



Published in final edited form as:

Cancer Discov. 2021 September ; 11(9): 2300–2315. doi:10.1158/2159-8290.CD-20-1542.

Transcriptional silencing of *ALDH2* confers a dependency on Fanconi anemia proteins in acute myeloid leukemia

Zhaolin Yang¹, Xiaoli S. Wu^{1,2}, Yiliang Wei¹, Sofya A. Polyanskaya¹, Shruti V. Iyer^{1,2}, 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, Christopher R. Vakoc^{1,6,*}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

²Genetics Program, Stony Brook University, Stony Brook, New York 11794, USA

³Laboratory of Genome Maintenance, The Rockefeller University, New York 10065, USA

⁴Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

⁵Department of Biochemistry and Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

⁶Lead contact

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 that 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.

*Correspondence: vakoc@cshl.edu; Christopher R. Vakoc, 1 Bungtown Rd, Cold Spring Harbor, NY 11724. 516-367-5030.

AUTHORSHIP CONTRIBUTIONS

Z.Y., X.S.W., Y.W., S.P., S.V.I., M.J., F.P.L., E.R.A., O.K., J.P.M., M.K., O.E.D., K.C., S.G., performed experiments and/or analyzed the data; E.H., W.R.M., M.E.F., A.S., and C.R.V. supervised the experiments and analysis; Z.Y. and C.R.V. wrote the manuscript.

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, non-homologous 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). Due to its therapeutic significance, the identification of synthetic lethal genetic interactions remains an important objective in the study of human cancer (5).

In recent years there has been a resurgent interest in the ubiquitination machinery as targets for cancer therapy. Ubiquitin is a 8.6 kDa protein that is covalently attached to lysine side chains as a form of post-translational 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) (7). With ~40 E2 and over 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 neo-substrates (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 crosslinks (ICLs) (14). A key step in the activation of the FA repair pathway is the mono-ubiquitination of FANCD2 and FANCI, which is performed by UBE2T/FANCT (an E2 conjugating enzyme) and FANCL (an E3 ligase) (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 crosslinks 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 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 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 (Figure 1A, Supplementary Table S1). Using this sgRNA library, we performed negative selection “dropout” screening in twelve Cas9-expressing human cancer cell lines, which included six AML and six solid tumor cell lines (Figure 1B, Supplementary Table S2). Many of the dependencies identified in these screens were pan-essential across the twelve lines, such as ANAPC11, CDC16, and RBX1 (Supplementary Figure 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 (Figure 1C–D). The known function of UBE2T and FANCL as E2 and E3 proteins, respectively, within the Fanconi anemia (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](https://depmap.org)), in which genome-wide CRISPR essentiality screening was performed in 729 cancer cell lines (Figure 1E, Supplementary Figure S1B) (21,22). In addition, these data revealed that several other FA pathway genes, including FANCI, FANCB and FANCG, 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 (Figure 2A). These experiments showed that targeting any of the three FA genes suppressed the fitness of 9 human leukemia lines, including 3 generated by retroviral transduction of oncogenes into human hematopoietic stem and progenitor cells (23). In contrast, the fitness of 5 other human leukemia lines and all of the solid tumor cell lines was less sensitive to targeting of FA genes (Figure 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 (Figure 2B, Supplementary Figure S2A–B).

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 (Figure 2C, Supplementary Figure S2C–F). 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 (Figure 2D, Supplementary Figure S2G–H). To ensure that this dependency was not influenced by DNA damage caused by CRISPR-Cas9, we also validated FANCD2 dependency in AML using RNAi-based knockdown (Figure 2E, Supplementary Figure S2I). 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 (Figure 2F, Supplementary Figure S2J, and Supplementary Figure S3A–C). 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 Figure S3D–G).

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 (Figure 3A–B). An Annexin V staining analysis of UBE2T- or FANCD2-deficient MOLM-13 cells also revealed evidence of apoptosis (Figure 3C–D). The levels of γ H2AX and of metaphase chromosome breaks, both markers of DNA damage, were also increased following FA gene inactivation (Supplementary Figure S4A–C). 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. While 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 (Figure 2A, Supplementary Figure S5A). Using RNA-seq analysis, we found that inactivation of UBE2T or FANCD2 led to significant upregulation of p53 target genes and suppression of leukemia stem cell signatures in TP53 wild-type MOLM-13 (Figure 3E, Supplementary Figure S5B). In addition, we observed induction of pro-apoptotic genes and suppression of DNA replication genes following FA protein inactivation, in accord with the phenotypes described above (Figure 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 disproportionately suppressed following FA protein inactivation (Figure 3E, Supplementary Figure S5B–C). 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 (Figure 3F–G, Supplementary Figure S6A–G). 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 aldehyde dehydrogenases (ALDH), which are comprised of 19 enzymes that oxidize diverse aldehydes into non-toxic acetates in a NAD- or NADP-dependent manner (19). We first made use of the Aldefluor assay, in which cells are treated with BODIPY-aminoacetaldehyde (BAAA), a cell-permeable fluorescent aldehyde that becomes trapped in cells following ALDH-mediated conversion to BODIPY-aminoacetate (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 (Figure 4A). One exception to this correlation was U937 cells, which are FA-dispensable but lacks 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 to FA-dispensable lines (Figure 4B, Supplementary Figure S7A). Using western blotting, we confirmed that ALDH1A1 and ALDH2 proteins are present at higher levels in the FA-dispensable versus the FA-dependent AML lines (Figure 4B). Of note, while ALDH1A1 and ALDH2 are both capable of oxidizing BAAA, each enzyme is known to have distinct cellular functions (29–35). ALDH1A1 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) (32).

Re-expression of ALDH2, but not ALDH1A1, renders 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 (Figure 4C). Aldefluor assays revealed that the lentivirally expressed proteins were enzymatically active, with both ALDH1A1 and ALDH2 causing oxidation of BAAA (Figure 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 (Figure 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 (Figure 4E, Supplementary Figure S7B–C). 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-AF9 and Nras^{G12D} cDNAs (37). Notably, this cell line retains an intact p53 pathway and expresses ALDH2, but not ALDH1A1 (Supplementary Figure S7D). We targeted ALDH2 using CRISPR in this cell line and confirmed loss of protein expression and loss of Aldefluor activity (Figure 4F–G). Notably, in competition-based cell fitness assays we observed that ALDH2-deficient RN2 cells became hypersensitive to the inactivation of FANCD2 relative to the control cells (Figure 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 non-hematopoietic cancer lineages (Supplementary Figure S7E–F). 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 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 dependency of MOLM-13 cells (Supplementary Figure S8A–D). As an additional control, we considered whether ALDH1A1 might be capable of rescuing the FA dependency if we forced its localization into the mitochondria using two different localization signals (Supplementary Figure S9A). Despite effective targeting to the mitochondria confirmed by cell fractionation and robust Aldefluor activity of these mutant proteins (Supplementary Figure S9B–C), the mitochondrial forms of ALDH1A1 were unable to rescue UBE2T dependence (Supplementary Figure 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 Figure S9E–H). 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 Figure S10A). In addition, an RNA-seq analysis suggested that the overall transcriptome of MOLM-13 cells was largely unaffected by re-expression of ALDH2 (Supplementary Figure 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 Figure S10C–D). 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 inter-strand DNA crosslinks.

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 Project (28,40), we detected an inverse correlation between levels of DNA methylation and *ALDH2* expression (Figure 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 (Figure 5B). Importantly, this same location was hypomethylated in normal human hematopoietic stem and progenitor cells (Figure 5B) (42). In addition, silencing of *ALDH2* correlated with diminished histone acetylation at this genomic region (Figure 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 (Figure 5D–E, Supplementary Figure S11A–B). Consistent with this finding, 5-azacytidine treated MOLM-13 cells were also significantly less sensitive to UBE2T inactivation (Supplementary Figure 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 Cancer Genome Atlas (TCGA) pan-cancer analysis, we found that AML had greater variability in *ALDH2* expression when compared to most other human cancer types (Figure 5F). Within this group of TCGA patient samples, we designated ALDH2-low and ALDH2-high AML, distinguished by a ~100-fold difference in ALDH2 expression (Figure 5G). Notably, we observe that a housekeeping gene *ACTB* only showed a 4-fold variance in expression across TCGA AML samples (Supplementary Figure 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 (Figure 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 Figure S11E–G). Using the TCGA datasets, we found that ALDH2-low AML samples are enriched for mutations in *NPM1*, *CEBPA*, and the *RUNX1-RUNX1T1* (AML1-ETO) translocation, whereas ALDH2-high AML is enriched for *TP53* and *KMT2A* alterations (Figure 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 crosslinks 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 reach 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 *Fancd2* 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/Fancd2* compound deficient mice become cleared through a p53-dependent mechanism (43). The redundant function of ALDH2 and FA proteins is also supported by evidence in humans, in which a combined hereditary deficiency of *ALDH2* and FA genes correlates with an accelerated rate of bone marrow failure when compared to 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 Aldefluor assays to human clinical samples and observed diminished ALDH activity in AML when compared to 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 arsenic 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 et al (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. While 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 byproducts 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/EBP α* and AML1-ETO proteins in AML cells (data now shown), thus raising the possibility that genetic alterations 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 while 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 pharmacological inhibition of UBE2T/FANCL-mediated ubiquitination as 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, PANC-1, MIAPaCa-2, small cell lung cancer (SCLC) lines NCI-H211, NCI-H82, DMS114, murine RN2c (MLL-AF9/Nras^{G12D} AML) were cultured in RPMI supplemented with 10% FBS. SEM cells were cultured in IMDM with 10% FBS. OCI-AML3 cells were cultured in alpha-MEM with 20% FBS. KASUMI-1 cells were cultured in RPMI supplemented with 20% FBS. NCI-H1048 cells were cultured in DMEM:F12 supplemented with 0.005 mg/mL insulin, 0.01 mg/mL transferrin, 30 nM sodium selenite, 10 nM hydrocortisone, 10 nM β -estradiol, 4.5 mM L-glutamine and 5% FBS. RH30, RD, RH4, CTR (rhabdomyosarcoma), and HEK293T were cultured in DMEM supplemented with 10% FBS and 4.5 mM 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 IL-3, 10 ng/mL IL-6. MLL-AF9/Nas^{G12D} and MLL-AF9/FLT3^{ITD} (engineered human AML) cells were cultured in IMDM with 20% FBS. Murine NIH-3T3 cells were cultured in DMEM with 10% bovine calf serum. 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 et al., 2013) (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 BsmBI-digested optimized sgRNA lentiviral expression vector LRG2.1T using a Gibson assembly kit (New England Biolabs; Cat. No. E2611). The LRG2.1T vector was derived from a 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 lenti plasmids together with helper packaging plasmids (VSVG and psPAX2) using polyethylenimine (PEI 25000; Polysciences; Cat. No. 23966-1) transfection reagent. HEK293T cells were plated in 10 cm culture dishes and were transfected when confluency reached ~80–90%. 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 min, and then added to the cells. The media was changed to fresh media 6–8 h post-transfection. The media containing lentivirus was collected at 24h, 48h and 72h post transfection and pooled together. Virus was filtered through 0.45 µm non-pyrogenic filter.

For shRNA knock-down 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 was changed to fresh media 6–8 h post-transfection. Retrovirus-containing supernatant was collected at 24, 48 and 72 h after transfection and pooled together. Virus was filtered through 0.45 µm non-pyrogenic 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 $600 \times g$ for 40 min 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 h post infection.

Plasmid construction: sgRNA and shRNA cloning

For CRISPR screening, the optimized sgRNA lentiviral expression vector (LRG2.1T) 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 LRG2.1T vector using BsmBI restriction site. LRCherry2.1 was derived from LRG2.1T 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 using the In-Fusion cloning system (Takara Bio; Cat. No. 121416). The CRISPR-resistant synonymous mutant of UBE2T and FANCL, and catalytic mutant of UBE2T and ALDH2 were cloned using PCR mutagenesis. shRNAs targeting FANCD2 and CDK1 were cloned into 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 transduction 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 post-infection as an initial reference time point. Cells were cultured for 12 population doublings and harvested for the final time point. Genomic DNA was extracted using QIAamp DNA mini kit (QIAGEN; Cat. No. 51304) according to the manufacture's instruction.

The extracted genomic DNA was used for sequencing library preparation. Briefly, the sgRNA cassette was PCR amplified from genomic DNA (~200 bp) using high-fidelity polymerase (Phusion master mix, Thermo Fisher; Cat. No. F531S). The PCR product was end-repaired with T4 DNA polymerase (NEB; Cat. No. B02025), DNA Polymerase I, Large (Klenow) fragment (NEB; Cat. No. M0210L) and T4 polynucleotide kinase (NEB; Cat. No. M0201L). A 3' A overhang was added to the ends of the blunted DNA fragment with Klenow Fragment (3'-5' exo; NEB; Cat. No. M0212L). The DNA fragments were then ligated with diversity-increased custom barcodes (Shi et al., 2015) (20), with Quick ligation kit (NEB; Cat. No. M2200L). The ligated DNA was PCR amplified with primers containing Illumina paired-end sequencing adaptors. 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 was de-multiplexed 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 \log_2 fold change in abundance of all sgRNA against a given domain was calculated. AML-specific dependency was determined by subtracting the average of \log_2 fold-change in non-AML cell lines from average \log_2 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 Table S1 and S2.

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 (<http://depmap.org/portal/>). For two class comparison in Fig. S7E, moderated t-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 Cancer Genome Atlas 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 LRG2.1T 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 (Miltenyl Biotec) instrument. GFP% (mCherry%) was then measured every two days (for leukemia cell lines) or every three days (for non-leukemia cell lines) over a time course. The relative 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 non-transduced 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; Cat. 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 min. The cell extracts were separated by SDS-PAGE (NuPAGE 4–12% Bis-Tris protein Gels, ThermoFisher), followed by transfer to nitrocellulose membrane using wet transfer at 30 V overnight. Membrane was blocked with 5% nonfat milk in TBST and incubate with primary antibody (1:500 dilution except FLAG antibody which is 1:1000 dilution) in 5% milk in room temperature for one hour. After incubation, membrane was washed for three times with TBST followed by incubation with secondary antibody for one hour at room temperature. After three times wash with TBST, membrane was then incubated with chemiluminescent HRP substrate (Thermo fisher; Cat. No. 34075). Primary antibodies used in this study included UBE2T (abcam; Cat. No. ab140611), FANCD2 (Novus Biotech; Cat. No. NB100-182SS), ALDH1A1 (Proteintech; Cat. No. 22109-1-AP), ALDH2 (Proteintech; Cat. No. 15310-1-AP), FLAG (Sigma; Cat. No. F1804), TP53 (Santa Cruz; sc-126), CDKN1A (Santa Cruz; Cat. 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. For experiments in

Fig. 2 and S2, MOLM-13-Cas9 cells were first transduced with a luciferase expressing cassette in Lenti-luciferase-P2A-Neo (Addgene #105621) vector, followed by G-418 (1 mg/mL) selection and then viral transduction with LRG2.1T-sgRNA-GFP vectors targeting the FANCD2 gene and negative control. Five replicates were performed for each sgRNA. On day 3 post infection with the sgRNA, the infection rate was checked by the percentage of GFP positive cells, and all samples had over 90% infection rate. 0.5 million cells were injected intravenous into sublethally irradiated (2.5Gy) NSG mice (Jax 005557). To detect the disease progression, mice were imaged with IVIS Spectrum system (Caliper Life Sciences) on day 10, 13, 16 and 19 post injection.

For experiments performed in Figure S3, MOLM-13/Cas9/luciferase cells were transduced with LRG2.1T-blasticidin, and cells were selected with blasticidin (10 µg/mL) from day 3 to day 6 post-infection. On day 6 post-infection, 0.1 million cells were injected intravenous into sublethally irradiated (2.5Gy) NSG mice. Mice were euthanized between day 20-day 25 post. Spleen and bone marrow were collected for total RNA (Trizol) and DNA extraction (QIAGEN; Cat. No. 51304) according to the manufacturer's instructions. RT-qPCR was used to measure the expression level of ALDH2 and UBE2T/FANCL/FANCD2 genes. To detect the CRISPR cutting sites, a ~400 bp region surrounding the CRISPR sgRNA binding site was PCR amplified and cloned into LentiV-neo-vector 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; Cat. No. 559619), with cells pulsed with BrdU for 1 hour at 37 °C. Cells were co-stained 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 manufacture's protocol, with DAPI stained for DNA content measurement (BD, FITC Annexin V Apoptosis Detection Kit; Cat. 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 min at room temperature, washed twice with PBS, permeabilized with 0.5% Triton X-100 (Sigma) in PBS for 10 min at room temperature, washed in PBS twice, incubated in 5% FBS/PBS for one hour at room temperature for blocking, incubated with primary antibody (anti-phospho-H2AX, clone JBW301; Millipore, Cat. No. 05-636) at 1:1000 dilution at 4 °C overnight, washed with 5% FBS/PBS for 5 min three times, incubated with secondary antibody (anti-mouse AF594; Invitrogen, Cat. No. A11005) at 1:2000 dilution for one hour at room temperature, washed with 5% FBS/PBS for 5 min 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 post infection, MOLM-13 or MV4-11 cells were arrested with colcemid (0.17 μg per ml of media) for 10 min, harvested, incubated for 12 min at 37 °C in 0.075 M KCl and fixed in freshly prepared methanol:glacial acetic 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 min before staining with 6% KaryoMAX Giemsa (Invitrogen #10092-013) in Gurr Buffer (Invitrogen #10582-013) for 3 min. After rinsing with fresh Gurr Buffer followed by distilled water, the slides were fully dried at 40 °C for 60 min 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; Cat.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 10 min at room temperature followed by centrifuge at $10,000 \times g$ for 15 min at 4 °C. The aqueous phase was added to equal volume of isopropanol and incubated for 10 min at room temperature. RNA was precipitated at $10,000 \times g$ for 10 min at 4 °C, 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 instruction. Briefly, polyA mRNA was selected from 2 μg of purified total RNA, and fragmented with fragmentation enzyme. First strand of cDNA was synthesized using Super Script II reverse transcriptase, and then 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 $\text{ng}/\mu\text{L}$. 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 counts tables were created by HTSeq v0.6.1 with customized gtf 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 by 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 \log_2 fold change calculated from DESeq2 as input for Pre-ranked GSEA analysis (1000 permutations) with all available signatures in the Molecular Signature Database v5.2 (MSigDB) and Leukemia-stem-cell signature from Somerville et al. (59). Top downregulated genes, defined as genes down regulated upon FANCD2 or UBE2T knockout with $p\text{-value} \leq 0.1$ and $\log_2\text{FoldChange} \leq -0.5$, were subjected to Gene Ontology (GO) term analysis with Metascape Express Analysis. Raw data can be obtained at the NCBI GEO database at accession GSE169586.

Aldefluor assay

The Aldefluor assay was performed following the instruction of the Aldefluor Kit (STEMCELL; Cat. No. 01700). Briefly, 0.5×10^6 fresh cell samples were collected and washed once in PBS buffer. The cells were then resuspended in Aldefluor Assay Buffer to 1×10^6 /mL. 5 μ L of DEAB reagent was added to the cell lines as negative control. 5 μ L of activated Aldefluor reagent was added to the control and test samples. The solution was mixed and incubated at 37 °C for 50 min. The cells were collected and resuspended in 500 μ L Aldefluor Assay Buffer, and subjected to flow cytometry assay for data acquisition. The experiments were performed in triplicate.

Mitochondria fractionation

Mitochondrial fractionation was performed using the Mitochondria Isolation Kit for Cultured Cells following the manufacture's instruction (Thermo Scientific; Cat 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 ChIP-seq analysis, raw reads were obtained from public GEO datasets MOLM-13(GSE63782), K562 (GSM1652918), MV4-11, 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 2.1.1.20160309) (60) software using 5% FDR cut off and broad peak option. Sequencing depth normalized ChIP-seq pileup tracks were visualization using the UCSC genome browser (61).

Nanopore sequencing

crRNA 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: ECI_S1014_v1_revA_11Dec2018). Guides were reconstituted to 100 μ M using TE (pH 7.5) and pooled into an equimolar mix. For each distinct sample, four identical reactions were prepared parallelly using 5 μ g gDNA each. Ribonucleoprotein complex (RNPs) assembly, genomic DNA dephosphorylation, and Cas9 cleavage were performed as described in (62). Affinity-based Cas9-Mediated Enrichment (ACME) using Invitrogen™ His-Tag Dynabeads™ was performed to pulldown Cas9 bound non-target DNA, increasing the proportion of on-target reads in the sample (Iyer et al. 2020) (63). The resultant product was cleaned up using 1X Ampure XP beads (Beckman Coulter; Cat. No. A63881), eluted in nuclease-free water, and pooled together. The ACME enriched sample was quantified using Qubit fluorometer (Thermo Fisher Scientific) and carried forward to the adapter ligation step as described by Iyer et al (63). Sequencing adaptors from the Oxford Nanopore Ligation Sequencing Kit (ONT; SQK-LSK109) were ligated to the target fragments using T4 DNA ligase (NEBNext Quick Ligation Module E6056). The sample was cleaned up using 0.3X Ampure XP beads (Beckman Coulter; Cat. No. A63881), washed with long-fragment buffer (LFB; ONT, SQK-LSK109), and eluted in 15 μ L of

elution buffer (EB; ONT, LSK109) for 30 min at room temperature. Resultant library was prepared for loading as described in the Cas-mediated PCR-free enrichment protocol from ONT (Version: ENR_9084_v109_revH_04Dec2018) by adding 25 μ L sequencing buffer (SQB; ONT, LSK109) and 13 μ L loading beads (LB; ONT, LSK109) to 12 μ L of the eluate. Each sample was run on a FLO-MIN106 R9.4.1 flow cell using the GridION sequencer.

Real time base calling was performed with Guppy v3.2, and files were synced to our Isilon 400NL storage server for further processing on the shared CSHL HPCC. Nanopolish v0.13.2 (41) was used to call methylation per the recommended workflow. Briefly, indexing was performed to match the ONT fastq read IDs with the raw signal level fast5 data. The ONT reads were then aligned to the human reference genome (UCSC hg38) using minimap2 v2.17 (64) and the resulting alignments were then sorted with samtools v0.1.19 (65). Nanopolish call-methylation was then used to detect methylated bases within the targeted regions – specifically 5-methylcytosine in a CpG context. The initial output file contained the position of the CpG dinucleotide in the reference genome and the methylation call in each read. A positive value for `log_lik_ratio` was used to indicate support for methylation, using a cutoff value of 2.0. The helper script `calculate_methylation.py` was then used to calculate the frequency of methylation calls by genomic position.

RT-qPCR 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 h treatment. For time-course treatment, cells were treated with 1 μ M 5-azacytidine and collected at different time points. Total RNA was extracted using TRIzol reagent as described above. 1–2 μ g of total RNA was treated with DNaseI and reverse transcribed to cDNA using qScript cDNA SuperMix (Quantabio; Cat. No. 84033), followed by RT-qPCR analysis with SYBR green PCR master mix (Thermo Fisher; Cat. No. 4309155) on a QuantStudio™ 7 Flex Real-Time PCR System. GAPDH was used as reference gene. The primers used in this study are listed in Supplementary Table S3.

Aligned enhanced reduced representation bisulfite sequencing (ERRBS)

myCpG files from AML (n=119) and normal (n=22) individuals were downloaded from GEO (accession number GSE98350) (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) (66) and R statistical software (version 3.5.1). Percent methylation for each CG for each donor, was calculated using the MethylKit ‘`percMethylation`’ function. Bedtools (67) intersect function was used to determine overlap with CpGi from hg19. The heatmap of the percent methylation of the cytosines covered within the ALDH2 CpGi was plotted using ComplexHeatmap (68), with the complex clustering method and Euclidian distances, and using light grey for NA values. For the boxplot of the average percent methylation for each sample at the CpGi, the mean methylation for each sample was calculated using all CG that were covered in that region, using R. Data was plotted using ggplot2 and significance calculated using the ggplot2 function ‘`stat_compare_means`’ with the Student’s t-test method. To correlate methylation with gene expression, processed

expression data that was generated for the above samples using the Affymetrix Human Genome 133 Plus2.0 GeneChips was downloaded (39,69). The average percent methylation of the regions covered within the CpGi versus RNA expression for ALDH2 was plotted using ggplot2 and fitted with a line calculated with a linear model. Pearson's correlation method was used to determine significance.

TCGA mutation analysis

Gene expression and mutation dataset for AML patients were extracted from TCGA pan cancer dataset through cBioportal. AML patients were ranked based on their ALDH2 expression. 40 patients that have the highest and lowest ALDH2 expression were categorized as ALDH2 high and ALDH2 low patient groups. Unpaired student t-test was performed for statistical significance analysis (p value).

Fluorometric measurements of endogenous aldehydes

Formaldehyde assay (ab272524) malondialdehyde assay (ab118970) kits were performed according to manufacturer's instructions.

Statistics

Error bars represent the mean plus or minus standard error of the mean, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank James C. Mulloy for sharing genetically engineered human AML cell lines.

FINANCIAL SUPPORT

This work was supported by Cold Spring Harbor Laboratory NCI Cancer Center Support grant 5P30CA045508. Additional funding was provided to C.R.V. by the Pershing Square Sohn Cancer Research Alliance, National Institutes of Health grants R01 CA174793 and P01 CA013106, a Leukemia & Lymphoma Society Scholar Award. A. S. is an HHMI Faculty Scholar and is supported by CA204127 from NIH and K99 HL150628 (M.J). M.E.F. is supported by a Leukemia & Lymphoma Society Scholar Award.

CONFLICT OF INTEREST DISCLOSURES

C.R.V. has received consulting fees from Flare Therapeutics, Roivant Sciences, and C4 Therapeutics, has served on the scientific advisory board of KSQ Therapeutics and Syros Pharmaceuticals, and has received research funding from Boehringer-Ingelheim. W.R.M. is a founder and shareholder of Orion Genomics and has received research support from Pacific Biosciences and support for attending meetings from Oxford Nanopore. A.S. has received research funding and partial salary support from Rocket Pharmaceuticals. None of these companies were involved in the planning or performance of this research.

REFERENCES

1. Kafri R, Springer M, Pilpel Y. Genetic Redundancy: New Tricks for Old Genes. *Cell*. 2009;136:389–92. [PubMed: 19203571]
2. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 2010;11:97–108. [PubMed: 20051986]
3. Cereda M, Mourikis TP, Ciccarelli FD. Genetic Redundancy, Functional Compensation, and Cancer Vulnerability. *Trends in Cancer*. 2016;2:160–2. [PubMed: 28741568]
4. O'Connor MJ. Targeting the DNA Damage Response in Cancer. *Molecular Cell*. 2015;60:547–60. [PubMed: 26590714]
5. Kaelin WG. The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nat Rev Cancer*. 2005;5:689–98. [PubMed: 16110319]
6. Komander D, Rape M. The Ubiquitin Code. *Annu Rev Biochem*. 2012;81:203–29. [PubMed: 22524316]
7. Grabbe C, Husnjak K, Dikic I. The spatial and temporal organization of ubiquitin networks. *Nat Rev Mol Cell Biol*. 2011;12:295–307. [PubMed: 21448225]
8. Senft D, Qi J, Ronai ZA. Ubiquitin ligases in oncogenic transformation and cancer therapy. *Nat Rev Cancer*. 2018;18:69–88. [PubMed: 29242641]
9. Hoeller D, Dikic I. Targeting the ubiquitin system in cancer therapy. *Nature*. 2009;458:438–44. [PubMed: 19325623]
10. Skaar JR, Pagan JK, Pagano M. SCF ubiquitin ligase-targeted therapies. *Nat Rev Drug Discov*. 2014;13:889–903. [PubMed: 25394868]
11. Hatakeyama STRIM proteins and cancer. *Nat Rev Cancer*. 2011;11:792–804. [PubMed: 21979307]
12. Kronke J, Udeshi ND, Narla A, Grauman P, Hurst SN, McConkey M, et al. Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. *Science*. 2014;343:301–5. [PubMed: 24292625]
13. Lu G, Middleton RE, Sun H, Naniong M, Ott CJ, Mitsiades CS, et al. The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. *Science*. 2014;343:305–9. [PubMed: 24292623]
14. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol*. 2016;17:337–49. [PubMed: 27145721]
15. Kottemann MC, Smogorzewska A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature*. 2013;493:356–63. [PubMed: 23325218]
16. Alcón P, Shakeel S, Chen ZA, Rappsilber J, Patel KJ, Passmore LA. FANCD2–FANCI is a clamp stabilized on DNA by monoubiquitination of FANCD2 during DNA repair. *Nat Struct Mol Biol*. 2020;27:240–8. [PubMed: 32066963]
17. Wang R, Wang S, Dhar A, Peralta C, Pavletich NP. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature*. 2020;580:278–82. [PubMed: 32269332]
18. Wang AT, Smogorzewska A. SnapShot: Fanconi Anemia and Associated Proteins. *Cell*. 2015;160:354–354.e1. [PubMed: 25594185]
19. Langevin F, Crossan GP, Rosado IV, Arends MJ, Patel KJ. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature*. 2011;475:53–8. [PubMed: 21734703]
20. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. 2015;33:661–7. [PubMed: 25961408]
21. Dempster JM, Rossen J, Kazachkova M, Pan J, Kugener G, Root DE, et al. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines [Internet]. *Cancer Biology*; 20197. Available from: <http://biorxiv.org/lookup/doi/10.1101/720243>
22. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49:1779–84. [PubMed: 29083409]

23. Wei J, Wunderlich M, Fox C, Alvarez S, Cigudosa JC, Wilhelm JS, et al. Microenvironment Determines Lineage Fate in a Human Model of MLL-AF9 Leukemia. *Cancer Cell*. 2008;13:483–95. [PubMed: 18538732]
24. Alpi AF, Pace PE, Babu MM, Patel KJ. Mechanistic Insight into Site-Restricted Monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Molecular Cell*. 2008;32:767–77. [PubMed: 19111657]
25. Machida YJ, Machida Y, Chen Y, Gurtan AM, Kupfer GM, D'Andrea AD, et al. UBE2T Is the E2 in the Fanconi Anemia Pathway and Undergoes Negative Autoregulation. *Molecular Cell*. 2006;23:589–96. [PubMed: 16916645]
26. Garaycochea JI, Crossan GP, Langevin F, Daly M, Arends MJ, Patel KJ. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature*. 2012;489:571–5. [PubMed: 22922648]
27. Hira A, Yabe H, Yoshida K, Okuno Y, Shiraishi Y, Chiba K, et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood*. 2013;122:3206–9. [PubMed: 24037726]
28. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019;569:503–8. [PubMed: 31068700]
29. Fan X, Molotkov A, Manabe S-I, Donmoyer CM, Deltour L, Foglio MH, et al. Targeted Disruption of Aldh1a1 (Raldh1) Provides Evidence for a Complex Mechanism of Retinoic Acid Synthesis in the Developing Retina. *MCB*. 2003;23:4637–48. [PubMed: 12808103]
30. Tomita H, Tanaka K, Tanaka T, Hara A. Aldehyde dehydrogenase 1A1 in stem cells and cancer. *Oncotarget*. 2016;7:11018–32. [PubMed: 26783961]
31. Chen Y, Koppaka V, Thompson DC, Vasilio V. Focus on Molecules: ALDH1A1: From lens and corneal crystallin to stem cell marker. *Experimental Eye Research*. 2012;102:105–6. [PubMed: 21536030]
32. Chen C-H, Ferreira JCB, Gross ER, Mochly-Rosen D. Targeting Aldehyde Dehydrogenase 2: New Therapeutic Opportunities. *Physiological Reviews*. 2014;94:1–34. [PubMed: 24382882]
33. Fernandez E, Koek W, Ran Q, Gerhardt GA, France CP, Strong R. Monoamine Metabolism and Behavioral Responses to Ethanol in Mitochondrial Aldehyde Dehydrogenase Knockout Mice. *Alcoholism Clin Exp Res*. 2006;30:1650–8.
34. Kitagawa K, Kawamoto T, Kunugita N, Tsukiyama T, Okamoto K, Yoshida A, et al. Aldehyde dehydrogenase (ALDH) 2 associates with oxidation of methoxyacetaldehyde; in vitro analysis with liver subcellular fraction derived from human and *Aldh2* gene targeting mouse. *FEBS Letters*. 2000;476:306–11. [PubMed: 10913633]
35. Ohsawa I, Nishimaki K, Murakami Y, Suzuki Y, Ishikawa M, Ohta S. Age-Dependent Neurodegeneration Accompanying Memory Loss in Transgenic Mice Defective in Mitochondrial Aldehyde Dehydrogenase 2 Activity. *Journal of Neuroscience*. 2008;28:6239–49. [PubMed: 18550766]
36. Nakahata K, Uehara S, Nishikawa S, Kawatsu M, Zenitani M, Oue T, et al. Aldehyde Dehydrogenase 1 (ALDH1) Is a Potential Marker for Cancer Stem Cells in Embryonal Rhabdomyosarcoma. Loeb DM, editor. *PLoS ONE*. 2015;10:e0125454. [PubMed: 25915760]
37. Zuber J, McJunkin K, Fellmann C, Dow LE, Taylor MJ, Hannon GJ, et al. Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat Biotechnol*. 2011;29:79–83. [PubMed: 21131983]
38. Nene A, Chen C-H, Disatnik M-H, Cruz L, Mochly-Rosen D. Aldehyde dehydrogenase 2 activation and coevolution of its ePKC-mediated phosphorylation sites. *J Biomed Sci*. 2017;24:3. [PubMed: 28056995]
39. 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. [PubMed: 28408400]

40. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483:603–7. [PubMed: 22460905]
41. Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods*. 2017;14:407–10. [PubMed: 28218898]
42. Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, et al. Directional DNA Methylation Changes and Complex Intermediate States Accompany Lineage Specificity in the Adult Hematopoietic Compartment. *Molecular Cell*. 2011;44:17–28. [PubMed: 21924933]
43. Garaycochea JI, Crossan GP, Langevin F, Mulderrig L, Louzada S, Yang F, et al. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature*. 2018;553:171–7. [PubMed: 29323295]
44. Gasparetto M, Pei S, Minhajuddin M, Khan N, Pollyea DA, Myers JR, et al. Targeted therapy for a subset of acute myeloid leukemias that lack expression of aldehyde dehydrogenase 1A1. *Haematologica*. 2017;102:1054–65. [PubMed: 28280079]
45. Gerber JM, Smith BD, Ngwang B, Zhang H, Vala MS, Morsberger L, et al. A clinically relevant population of leukemic CD34⁺CD38⁻ cells in acute myeloid leukemia. *Blood*. 2012;119:3571–7. [PubMed: 22262762]
46. Hoang VT, Buss EC, Wang W, Hoffmann I, Raffel S, Zepeda-Moreno A, et al. The rarity of ALDH⁺ cells is the key to separation of normal versus leukemia stem cells by ALDH activity in AML patients: Aldehyde dehydrogenase in acute myeloid leukemia. *Int J Cancer*. 2015;137:525–36. [PubMed: 25545165]
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 Leukemic Stem Cells. *Pant AB, editor. PLoS ONE*. 2013;8:e78897. [PubMed: 24244383]
48. Guo J-M, Liu A-J, Zang P, Dong W-Z, Ying L, Wang W, et al. ALDH2 protects against stroke by clearing 4-HNE. *Cell Res*. 2013;23:915–30. [PubMed: 23689279]
49. Czerwińska J, Poznański J, Dąbski J, Bukowy Z, Bohr VA, Tudek B, et al. Catalytic activities of Werner protein are affected by adduction with 4-hydroxy-2-nonenal. *Nucleic Acids Research*. 2014;42:11119–35. [PubMed: 25170083]
50. Voulgaridou G-P, Anastopoulos I, Franco R, Panayiotidis MI, Pappa A. DNA damage induced by endogenous aldehydes: Current state of knowledge. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2011;711:13–27. [PubMed: 21419140]
51. Cremona CA, Behrens A. ATM signalling and cancer. *Oncogene*. 2014;33:3351–60. [PubMed: 23851492]
52. Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA. DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer*. 2008;8:193–204. [PubMed: 18256616]
53. Sullivan KD, Padilla-Just N, Henry RE, Porter CC, Kim J, Tentler JJ, et al. ATM and MET kinases are synthetic lethal with nongenotoxic activation of p53. *Nat Chem Biol*. 2012;8:646–54. [PubMed: 22660439]
54. Santos MA, Faryabi RB, Ergen AV, Day AM, Malhowski A, Canela A, et al. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature*. 2014;514:107–11. [PubMed: 25079327]
55. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827–32. [PubMed: 23873081]
56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. [PubMed: 23104886]
57. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. [PubMed: 25516281]
58. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn J, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31:46–53. [PubMed: 23222703]
59. Somerville TCP, Matheny CJ, Spencer GJ, Iwasaki M, Rinn JL, Witten DM, et al. Hierarchical Maintenance of MLL Myeloid Leukemia Stem Cells Employs a Transcriptional Program Shared

- with Embryonic Rather Than Adult Stem Cells. *Cell Stem Cell*. 2009;4:129–40. [PubMed: 19200802]
60. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012;7:1728–40. [PubMed: 22936215]
61. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002;12:996–1006. [PubMed: 12045153]
62. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted Nanopore Sequencing with Cas9 for studies of methylation, structural variants, and mutations [Internet]. *Genomics*; 2019. Available from: <http://biorxiv.org/lookup/doi/10.1101/604173>
63. Iyer SV, Goodwin S, Kramer M, McCombie WR. Understanding genetic variation in cancer using targeted nanopore long read sequencing [abstract]. In: Proceedings of the Annual Meeting of the American Association for Cancer Research 2020. 2020 Apr 27–28 and Jun 22–24. Philadelphia (PA): AACR; Cancer Res 2020;80(16 Suppl):Abstract nr 1360.
64. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Biol I*, editor. *Bioinformatics*. 2018;34:3094–100. [PubMed: 29750242]
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9. [PubMed: 19505943]
66. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012;13:R87. [PubMed: 23034086]
67. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2. [PubMed: 20110278]
68. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9. [PubMed: 27207943]
69. Verhaak RGW, Wouters BJ, Erpelinck CAJ, Abbas S, Beverloo HB, Lugthart S, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94:131–4. [PubMed: 18838472]
70. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005;102:15545–50.
71. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N Engl J Med*. 2013;368:2059–74. [PubMed: 23634996]

STATEMENT OF 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.

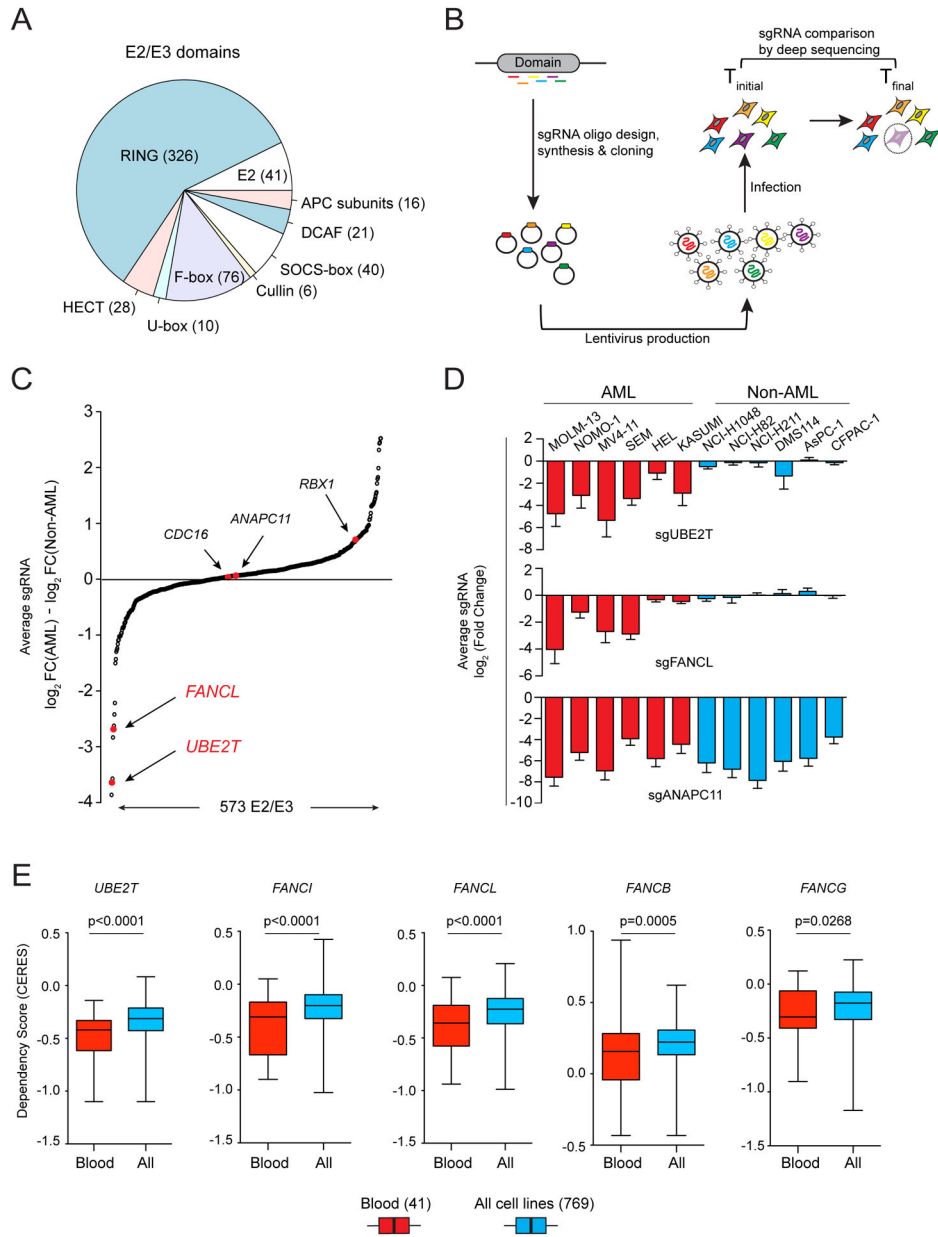


Figure 1. Domain-Focused CRISPR screening identifies UBE2T and FANCL as AML-biased dependencies.

(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 \log_2 fold-change difference comparing six AML and six non-AML cell lines in the pooled CRISPR screening. (D) Average \log_2 fold-change for sgRNAs targeting UBE2T, FANCL, and ANAPC11 in the twelve screened cell lines. (E) Dependency scores of five FA genes extracted from Project Achilles (20Q2) (21,22). The boxplots 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’s t-test. All sgRNA experiments were performed in Cas9-expressing cell lines.

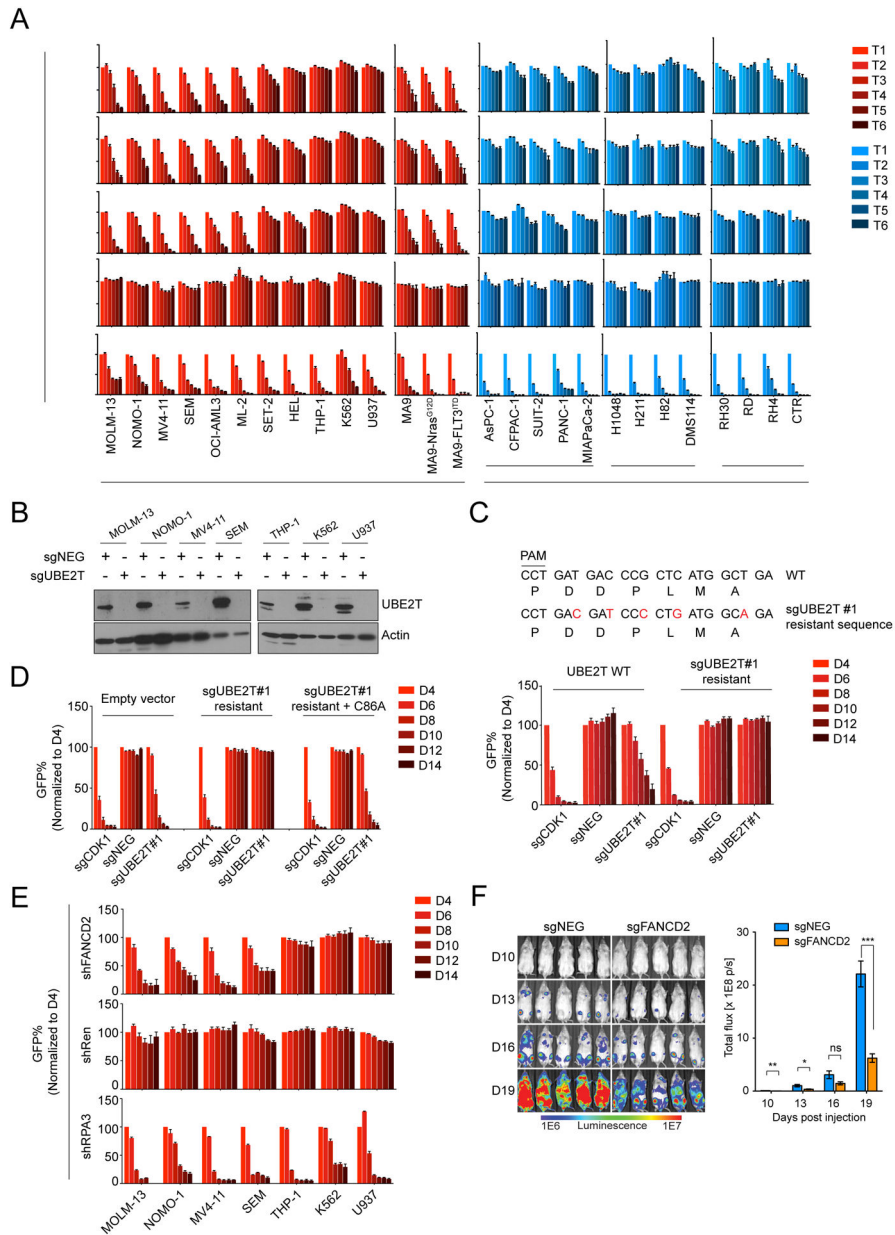


Figure 2. FA proteins are AML-biased dependencies *in vitro* and *in vivo*.

(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 non-leukemia lines. Rate of GFP depletion indicates loss of cell fitness caused by Cas9/sgRNA-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 co-transduction with UBE2T WT or sgRNA-resistant UBE2T cDNA. The sgRNA-resistant cDNA mutagenesis strategy is

depicted. (n=3). **(D)** Competition-based assays measuring sgRNA (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 FANCD2. (n=3). **(F)** (Left) Bioluminescence imaging of NSG mice transplanted with luciferase+/Cas9+ MOLM-13 cells infected with either sgNEG or sgFANCD2. (Right) Quantification of bioluminescence intensity. (n=5). All bar graphs represent the mean \pm SEM. p-value is calculated by unpaired Student's t-test. ***p < 0.001, **p < 0.01, *p < 0.05. All sgRNA experiments were performed in Cas9-expressing cell lines.

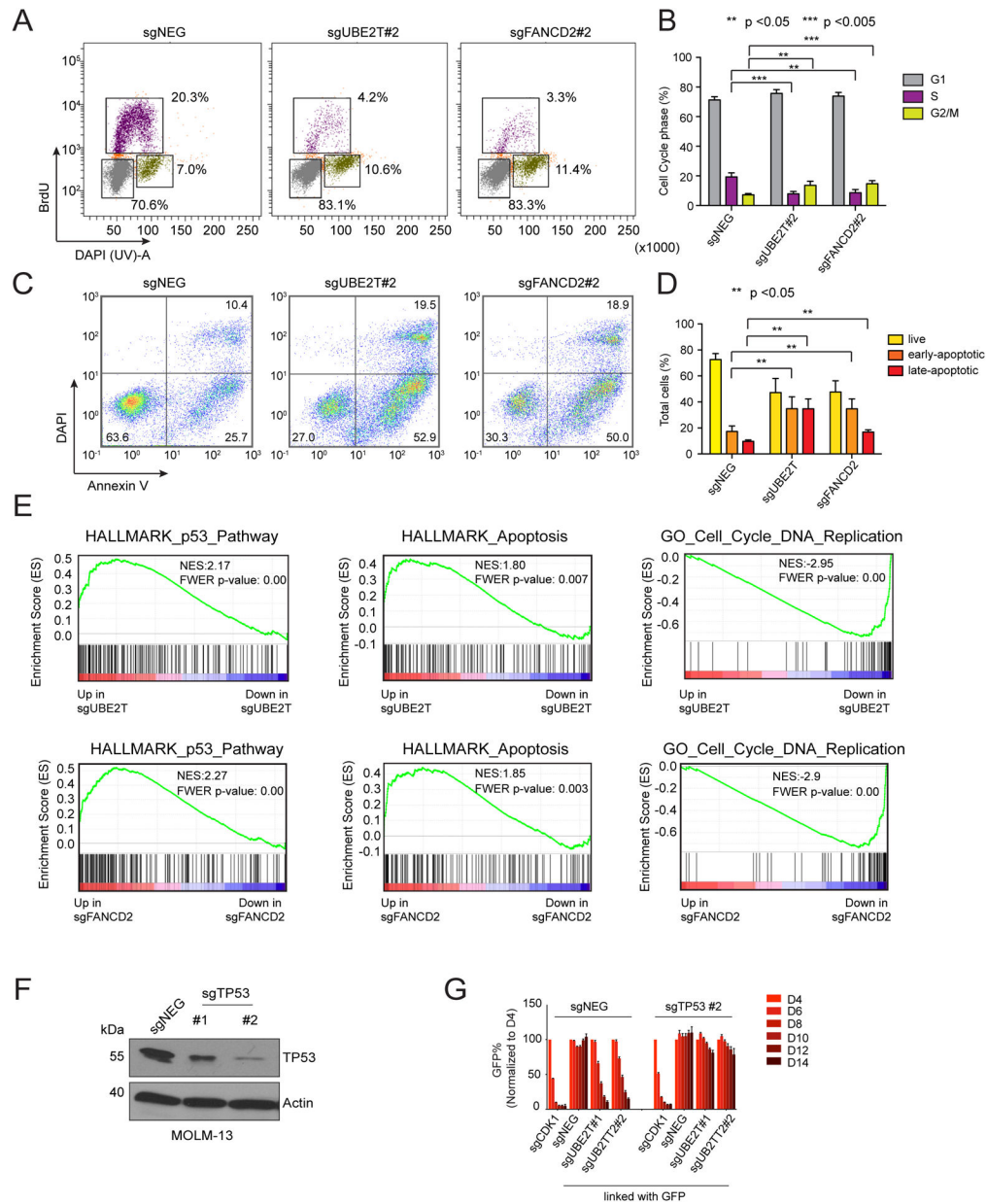


Figure 3. Inactivation of Fanconi anemia genes in AML leads to p53-induced cell cycle arrest and apoptosis.

(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 different cell cycle stages, average of 3 biological replicates. Paired Student's t-test was applied to calculate p values. (C) Representative flow cytometry analysis of DAPI (indicating permeable dead cells) and annexin-V staining (a pre-apoptotic cell marker) following lentiviral transduction of MOLM-13 cells (day 6). (D) Quantification of live and apoptotic cells, average of 3 biological replicates. Paired Student's t-test was applied to calculate p values. (E) Gene set enrichment analysis (GSEA) 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 \pm SEM. All sgRNA experiments were performed in Cas9-expressing cell lines.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

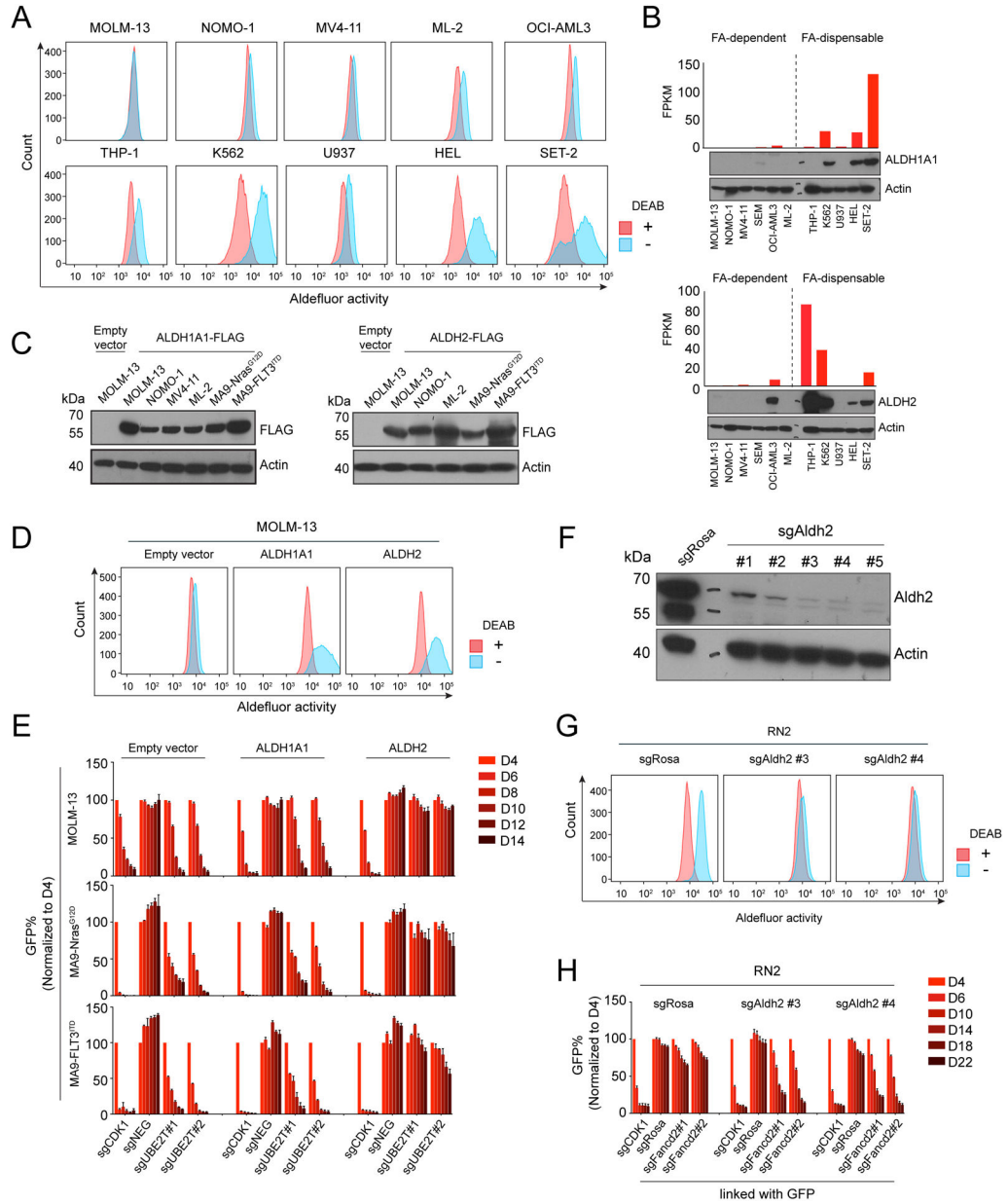


Figure 4. Silencing of ALDH2 expression in AML confers Fanconi anemia protein dependency
(A) Representative flow cytometry analysis of Aldefluor assays, in which AML cell lines were treated with 3.75 μ M BAAA, with or without the ALDH inhibitor DEAB (15 μ M). **(B)** RNA and protein levels of *ALDH1A1* 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 transduced with the indicated lentiviral cDNA constructs. **(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 \pm SEM, (n=3). All sgRNA experiments were performed in Cas9-expressing cell lines.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

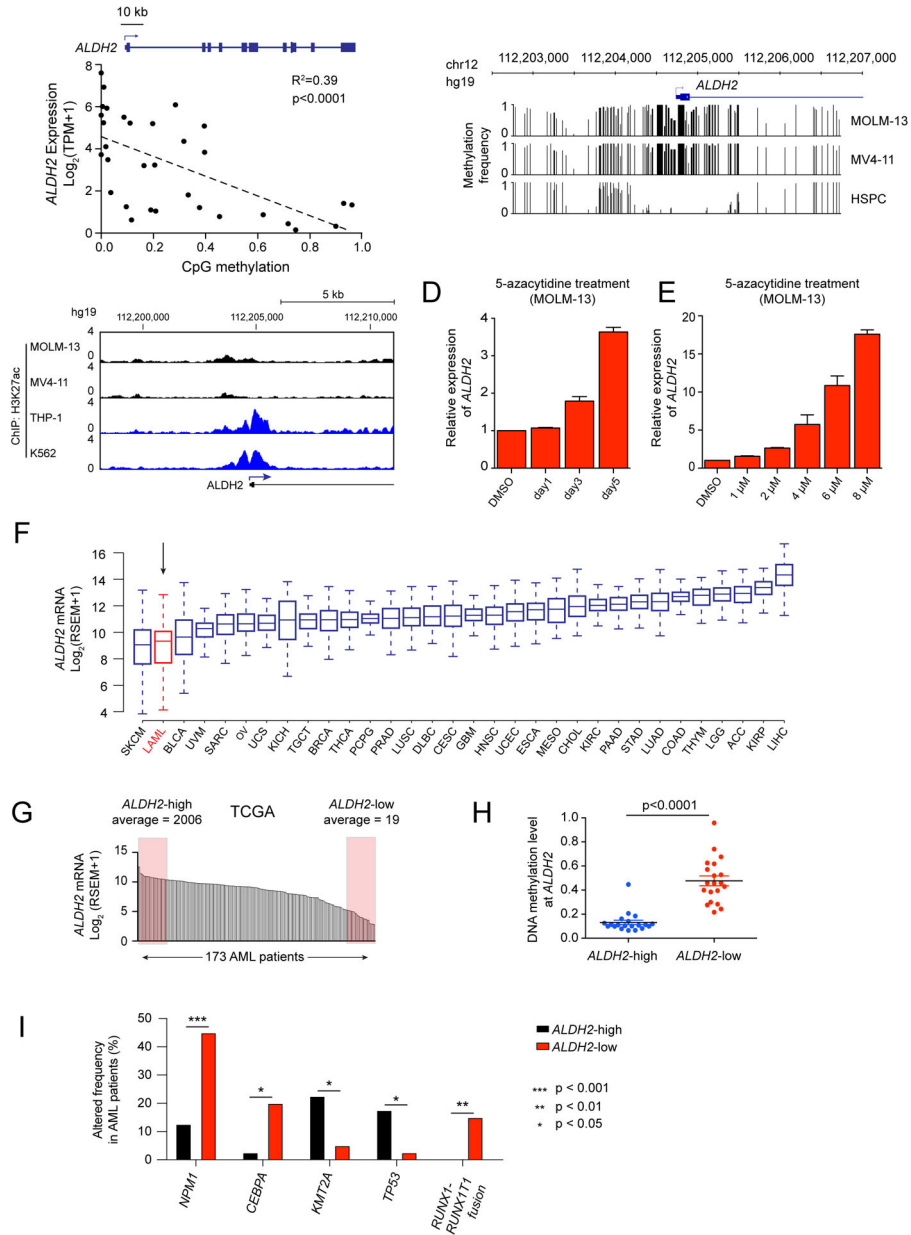


Figure 5. Silencing of *ALDH2* in human AML cell lines and patient samples is associated with promoter DNA hypermethylation.

(A) Scatterplot analysis comparing *ALDH2* mRNA levels and DNA methylation levels at the *ALDH2* promoter across 32 AML cell lines (data was obtained from the Cancer Cell Line Encyclopedia) (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 (42)). (C) ChIP-seq analysis of H3K27ac enrichment at the *ALDH2* promoter. (D-E) RT-qPCR analysis of *ALDH2* mRNA levels in MOLM-13 cells following treatment with 5-azacytidine. In (D), 1 μ M 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 is calculated by unpaired Student's t-test. **(I)** Genetic alterations that correlate with ALDH2 silencing in human AML patients. 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's t test was performed to determine statistical significance, which revealed the five genes shown as being significantly correlated or anti-correlated with ALDH2 silencing. All bar graphs represent the mean \pm SEM (n=3).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript