13 cells became resistant to growth arrest caused by UBE2T inactivation, whereas ALDH1A1-expressing cells remained sensitive to UBE2T targeting (Fig. 4E). To evaluate the generality of this result, we expressed ALDH1A1 and ALDH2 in two other FA-dependent AML contexts, which likewise showed that ALDH2, but not ALDH1A1, alleviated the dependency on FA proteins (Fig. 4E; Supplementary Fig. S7B and S7C). To address whether inactivation of ALDH2 is sufficient to confer FA dependence in AML, we made use of the murine AML cell line RN2, which was derived by retroviral transformation of hematopoietic stem and progenitor cells with the MLL-ARF and NRASCD112 CDNs (37). Notably, this cell line retains an intact p53 pathway and expresses ALDH2, but not ALDH1A1 (Supplementary Fig. S7D). We targeted ALDH2 using CRISPR in this cell line and confirmed loss of protein expression and loss of Aldefluor activity (Fig. 4F and G). Notably, in competition-based cell fitness assays, we observed that ALDH2-deficient RN2 cells became hypersensitive to the inactivation of FANC2 relative to the control cells (Fig. 4H). By segregating all human cancer solid tumor cell lines profiled in Project Achilles based on ALDH2 expression, we find that the FA gene essentiality is significantly correlated with low ALDH2 expression in several nonhematopoietic cancer lineages (Supplementary Fig. S7E and S7F). This suggests that the link between ALDH2 silencing and FA dependency extends to other cancer types beyond AML. Together, these experiments suggest that loss of ALDH2 expression confers dependency on FA proteins in AML.

We next investigated whether the catalytic function of ALDH2 was needed for the bypass of FA pathway dependency by comparing the wild-type ALDH2 cDNA with the E268K

---

**Figure 2.** FA proteins are AML-based dependencies in vitro and in vivo. A, Competition-based proliferation assays performed in 27 human cancer cell lines by lentiviral infection with the indicated sgRNAs linked to GFP Timepoints (T) were collected every 2 days for leukemia cells or every 3 days for nonleukemia lines. Rate of GFP depletion indicates loss of cell fitness caused by Cas9sgRNA-mediated genetic mutations. PDAC, pancreatic ductal adenocarcinoma; SCLC, small-cell lung cancer; n = 3. B, Western blotting performed on whole-cell lysates prepared from indicated cell lines, performed on day 4 following lentiviral transduction with indicated sgRNAs. C, Competition-based proliferation assays comparing impact of sgRNAs (linked with GFP) on MOLM-13 cell fitness, in the context of cotransduction with UBE2T WT or sgRNA-resistant UBE2T cDNA. The sgRNA-resistant CDA1 mutagenesis strategy is depicted (n = 3). D, Competition-based assays measuring sRNA (linked to GFP) effects on MOLM-13 cell fitness, in the presence of the indicated cDNAs (n = 3). E, Competition-based proliferation assays in MOLM-13 cells lentivirally infected with indicated shRNA (linked to GFP) targeting FANC2 (n = 3). F, Left, bioluminescence imaging of NSG mice transplanted with luciferase/Cas9 MOLM-13 cells infected with either sgNEG or sgFANC2 Right, quantification of bioluminescence intensity (n = 5). All bar graphs represent the mean ± SEM. P value is calculated by unpaired Student t test. *** P ≤ 0.001, ** P < 0.01, * P < 0.05. All sgRNA experiments were performed in Cas9-expressing cell lines.
Figure 3. Inactivation of FA genes in AML leads to p53-induced cell-cycle arrest and apoptosis. A, Representative flow cytometry analysis of BrdU incorporation and DNA content to infer cell status following lentiviral transduction of MOLM-13 cells with the indicated sgRNAs (day 6). B, Quantification of live and apoptotic cells, average of three biological replicates. Paired Student t test was applied to calculate P values. C, Representative flow cytometry analysis of DAPI (indicating permeable dead cells) and Annexin-V staining (a preapoptotic cell marker) following lentiviral transduction of MOLM-13 cells with the indicated sgRNAs (day 6). D, Quantification of live and apoptotic cells, average of three biological replicates. Paired Student t test was applied to calculate P values. E, Gene set enrichment analysis of RNA-seq data obtained from MOLM-13 cells lentivirally transduced with the indicated sgRNAs (70). Normalized enrichment score (NES) and family-wise error rate (FWER) P value are shown. F, Western blot analysis performed on lysates obtained from MOLM-13 cells on day 6 following sgRNA transduction. G, Competition-based proliferation assays in MOLM-13 cells following sequential sgRNA transduction. Negative sgRNA or TP53 sgRNAs were infected first, selected with neomycin, followed by transduction with the sgRNAs indicated at the bottom of the graph (linked with GFP). n = 3. All bar graphs represent the mean ± SEM. All sgRNA experiments were performed in Cas9-expressing cell lines.
Figure 4. Silencing of ALDH2 expression in AML confers FA protein dependency. **A**, Representative flow cytometry analysis of Aldefluor assays, in which AML cell lines were treated with 3.75 μmol/L BAAA, with or without the ALDH inhibitor DEAB (15 μmol/L). **B**, RNA and protein levels of ALDH1A1 or ALDH2 measured by RNA-seq (top) and Western blot (bottom) in the indicated cell lines. **C**, Western blot analysis of whole-cell lysates from cell lines measured by RNA-seq (top) and Western blot (bottom) in the indicated cell lines. **D**, Representative Aldefluor analysis of MOLM-13 cells expressing FLAG-tagged ALDH1A1 or ALDH2. **E**, Competition-based proliferation assay in MOLM-13 expressing the indicated cDNAs, followed by transduction with GFP-linked sgRNA. **F**, Western blot analysis of ALDH2 expression in RN2 cells following lentiviral transduction with the indicated sgRNAs. **G**, Aldefluor assay performed in RN2 cells following lentiviral transduction with the indicated sgRNAs. **H**, Competition-based proliferation assays performed in RN2 cells (control vs. Aldh2 knockout) with the indicated sgRNAs linked to GFP. All bar graphs represent the mean ± SEM, (n = 3). All sgRNA experiments were performed in Cas9-expressing cell lines.
and the C302A alleles (38). Importantly, these two mutant proteins were expressed normally, but lacked any detectable Aldefluor activity. In addition, both mutants were unable to rescue the UBE2T dependence of MOLM-13 cells (Supplementary Fig. S8A–S8D). As an additional control, we considered whether ALDH1A1 might be capable of rescuing the FA dependence if we forced its localization into the mitochondria using two different localization signals (Supplementary Fig. S9A). Despite effective targeting to the mitochondria confirmed by cell fractionation and robust Aldefluor activity of these mutant proteins (Supplementary Fig. S9B and S9C), the mitochondrial forms of ALDH1A1 were unable to rescue UBE2T dependence (Supplementary Fig. S9D). Other ALDH enzymes (ALDH1A2, ALDH1B1, ALDH6A1, ALDH7A, or ALDH1A3) were also tested in this assay, but were unable to rescue the UBE2T dependence when lentivirally expressed in MOLM-13 cells (Supplementary Fig. S9E–S9H). Taken together, these data suggest a unique capability of ALDH2 to detoxify a specific subset of endogenous aldehydes that drive FA protein dependency in AML.

Interestingly, restoration of ALDH2 expression in MOLM-13 cells led to no detectable impact on cell proliferation in vitro (Supplementary Fig. S10A). In addition, an RNA-seq analysis suggested that the overall transcriptome of MOLM-13 cells was largely unaffected by reexpression of ALDH2 (Supplementary Fig. S10B). Using fluorometric assays to measure endogenous aldehyde levels, we found that restoring ALDH2 expression, but not ALDH1A1 expression, in MOLM-13 cells led to a reduction of malondialdehyde, whereas expression of either enzyme was able to suppress formaldehyde levels (Supplementary Fig. S10C and S10D). While additional aldehyde substrates of ALDH2 might also be relevant, our findings suggest that loss of ALDH2 in AML leads to an accumulation of malondialdehyde, which is known to cause interstrand DNA cross-links.

Silencing and Hypermethylation of ALDH2 Occurs in a Recurrent Manner in Human AML

We next investigated whether ALDH2 silencing is associated with DNA hypermethylation, which is known to be aberrantly distributed across the AML genome (39). In AML cell lines profiled in the Cancer Cell Line Encyclopedia (CCLE) project (28, 40), we detected an inverse correlation between levels of DNA methylation and ALDH2 expression (Fig. 5A). To confirm this finding, we analyzed DNA methylation at the ALDH2 promoter in MOLM-13 and MV4–11 cells using Nanopore sequencing (41), which confirmed dense hypermethylation in the vicinity of the ALDH2 promoter (Fig. 5B). Importantly, this same location was hypomethylated in normal human hematopoietic stem and progenitor cells (Fig. 5B; ref. 42). In addition, silencing of ALDH2 correlated with diminished histone acetylation at this genomic region (Fig. 5C). Using an inhibitor of DNA methyltransferase activity 5-azacytidine, we confirmed a time- and dose-dependent increase in ALDH2 expression in MOLM-13 and MV4–11 cells following compound exposure (Fig. 5D and E; Supplementary Fig. S11A and S11B). Consistent with this finding, 5-azacytidine-treated MOLM-13 cells were also significantly less sensitive to UBE2T inactivation (Supplementary Fig. S11C). These results suggest that silencing of ALDH2 in AML cell lines is associated with the acquisition of DNA hypermethylation.

We next investigated whether silencing and hypermethylation of ALDH2 occurs in human AML patient specimens. By analyzing RNA-seq data obtained from diverse tumors included in the TCGA pan-cancer analysis, we found that AML had greater variability in ALDH2 expression when compared with most other human cancer types (Fig. 5F). Within this group of TCGA patient samples, we designated ALDH2-low and ALDH2-high AML, distinguished by an approximately 100-fold difference in ALDH2 expression (Fig. 5G). Notably, we observe that a housekeeping gene, ACTB, showed only a 4-fold variance in expression across TCGA AML samples (Supplementary Fig. S11D). In accord with cell line observations, ALDH2-low AML samples in the TCGA possessed elevated levels of DNA methylation at ALDH2 relative to ALDH2-high AML samples (Fig. 5H). We further confirmed the aberrant hypermethylation and silencing of ALDH2 in a subset of AML in an independent collection of patient samples (39), whereas normal bone marrow cells remained hypomethylated (Supplementary Fig. S11E–S11G). Using the TCGA datasets, we found that ALDH2-low AML samples are enriched for mutations in NPM1, CEBPA, and the RUNXI–RUNXIT1 (AML1–ETO) translocation, whereas ALDH2-high AML is enriched for TP53 and KMT2A alterations (Fig. 5I). Together, these findings suggest that DNA hypermethylation and silencing of ALDH2 occurs in a recurrent manner in human AML.

DISCUSSION

Using a genetic screen focused on the ubiquitination machinery, we uncovered a role for FA proteins as dependencies in AML. We account for this observation by the aberrant expression of ALDH2, an enzyme that oxidizes aldehydes in normal tissues but becomes epigenetically silenced in this disease context. We propose that silencing of ALDH2 in AML leads to an accumulation of endogenous aldehydes, which in turn causes the formation of DNA cross-links that necessitate repair by FA proteins to maintain cell fitness. Upon inactivation of FA genes in ALDH2-deficient AML, the level of aldehyde-induced DNA damage reaches a threshold that triggers p53-mediated cell-cycle arrest and cell death. This study reinforces how aberrant gene silencing can disable redundant pathways and can lead to acquired dependencies in cancer.

The synthetic lethal interaction between ALDH2 and FA genes is well supported by observations in mice, in which the combined deficiency of Aldh2 and Fancl2 leads to developmental defects, a predisposition to cancer, and a hypersensitivity to exogenous aldehydes (19). These phenotypes were thought to be uniquely present in normal hematopoietic stem and progenitor cells (26), but our study now shows that this genetic interaction extends to malignant myeloid cells. In normal mouse hematopoietic stem cells, aldehyde-induced DNA damage leads to the formation of double-stranded DNA breaks, which ultimately causes a stem-cell attrition phenotype that resembles the clinical presentation of germline FA deficiency in humans (43). Similar to our observations in AML, hematopoietic stem cells in Aldh2/Fancl2 compound-deficient mice become cleared through a p53-dependent mechanism (43). The redundant function of ALDH2 and FA proteins is also supported by evidence...
Figure 5. Silencing of ALDH2 in human AML cell lines and patient samples is associated with promoter DNA hypermethylation. A, Scatter plot analysis comparing ALDH2 mRNA levels and DNA methylation levels at the ALDH2 promoter across 32 AML cell lines (data was obtained from the CCLE; refs. 28, 40). P value is calculated by linear regression. B, DNA methylation analysis of the ALDH2 promoter in leukemia cell lines (Nanopore sequencing) or in normal human hematopoietic stem and progenitor cells (bisulfite sequencing from ref. 42). C, Chromatin immunoprecipitation sequencing analysis of H3K27ac enrichment at the ALDH2 promoter. D and E, RT-qPCR analysis of ALDH2 mRNA levels in MOLM-13 cells following treatment with 5-azacytidine. In D, 1 μmol/L concentration was used. In E, a 36-hour timepoint was used. F, ALDH2 mRNA levels extracted from the TCGA Pan-Can Atlas (71). G, ALDH2 mRNA levels in 173 AML patient samples from TCGA. Shown are the samples classified as ALDH2-high and ALDH2-low. H, Comparison of CpG DNA methylation level at the ALDH2 promoter of indicated AML patient samples. P value was calculated by unpaired Student t test. I, Genetic alterations that correlate with ALDH2 silencing in human patients with AML. ALDH2-low and ALDH2-high patient samples were extracted from the TCGA Pan-Cancer database, and the genetic alteration frequencies were evaluated in these two groups of patients. An unpaired Student t test was performed to determine statistical significance, which revealed the five genes shown as being significantly correlated or anticorrelated with ALDH2 silencing. All bar graphs represent the mean ± SEM (n = 3).
in humans, in which a combined hereditary deficiency of ALDH2 and FA genes correlates with an accelerated rate of bone marrow failure when compared with FA patients with intact ALDH2 function (27). Thus, our study builds upon prior work by revealing a clinical context in which the ALDH2/FA protein redundancy is disabled in humans and could be exploited to eliminate AML in vivo.

Several prior studies have applied Aldelfluor assays to human clinical samples and observed diminished ALDH activity in AML when compared with normal hematopoietic stem and progenitor cells, in accord with our own findings (44–47). Loss of ALDH1A1 expression was previously observed in AML, which correlated with favorable prognosis and was found to render cells hypersensitive to toxic ALDH substrates, such as arsanic trioxide (44). Our study validates ALDH1A1 silencing in AML cell lines; however, our functional experiments demonstrate that this event is unrelated to FA dependency. Collectively, our study and the work of Gasparetto and colleagues (44) suggest distinct functional consequences upon loss of ALDH1A1 versus ALDH2 in AML.

Our study points to the existence of an endogenous genotoxic aldehyde species that is uniquely oxidized by ALDH2 in AML. Although high reactivity precluded us from performing an unbiased assessment of aldehyde species in AML cells, our targeted analysis points to malondialdehyde as a potential substrate unique to ALDH2. However, it is also known that 4-HNE levels are elevated in ALDH2-deficient mice and humans (35, 48). 4-HNE and malondialdehyde are both by-products of lipid peroxidation and can form toxic adducts with DNA and proteins in a variety of cellular pathologies (49, 50). An important objective for future investigation will be to establish definitively the endogenous aldehyde source that drives the demand for FA and ALDH2 in AML.

Our experiments suggest that ALDH2 silencing has no measurable impact on AML cell fitness, owing to compensation via FA proteins. Why then is ALDH2 recurrently silenced in AML? We propose at least three possibilities. First, loss of ALDH2 expression might be under positive selection to confer a critical metabolic adaptation for the early in vivo expansion of an AML clone, while at later stages of AML progression (reflected by our cell lines), ALDH2 silencing is no longer relevant for cell proliferation. A second possibility is that loss of ALDH2 promotes the genetic evolution of AML by increasing the probability of acquiring aldehyde-induced genetic mutations. A third hypothesis is that ALDH2 is merely a gene susceptible to DNA hypermethylation, with epigenetic silencing occurring as a passenger event in this disease. Irrespective of these different scenarios, our study demonstrates how epigenetic silencing can lead to acquired dependencies in cancer. Interestingly, we find that the ALDH2 locus is occupied by C/EBPα and AML1–ETO proteins in AML cells (data not shown), thus raising the possibility that genetic alteration of these transcription factors is linked to ALDH2 silencing.

Our study reveals a paradox: a germline deficiency of FA genes leads to an elevated risk of AML formation while sporadic AML can acquire a dependency on FA proteins to sustain cell proliferation and survival. The contextual nature of this pathway is reminiscent of other DNA repair regulators, such as ATM, which act to protect normal tissues from cancer-causing somatic mutations, whereas inhibition of this pathway can hypersensitize transformed cells to DNA-damaging agents (51–53). Considering the age-dependent onset of symptoms in FA patients, a possibility exists that acute and reversible inhibition of the FA pathway with small molecules may have a therapeutic index in AML, as has been demonstrated for targeting of other DNA repair proteins (54). Therefore, our study provides justification for evaluating pharmacologic inhibition of UBE2T/FANCL-mediated ubiquitination as a therapeutic approach for eliminating ALDH2-deficient cancers.

METHODS

Cell Lines

All cell lines were authenticated using STR profiling. Leukemia lines MOLM-13, NOMO-1, MV4–11, ML-2, HEL, SET-2, THP-1, U937, K562, pancreatic ductal adenocarcinoma (PDAC) lines AsPc-1, CFPAC-1, SUIT-2, PAN-C, MIAPaCa-2, small-cell lung cancer (SCLC) lines NCI-H211, NCI-H82, DMS114, and murine RN2c (MLL-AF9/NRASG12D AML) were cultured in RPMI supplemented with 10% FBS. SEM cells were cultured in Iscove’s Modified Dulbecco’s Medium (IMDM) with 10% FBS. OCI-AML3 cells were cultured in αMEM with 20% FBS. KASUMI-1 cells were cultured in RPMI supplemented with 20% FBS. NCI-H1058 cells were cultured in DMEM:F12 supplemented with 0.005 mg/mL insulin, 0.01 mg/mL transferrin, 30 nmol/L sodium selenite, 10 mmol/L hydrocortisone, 10 mmol/L β-estradiol, 4.5 mmol/L L-glutamine, and 5% FBS. RH30, RD, RH4, CTR (rhombodysomocarcinoma), and HEK293T were cultured in DMEM supplemented with 10% FBS and 4.5 mmol/L L-glutamine. MLL-AF9 (MA9, engineered human AML) cells were cultured in IMDM supplemented with 20% FBS, 10 ng/mL SCF, 10 ng/mL TPO, 10 ng/mL FLT3L, 10 ng/mL IL3, 10 ng/mL IL6. MLL-AF9/NRASG12D, and MLL-AF9/FLT3ITD (engineered human AML) cells were cultured in IMDM with 20% FBS. Murine NIH-3T3 cells were cultured in DMEM with 10% FCS. Penicillin/streptomycin was added to all media. All cell lines were cultured at 37°C with 5% CO2 and were confirmed Mycoplasma-negative. All experiments were performed within one month of thawing a cryopreserved vial of cells.

Construction of E2 and E3 Domain-Focused sgRNA Library

The list of E2 and E3 genes was retrieved from HUGO Gene Nomenclature Committee (HGNC) resource. The E2 and E3 enzymatic functional domain annotation was retrieved from the NCBI Conserved Domains Database. Six to ten independent sgRNAs were designed against each individual domain based while minimizing off-target effects (Hsu and colleagues, 2013; ref. 55). The domain targeting and spike-in positive/negative control sgRNA oligonucleotides were synthesized using an array platform (Twist Bioscience) and then amplified by PCR. The PCR products were cloned into the BamHI-digested optimized sgRNA lentiviral expression vector LRG2.1T using a Gibson Assembly Kit (New England Biolabs, catalog no. E2611). The LRG2.1T vector was derived from the lentiviral U6-sgRNA-EFS-GFP expression vector (Addgene: #65656). The pooled plasmids library was subjected to deep-sequencing analysis on a MiSeq instrument (Illumina) to verify the identity and representative of sgRNAs in the library. It was confirmed that 100% of the designed sgRNAs were cloned in the LRG2.1T vector and that the abundance of >95% of individual sgRNA constructs was within 5-fold of the mean. The sgRNA sequences used in this study are provided in Supplementary Tables S1 and S2.

Virus Production and Transduction

Lentivirus was produced in HEK293T cells by transfecting lentivirus plasmids together with helper packaging plasmids (VSVG and psPAX2) using polyethylenimine (PEI 25000; Polysciences; catalog number 21662). The lentiviral supernatant was collected 48 h after transfection and filtered through a 0.45-μm filter. Lentiviral supernatant was then used to infect cultured cells.

Cancer Discovery | September 2021

Downloaded from cancerdiscovery.aacrjournals.org on September 15, 2021. © 2021 American Association for Cancer Research.
no. 23966–1) transfection reagent. HEK293T cells were plated in 10-cm culture dishes and were transfected when confluency reached approximately 80% to 90%. For pooled screens, five plates of HEK293T were used to ensure the representation of the library. For one 10-cm dish of HEK293T cells, 10 μg of plasmid DNA, 5 μg of pVSVG, and 7.5 μg psPAX2 and 64 μL of 1 mg/mL PEI were mixed, incubated at room temperature for 20 minutes, and then added to the cells. The media was changed to fresh media 6 to 8 hours posttransfection. The media-containing lentivirus was collected at 24, 48, and 72 hours posttransfection and pooled together. Virus was filtered through 0.45-μm nonpyrogenic filter.

For shRNA knockdown experiments, retrovirus was produced in Plat-E cells, which were transfected with retroviral DNA (MLS-E plasmid), VSVG, and Eco helper plasmids in a ratio of 10:1:1.5. The media-containing lentivirus was collected at 24, 48, and 72 hours after transfection and pooled together. Virus was filtered through 0.45-μm nonpyrogenic filter.

For both lentivirus and retrovirus infection, target cells were mixed with corresponding volume of virus supplemented with 4 μg/mL polybrene, and then centrifuged at 600 × g for 40 minutes at room temperature. If selection was needed for stable cell line establishment, corresponding antibiotics (1 μg/mL puromycin, 1 mg/mL G418) were added 72 hours postinfection.

**Plasmid Construction: sgRNA and shRNA Cloning**

For CRISPR screening, the optimized sgRNA lentiviral expression vector (LRG2.1) and the lentiviral human codon-optimized _Streptococcus pyogenes_ Cas9 vector (LentiV_Cas9_Puro, Addgene: 108100) were used. For the competition-based proliferation assays, sgRNAs were cloned into the LentiCR2.1 vector using BamHI restriction site. LRCherry2.1 was derived from LRG2.1 by replacing GFP with mCherry CDS. For the cDNA rescue experiments, the cDNA of UBE2T, FANCL, ALDH1A1, or ALDH2 were cloned into the LentiV_Neo vector expressing cancer cell lines, which were established by infection with the mirE-based retroviral shRNA expression vector MLS-E.

**Pooled Negative Selection CRISPR Screening and Data Analysis**

CRISPR-based negative selection screening was performed in Cas9-expressing cancer cell lines, which were established by infection with LentiV-Cas9-Puro vector and selected with puromycin. The lentivirus of the pooled sgRNA library targeting the functional domains of ubiquitination genes was produced as described above. Multiplicity of infection (MOI) was set to 0.3–0.4 to ensure a single sgRNA transcription per cell. To maintain the representation of sgRNAs during the screen, the number of sgRNA-infected cells was kept to 800 times the number of sgRNAs in the library. Cells were harvested at three days postinfection as an initial reference time point. Cells were cultured for 12 days and Data Analysis

**Analysis of Genetic Dependencies and Gene Expression in DepMap and Other Data Sets**

Genetic dependency (CRISPR, Avana) data, RNA-seq gene expression (CCLE) data, and DNA methylation data (CCLE, promoter 1kb upstream TSS) from cancer cell lines were extracted from the DepMap Public Project Achilles 21Q1 database (https://depmap.org/portal/). For two-class comparison in Supplementary Fig. S7E, moderated test was performed on DepMap to identify differentially dependent genes for each group. RNA expression data with genomic information in AML patient and other tumor patient samples was extracted from the TCGA database via cbioportal (https://www.cbioportal.org/). Pearson correlation coefficient was calculated with Python 3.8.5.

**Competition-Based Cell Proliferation Assay**

Cas9-expressing cell lines were lentivirally transduced with LRG2.1 sgRNA linked with GFP or mCherry reporter. Percentage of GFP-positive cell population was measured at day 3 or day 4 as initial time point using a Guava EasyCyte HT (Millipore) or a MACSQuant (Miltenyi Biotec) instrument. GFP% (mCherry%) was then measured every two days (for leukemia cell lines) or every three days (for nonleukemia cell lines) over a time course. The change in the GFP% (or mCherry%) percentage at each time point was then normalized to initial time point GFP% (or mCherry%). This relative change was used to assess the impact of individual sgRNAs on cellular proliferation, which reflects cells with a genetic knockout being outcompeted by nontransduced cells in the cell culture.

**Western Blot Analysis**

Cells were collected and washed once with PBS. Cell pellets were resuspended in RIPA buffer (Thermo Scientific; catalog no. 89901) and sonicated to fragment chromatin. Cell lysate was mixed with SDS-loading buffer containing 2-mercaptoethanol and boiled at 95°C for 5 minutes. The cell extracts were separated by SDS-PAGE (NuPAGE 4%–12% Bis-Tris protein Gels, Thermo Fisher Scientific), followed by transfer to a nitrocellulose membrane using wet transfer at 30 V overnight. The membrane was blocked with 5% nonfat milk in TBST and incubate with primary antibody (1:500 dilution except FLAG antibody which is 1:1,000 dilution) in 5% milk in room temperature for 1 hour. After incubation, the membrane was washed for three times with TBST followed by incubation with secondary antibody for 1 hour at room temperature. After washing with TBST three times, the membrane was then incubated with chemiluminescent HRP substrate (Thermo Fisher; catalog no. 34075). Primary antibodies

barcodes (Shi and colleagues, 2015; ref. 20), with Quick Ligation Kit (NEB; catalog no. M2200L). The ligated DNA was PCR amplified with primers containing Illumina paired-end sequencing adapters. The final libraries were quantified using bioanalyzer Agilent DNA 1000 (Agilent 5067–1504) and were pooled together in equal molar ratio for paired-end sequencing using the MiSeq platform (Illumina) with MiSeq Reagent Kit V3 150-cycle (Illumina).

The sequencing data were demultiplexed and trimmed to contain only the sgRNA sequence. The sgRNA sequences were mapped to a reference sgRNA library to discard any mismatched sgRNA sequences. The read counts of each sgRNA were calculated. The following analysis was performed with a custom Python script: sgRNAs with read counts less than 50 in the initial time point were discarded; the total read counts were normalized between samples; artificial one count was assigned to sgRNAs that have zero read count at final time point; the average log₂ fold change in abundance of all sgRNA against a given domain was calculated. AML-specific dependency was determined by subtracting the average of log₂ fold-change in non-AML cell lines from average log₂ fold-change in AML cell lines, and the score was ranked in ascending order. The E2/E3 CRISPR screening library and data are provided in Supplementary Tables S1 and S2.

**Western Blot Analysis**

Cells were collected and washed once with PBS. Cell pellets were resuspended in RIPA buffer (Thermo Scientific; catalog no. 89901) and sonicated to fragment chromatin. Cell lysate was mixed with SDS-loading buffer containing 2-mercaptoethanol and boiled at 95°C for 5 minutes. The cell extracts were separated by SDS-PAGE (NuPAGE 4%–12% Bis-Tris protein Gels, Thermo Fisher Scientific), followed by transfer to a nitrocellulose membrane using wet transfer at 30 V overnight. The membrane was blocked with 5% nonfat milk in TBST and incubate with primary antibody (1:500 dilution except FLAG antibody which is 1:1,000 dilution) in 5% milk in room temperature for 1 hour. After incubation, the membrane was washed for three times with TBST followed by incubation with secondary antibody for 1 hour at room temperature. After washing with TBST three times, the membrane was then incubated with chemiluminescent HRP substrate (Thermo Fisher; catalog no. 34075). Primary antibodies
used in this study included UBET2 (Abcam; catalog no. ab140611), FANCDC2 (Novus Biologicals; catalog no. NB100-182SS), ALDH1A1 (ProteinTech; catalog no. 22109-1-AP), ALDH2 (ProteinTech; catalog no. 15310-1-AP), FLAG (Sigma; catalog no. F1804), TP53 (Santa Cruz Biotechnology; sc-126), CDKN1A (Santa Cruz Biotechnology; catalog no. SC-71811).

**In Vivo Transplantation of MOLM-13 Cells into NSG Mice**

Animal procedures and studies were conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) at Cold Spring Harbor Laboratory (Cold Spring Harbor, NY). For experiments in Fig. 2 and Supplementary Fig. S2, MOLM-13-Cas9 cells were first transduced with a luciferase-expressing cassette in Lentiluciferase-P2A-Neo (Addgene #105621) vector, followed by G418 (1 mg/mL) selection and then viral transduction with LRG2.1T-sgRNA-GFP vectors targeting the FANCDC2 gene and negative control. Five replicates were performed for each sgRNA. On day 3 postinfection with the sgRNA, the infection rate was checked by the percentage of GFP-positive cells, and all samples had more than 90% infection rate. 0.5 million cells were injected intravenously into sublethally irradiated (2.5 Gy) NSG mice (Jax 005577). To detect the disease progression, mice were imaged with IVIS Spectrum system (Caliper Life Sciences) on day 10, 13, 16, and 19 postinfection.

For experiments performed in Supplementary Fig. S3, MOLM-13-Cas9/luciferase cells were transduced with LRG2.1T-lasticidin, and cells were selected with blasticidin (10 μg/mL) from day 3 to day 6 postinfection. On day 6 postinfection, 0.1 million cells were injected intravenously into sublethally irradiated (2.5 Gy) NSG mice. Mice were euthanized between day 20 and day 25 post infection. On day 6 postinfection, 0.1 million cells were injected intravenously into sublethally irradiated (2.5 Gy) NSG mice. Mice were euthanized between day 20 and day 25 post infection. Spleen and bone marrow were collected for total RNA (TRizol) and DNA extraction (QIAGEN; catalog no. 51304) according to the manufacturer’s instructions. qRT-PCR was used to measure the expression level of ALDH2, UBET2, FANCL, and FANCDC2. To detect the CRISPR cutting sites, an approximately 400-bp region surrounding the CRISPR sgRNA-binding site was PCR amplified and cloned into Lentivector for Sanger sequencing. The primer sequences used for this analysis are provided in Supplementary Table S3.

**Cell-Cycle Arrest and Apoptosis Analysis**

Cell-cycle analysis was performed according to the manufacturer’s protocol (BD, FITC BrdU Flow Kit; catalog no. 559619), with cells pulsed with BrdU for 1 hour at 37°C. Cells were contained with 4’,6-diamidino-2-phenylindole (DAPI) for DNA content measurement, and analyzed with a BD LSRFortessa flow cytometer (BD Biosciences) and FlowJo software (TreeStar). Annexin V apoptosis staining was performed according to the manufacturer’s protocol, with DAPI stained for DNA content measurement (BD, FITC Annexin V Apoptosis Detection Kit; catalog no. 556547). The experiments were performed in triplicate.

**Immunofluorescence for Phospho-H2AX Foci**

Cells were harvested four days after lentiviral spin-infection and spun onto a slide using a Shandon Cytospin 2 centrifuge. Cells were washed once in PBS, fixed with 3.7% formaldehyde (Sigma) in PBS for 10 minutes at room temperature, washed twice with PBS, permeabilized with 0.5% Triton X-100 (Sigma) in PBS for 10 minutes at room temperature, washed in PBS twice, incubated in 5% PBS/PBS for 1 hour at room temperature for blocking, incubated with primary antibody (anti-phospho-H2AX, clone JW301; Millipore, catalog no. 05-636) at 1:1,000 dilution at 4°C overnight, washed with 5% PBS/PBS for 5 minutes three times, incubated with secondary antibody (anti-mouse AF594; Invitrogen; catalog no. A11005) at 1:2,000 dilution for 1 hour at room temperature, washed with 5% PBS/PBS for 5 minutes three times, once with PBS, and mounted with DAPI Fluoromount-G (SouthernBiotech). Images were obtained using Zeiss Axio Observer A1 microscope and AxioVision 4.9.1. software. Data were analyzed by CellProfiler 3.1.8.

**Radial Chromosome Formation Assay**

At day 4 and day 6 postinfection, MOLM-13 or MV4–11 cells were arrested with colcemid (0.17 μg per mL of media) for 10 minutes, harvested, incubated for 12 minutes at 37°C in 0.075 mol/L KCl and fixed in freshly prepared methanol-glacial acid acid (3:1 vol/vol). Cells were stored at 4°C and, when needed, dropped onto wet slides and air dried at 40°C for 60 minutes before staining with 6% KaryoMAX Giemsa (Invitrogen #10092-013) in Gurr Buffer (Invitrogen #10582-013) for 3 minutes. After rinsing with fresh Gurr Buffer followed by distilled water, the slides were fully dried at 40°C for 60 minutes and scanned using the Metasystems Metafer application. Breakage analysis was blinded.

**RNA-seq Library Construction**

Total RNA of each cell line was extracted using TRizol reagent (Thermo Scientific; catalog no. 15596018) according to the manufacturer’s instructions. Briefly, 3 million cells were lysed with 1 mL of TRizol and 200 μL chloroform and incubated for 15 minutes at room temperature followed by centrifugation at 10,000 × g for 15 minutes at 4°C. The aqueous phase was added to equal volume of iso-propanol and incubated for 10 minutes at room temperature. RNA was precipitated at 10,000 × g for 10 minutes at 4°C, and the pellet was washed once with 75% ethanol and dissolved in DEPC-treated water. RNA-seq libraries were constructed using TruSeq sample preparation kit V2 (Illumina) according to the manufacturer’s instructions. Briefly, polyA mRNA was selected from 2 μg of purified total RNA and fragmented with fragmentation enzyme. The first strand of cDNA was synthesized using Super Script II reverse transcriptase, and then the second strand was synthesized. Double-stranded cDNA was end-repaired, 3′-adenylated, ligated with indexed adaptor, and then PCR-amplified. The quantity of the RNA-seq library was determined by Nanodrop, and the average quantity of RNA-seq libraries ranged from 40 to 80 ng/μL. The same molar amount of RNA-seq library was pooled together and analyzed by single-end sequencing using NextSeq platform (Illumina) with single-end reads of 75 bases.

**RNA-seq Data Analysis**

Sequencing reads were mapped into reference genome hg38 using STAR v2.5.2 with default parameters (56). Read count tables were created by HTSeq v0.6.1 with customed gff file containing protein-coding genes only. Differentially expressed genes were analyzed using DESeq2 with replicate (57) using default parameters. RPKMs (reads per kilobase per million mapped reads) were calculated using Cuffdiff v2.2.1 with default parameters (58). Genes with RPKMs of more than three in the control were considered as expressed and used in the subsequent analysis. Genes were ranked by their log2 fold change calculated from DESeq2 as input for Pre-ranked gene set enrichment analysis (1,000 permutations) with all available signatures in the Molecular Signature Database v5.2 (MSigDB) and Leukemia-stem-cell signature from Somervaille and colleagues (59). Top down-regulated genes, defined as genes downregulated upon FANCDC2 or UBET2 knockout with P ≤ 0.1 and log2FoldChange ≤ –0.5, were subjected to Gene Ontology (GO) term analysis with Metascape Express Analysis. Raw data can be obtained at the NCBI Gene Expression Omnibus (GEO) database at accession GSE169586.

**Aldefluor Assay**

The Aldefluor assay was performed following the instruction of the Aldefluor Kit (STEMCELL; catalog no. 01700). Briefly, 0.5 × 10⁵ fresh cell samples were collected and washed once in PBS buffer. The cells were then resuspended in Aldefluor Assay Buffer to 1 × 10⁶/mL. Five
microliters of DEAB reagent was added to the cell line as negative control. Five microliters of activated Aldefluor reagent was added to the control and test samples. The solution was mixed and incubated at 37°C for 50 minutes. The cells were collected and resuspended in 500 μL Aldefluor Assay Buffer, and subjected to flow cytometry assay for data acquisition. The experiments were performed in triplicate.

Mitochondrial Fractionation

Mitochondrial fractionation was performed using the Mitochondria Isolation Kit for Cultured Cells following the manufacturer's instructions (Thermo Scientific; catalog no. 89874). Briefly, cell membrane was first disrupted by three freeze and thaw cycles, and mitochondria fraction was collected by centrifugation. The collected mitochondrial pellet was resuspended in RIPA buffer for further Western blot analysis.

ChIP-seq Analysis

For chromatin immunoprecipitation sequencing (ChIP-seq) analysis, raw reads were obtained from public GEO datasets MGML-13 (GSE67392), K562 (GSM1652918), MV-4-11, THP-1 (GSE79899), and mapped to the human genome (hg19) using Bowtie2 (version 2.2.3) software (Langmead and Salzberg, 2012) using sensitive settings. Duplicate reads were removed prior to peak calling using MACS2 (version 2.1.2.10160309; ref. 60) software using 5% FDR cutoff and broad peak option. Sequencing depth normalized ChIP-seq pileup tracks were visualized using the UCSC genome browser (61).

Nanopore Sequencing

cRNA guides specific to the regions of interest (ROI) were designed as per recommended guidelines described in the Nanopore infosheet on Targeted, amplification-free DNA sequencing using CRISPR/Cas (Version: ECl_S1014_v1_revA_11Dec2018). Guides were reconstituted to 100 μmol/L using TE (pH 7.5) and pooled into an equimolar mix. For each distinct sample, four identical reactions were prepared parallelly using 5 μg gDNA each. Ribonucleoprotein complex (RNP) assembly, genomic DNA dephosphorylation, and Cas9 cleavage were performed as described in ref. 62. Affinity-based Cas9-Mediated Enrichment (ACME) using Invitrogen His-Tag Dynabeads was performed to pulldown Cas9 bound nontarget DNA, increasing the proportion of on-target reads in the sample (Iyer and colleagues 2020; ref. 63). The resultant product was cleaned up using 1X Ampure XP beads (Beckman Coulter; catalog no. A63881), eluted in nucleic-acid-free water, and pooled together. The ACME-enriched sample was quantified using Qubit fluorometer (Thermo Scientific) and carried forward to the adapter ligation step as described by Iyer and colleagues (63).

qRT-PCR Analysis Following 5-Azacytidine Treatment

For dose-dependent experiment, 5-azacytidine was added to cell culture with different concentrations, and cells were collected after 36-hour treatment. For time-course treatment, cells were treated with 1 μmol/L 5-azacytidine and collected at different time points. Total RNA was extracted from each sample at the percent methylation of the cytosines covered within the ALDH2 CpG was plotted using ComplexHeatmap (68), with the complex clustering method and Euclidian distances, and using light gray for NA values. For the box plot of the average percent methylation for each sample at the CGi, the mean methylation for each sample was calculated using all CGs that were covered in that region, and the percent methylation for each sample was calculated using all CGs that were present in at least 5 samples per condition (meth.min = 5). The average percent methylation for each sample at the CGi was calculated using the MethylKit “percMethylation” function. Bedtools (67) intersect function was used to determine overlap with CpG from hg19.

Aligned Enhanced Representation Bisulfite Sequencing

myCpG files from AML (n = 119) and normal (n = 22) individuals were downloaded from GEO (accession number GSE98350; ref. 39). After filtering and normalizing by coverage, a methylBase object containing the methylation information and locations of cytosines that were present in at least 5 samples per condition (meth.min = 5) was generated using MethylKit (version 1.9.3; ref. 66) and R statistical software (version 3.5.1). Percent methylation for each Cytosine- Guanine dinucleotide (CG) for each donor was calculated using the MethylKit “percMethylation” function. Bedtools (67) intersect function was used to determine overlap with CpG from hg19. The heat map of the percent methylation of the cytosines covered within the ALDH2 CpG was plotted using ComplexHeatmap (68), with the complex clustering method and Euclidian distances, and using light gray for NA values. For the box plot of the average percent methylation for each sample at the CGi, the mean methylation for each sample was calculated using all CGs that were covered in that region, and the percent methylation for each sample was calculated using all CGs that were present in at least 5 samples per condition (meth.min = 5). The average percent methylation for each sample at the CGi was calculated using the MethylKit “percMethylation” function. Bedtools (67) intersect function was used to determine overlap with CpG from hg19.
Statistical Analysis

Error bars represent the mean ± SEM, and n refers to the number of biological repeats. Statistical significance was evaluated by P value using GraphPad Prism software or Scipy package from Python 3.8.5 as indicated in the figure legends. For Kaplan–Meier survival curves, the log rank (Mantel–Cox) test was used to estimate median overall survival and statistical significance.

Authors’ Disclosures

S.V. Iyer reports non-financial support from Oxford Nanopore Technologies Ltd. outside the submitted work. W. McCombie reports grants from NIH during the conduct of the study; non-financial support from Pacific Bioscience; personal fees and other support from Orion Genomics, and non-financial support and other support from Oxford Nanopore outside the submitted work. M.E. Figueroa reports grants from the NIH-NCI, the Pershing Square Sohn Cancer Research Alliance, and the Leukemia & Lymphoma Society during the conduct of the study. A. Smogorzewska reports grants from Rocket Pharmaceuticals outside the submitted work. C.R. Vakoc reports grants from the NIH-NCI, the Pershing Square Sohn Cancer Research Alliance, and the Leukemia & Lymphoma Society during the conduct of the study; personal fees from Roivant Sciences, C4 Therapeutics, KSQ Therapeutics, Syros Pharmaceuticals, and Flare Therapeutics, and grants from Boehring-Ingelheim outside the submitted work. No disclosures were reported by the other authors.

Authors’ Contributions


Acknowledgments

This work was supported by Cold Spring Harbor Laboratory NCI Cancer Center Support grant SP30CA045508. Additional funding was provided to C.R. Vakoc by the Pershing Square Sohn Cancer Research Alliance, NIH grants R01 CA174793 and P01 CA013106; a Leukemia & Lymphoma Society Scholar Award. A. Smogorzewska is an HHMI Faculty Scholar and is supported by CA204127 from NIH Leukemia & Lymphoma Society Scholar Award. A. Smogorzewska is supported by a Leukemia & Lymphoma Society Scholar Award. M. Jung reports grants from Rhode Island Hospital, personal fees from National Cancer Institute, NIH grants R01 CA174793 and P01 CA013106, a Keating family HHMI Faculty Scholar Award. We thank James C. Mulloy and K99 HL150628 (M. Jung). M.E. Figueroa is supported by a Leukemia & Lymphoma Society Scholarship. A. Smogorzewska reports grants from the National Science Foundation, personal fees from Cold Spring Harbor Laboratory NCI Cancer Research, personal fees from Janssen Research and Development, L.L.C., personal fees from Roivant Sciences, C4 Therapeutics, KSQ Therapeutics, Syros Pharmaceuticals, and Flare Therapeutics, and grants from Boehringe-Ingelheim outside the submitted work. No disclosures were reported by the other authors.

REFERENCES


Received October 23, 2020; revised March 23, 2021; accepted April 21, 2021; published first April 23, 2021.

Downloaded from cancerdiscovery.aacrjournals.org on September 15, 2021. © 2021 American Association for Cancer Research.
Fanconi Anemia Pathway Dependency in AML


47. Schuurhuis GJ, Meel MH, Wouters F, Min LA, Terwijn M, de Jonge NA, et al. Normal hematopoietic stem cells within the AML bone marrow have a distinct and higher ALDH activity level than co-existing leukemia stem cells. PLoS One 2013;8:e78897.


Transcriptional Silencing of ALDH2 Confers a Dependency on Fanconi Anemia Proteins in Acute Myeloid Leukemia

Zhaolin Yang, Xiaoli S. Wu, Yiliang Wei, et al.


Access the most recent version of this article at: doi:10.1158/2159-8290.CD-20-1542
Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2021/04/24/2159-8290.CD-20-1542.DC1

This article cites 68 articles, 10 of which you can access for free at: http://cancerdiscovery.aacrjournals.org/content/11/9/2300.full#ref-list-1

Sign up to receive free email-alerts related to this article or journal.
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/11/9/2300. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.