Supplementary Material for

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

Zhaolin Yang[1], Xiaoli S. Wu[1,2], Yiliang Wei[1], Sofya A. Polyanskaya[1], Shruti V. Iyer[1,2], Moonjung Jung[3], Francis P. Lach[3], Emmalee R. Adelman[4], Olaf Klingbeil[1], Joseph P. Milazzo[1], Melissa Kramer[1], Osama E. Demerdash[1], Kenneth Chang[1], Sara Goodwin[1], Emily Hodges[5], W. Richard McCombie[1], Maria E. Figueroa[4], Agata Smogorzewska[3], and Christopher R. Vakoc[1,6*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Genetics Program, Stony Brook University, Stony Brook, New York 11794, USA
[3]Laboratory of Genome Maintenance, The Rockefeller University, New York 10065, USA
[4]Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL 33136, USA
[5]Department of Biochemistry and Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN 37232, USA
[6]Lead contact
*Correspondence: vakoc@cshl.edu; Christopher R. Vakoc, 1 Bungtown Rd, Cold Spring Harbor, NY 11724

Supplementary Figures with Legends

Supplementary Figures S1. Domain-focused CRISPR screening identifies UBE2T and FANCL as AML dependencies.

Supplementary Figures S2. Validation of FA dependencies in AML maintenance.

Supplementary Figures S3. In vivo evaluation of FANCL, UBE2T, and FANCD2 dependency in MOLM-13 cells.

Supplementary Figures S4. Targeting of FA genes leads to DNA damage.

Supplementary Figures S5. Additional experiments evaluating connection between FA dependency and TP53.

Supplementary Figures S6. TP53 and CDKN1A inactivation alleviates FA dependency.

Supplementary Figures S7. Additional analysis linking ALDH2 silencing with FA dependency in AML.

Supplementary Figures S8. The catalytic activity of ALDH2 is required to bypass FA protein dependency in AML.

Supplementary Figures S9. Re-expression of ALDH2 in MOLM-13 leads to minimal changes in cell proliferation and transcription, but leads to suppression of endogenous aldehydes.

Supplementary Figures S10. Re-expression of ALDH2 in MOLM-13 leads to minimal changes in cell proliferation and transcription, but leads to suppression of endogenous aldehydes.

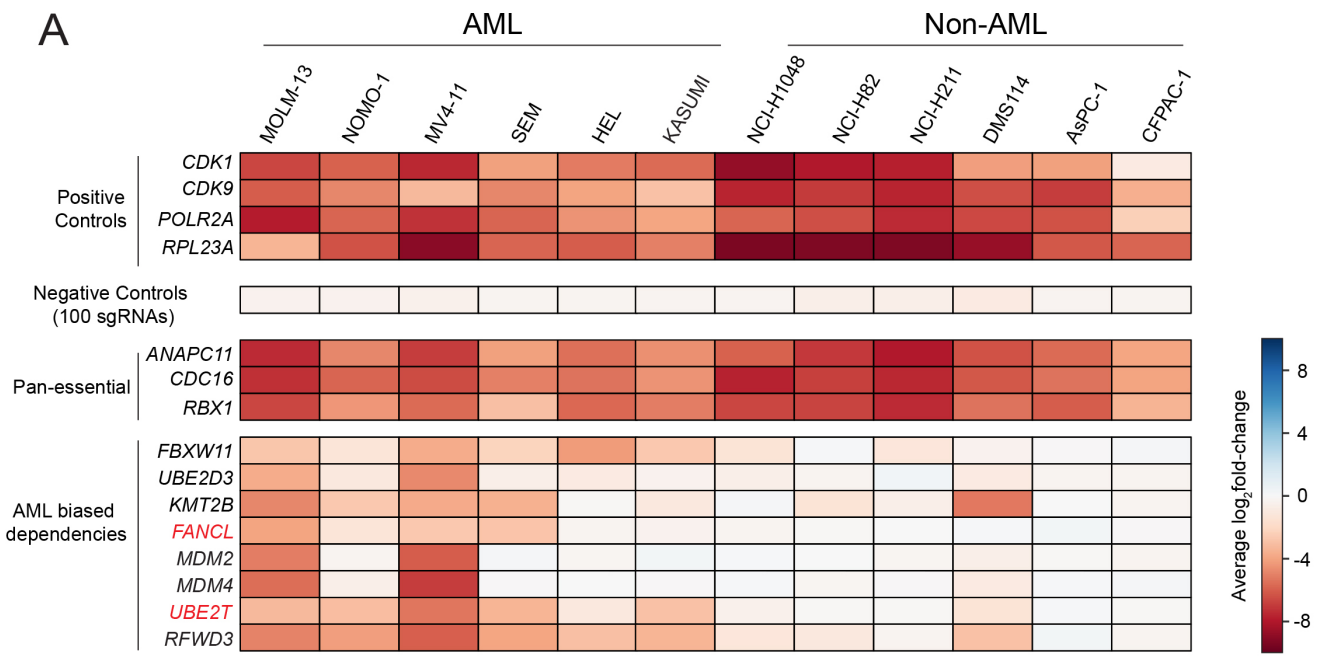
Supplementary Figures S11. ALDH2 expression is silenced by aberrant DNA methylation in AML.

Supplementary Tables

Supplementary Table S1. The E2/E3 sgRNA guide sequences.

Supplementary Table S2. Sequences of additional sgRNA and primers

Supplementary Table S3. CRISPR screening data using the E2/E3 sgRNA library in 12 cell lines.

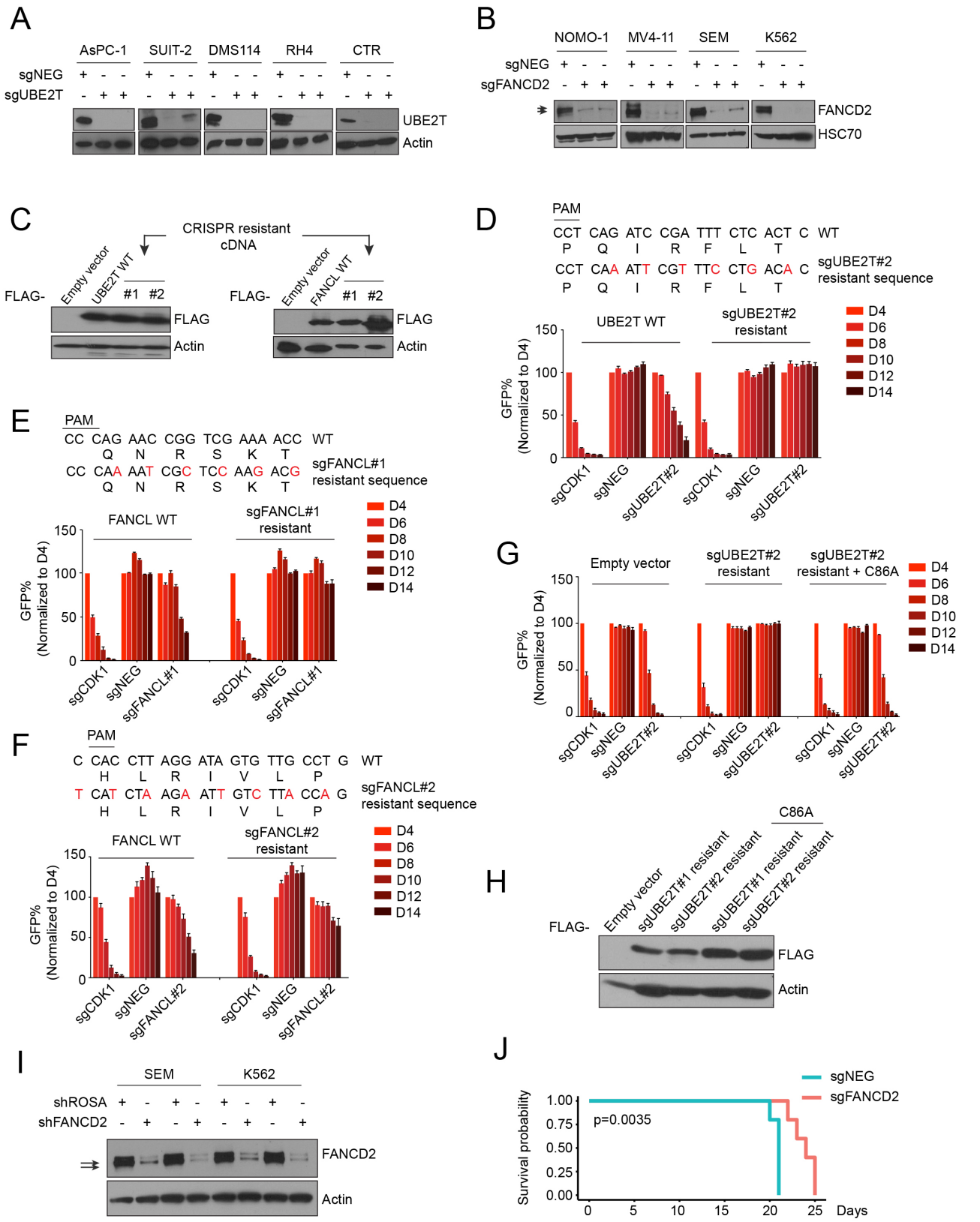


B

Top 20 Co-dependencies of UBE2T CRISPR across 769 cell lines

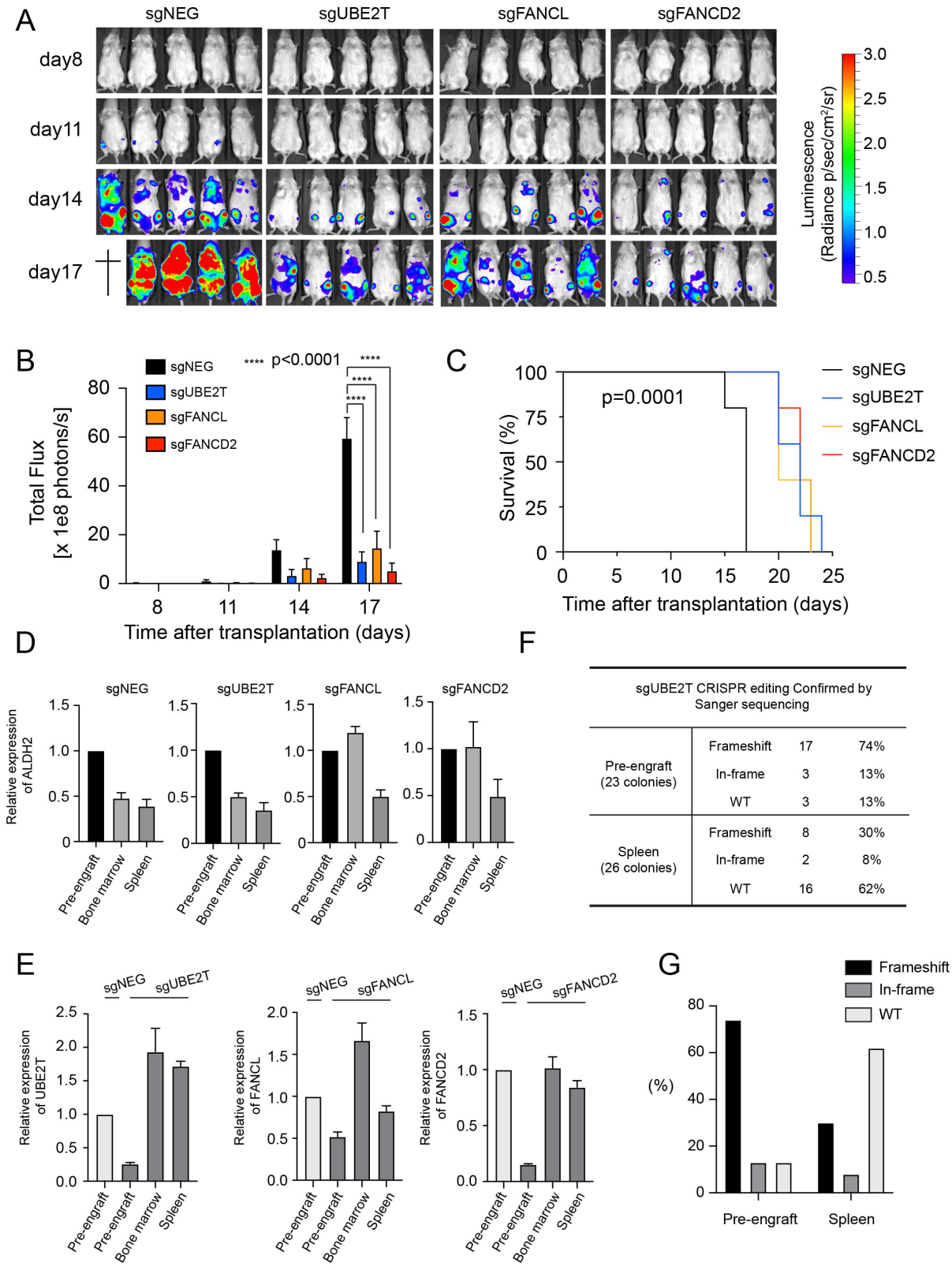
	Gene	Pearson correlation		Gene	Pearson correlation		Gene	Pearson correlation		Gene	Pearson correlation
1	***FANCI***	0.63	6	*BRIP1*	0.45	11	*RFWD3*	0.35	16	*ERCC1*	0.30
2	***FANCL***	0.58	7	***FANCD2***	0.43	12	*ERCC4*	0.35	17	*FAAP24*	0.21
3	***FANCB***	0.51	8	*FAAP100*	0.42	13	***FANCM***	0.34	18	*RAD51D*	0.20
4	***FANCG***	0.49	9	***FANCA***	0.42	14	***FANCC***	0.31	19	*CDC73*	0.210
5	***FANCF***	0.47	10	***FANCE***	0.36	15	*SLX4*	0.30	20	*PPP1R35*	0.19

Supplementary Figure S1. Domain-focused CRISPR screening identifies UBE2T and FANCL as AML dependencies. **(A)** Summary of the fold-depletion of spike-in positive and negative control sgRNAs in the pooled ubiquitination domain-focused CRISPR screens in 12 cell lines. In addition, the average $\log_2$ fold-change is shown for 3 pan-essential genes and for the top 8 AML-biased dependencies nominated from the screens. **(B)** The list of top co-dependencies of UBE2T across 769 cell lines from Project Achilles (20Q3). The genes that belong to Fanconi anemia pathway are highlighted in bold.



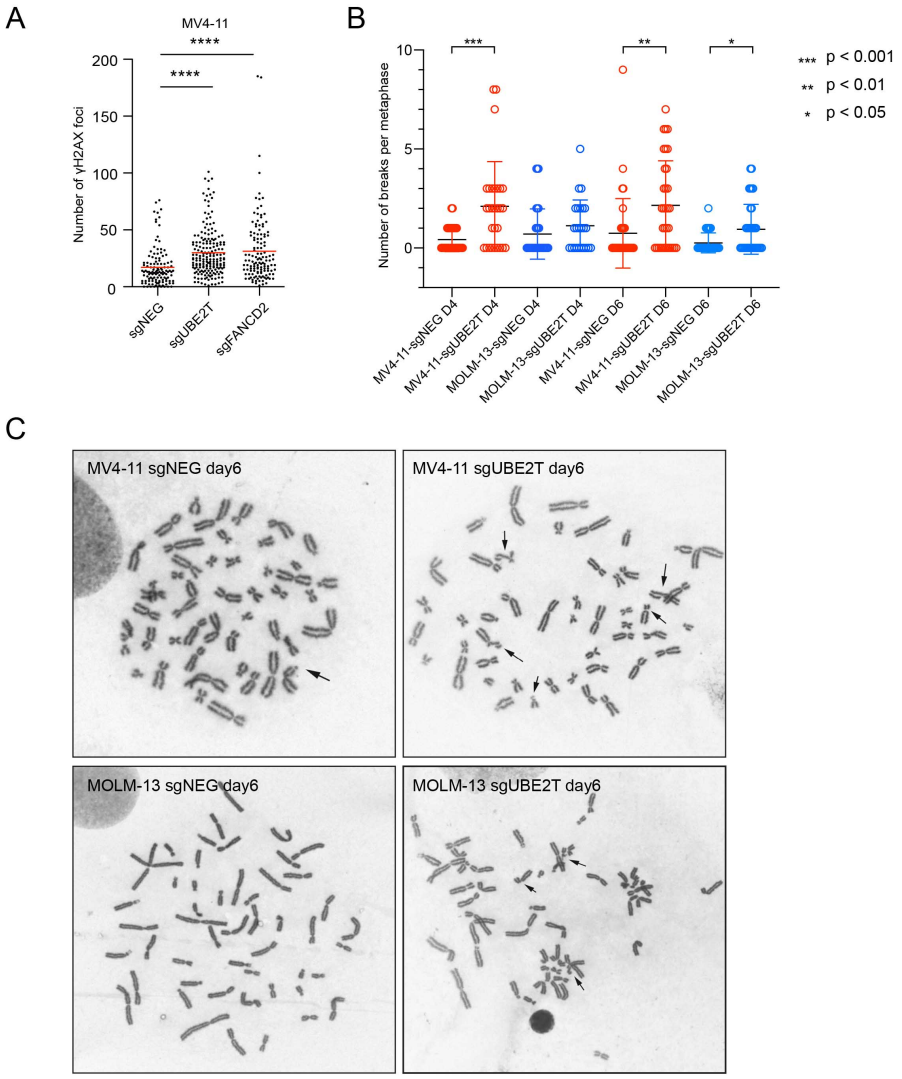
Supplementary Figure S2. Validation of FA dependencies in AML maintenance. (A-B) Western blot performed on whole cell lysates of the indicated cell lines following lentiviral transduction with indicated sgRNAs. **(C)** Western blot analysis in MOLM-13 cells of FLAG-tagged UBE2T (left) or FANCL (right) with wild-type sequence or harboring silent mutations of the sgRNA recognition site. **(D-G)** Competition-based proliferation assay in MOLM-13 cells expressing the indicated cDNAs (wild or CRISPR-resistant, labeled at the top) transduced with GFP-linked sgRNAs (labeled at the bottom). (n=3). **(H)**

Western blot analysis of FLAG-tagged UBE2T in MOLM-13 cells, carrying wild-type or various mutations. **(I)** Western blot analysis following transduction with shRNAs targeting FANCD2. **(J)** Kaplan-Meier survival curves of NSG recipient mice transplanted with MOLM-13 cell infected with indicated sgRNA. Each sgRNA group contains 5 mice. A log-rank (Mantel-Cox) statistical test was used to calculate the p-value. All bar graphs represent the mean ± SEM.



Supplementary Figure S3. In vivo evaluation of FANCL, UBE2T, and FANCD2 dependency in MOLM-13 cells. A) Bioluminescence imaging of NSG mice transplanted with luciferase+/Cas9+ MOLM-13 cells infected with indicated sgRNAs (LRG 2.1T-Blast). **B)** Quantification of bioluminescence intensity. (n=5). All bar graphs represent the mean ± SEM. p-value is calculated using unpaired Student's t-test. **C)** Kaplan-Meier survival curves of NSG recipient mice transplanted with MOLM-13 cell infected with indicated sgRNA. Each sgRNA group contains 5 mice. A log-rank (Mantel-Cox) statistical test

was used to calculate the p-value. All bar graphs represent the mean ± SEM. **D)** RT-qPCR measurement of ALDH2 mRNA levels (normalized to Beta2-microglobulin) in pre-transplanted and in cells extracted from spleen/bone marrow MOLM-13 cells expressing the indicated sgRNAs (selected with blasticidin). **E)** RT-qPCR measurement of UBE2T/FANCL/FANCD2 mRNA levels (normalized to Beta2-microglobulin) in pre-transplanted and in cells extracted from spleen/bone marrow MOLM-13 cells expressing the indicated sgRNAs (selected with blasticidin). sgRNA expression reduces each gene (presumably via non-sense mediated decay), but the expression becomes restored in d20-25 leukemia cells collected from bone marrow/spleen. **F)** PCR-amplified sgRNA target sites were cloned and Sanger sequenced to quantify relative abundance of WT and mutated sequences. **G)** Summary of mutation abundance from (F).



Supplementary Figure S4. Targeting of FA genes leads to DNA damage. (A) The numbers of phospho-H2AX foci in MV4-11-Cas9 cells infected with the indicated sgRNAs. Data were combined from three independent experiments. (sgNEG n=121, sgUBE2T n=184, and sgFANCD2 n=129). Statistical analysis was performed using one-way ANOVA followed by Dunnett's multiple comparison test (**** $p < 0.0001$; ns, not significant). **(B)** Quantification of metaphase breaks following inactivation of UBE2T in MOLM-13 and MV4-11 cells. Breakage analysis was blinded. p values were calculated using a one-way ANOVA test. **(C)** Select representative figures of metaphase spreads.

A

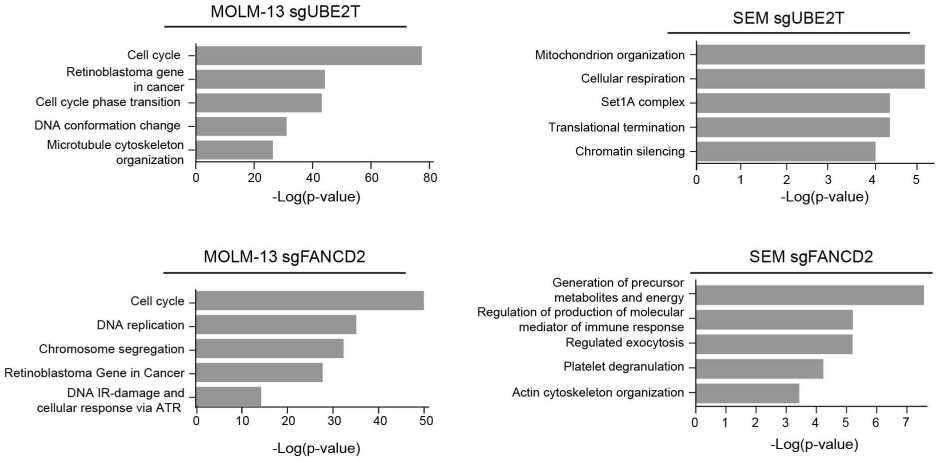
TP53 mutation status in AML cell lines

Cell line	CCLE	Sanger Sequencing	ALDH2 Expression	FA pathway depencency
MOLM-13	WT	WT	No	*****
NOMO-1	frameshift-del	frameshift-del	No	****
MV4-11	WT	WT	No	*****
SEM	R248Q	R248Q	No	****
Kasumi-1	R248Q	R248Q	No	***
THP-1	frameshift-del	frameshift-del	Yes	No
K562	frameshift-ins	frameshift-ins	Yes	No
HEL	M133K	M133K	Yes	No
SET-2	R248W	R248W	Yes	No
MA9	N/A	WT	No	****
MA9-Nras^{G12D}	N/A	WT	No	*****
MA9-FLT3^{ITD}	N/A	WT	No	*****
RN2	N/A	WT	Yes	No

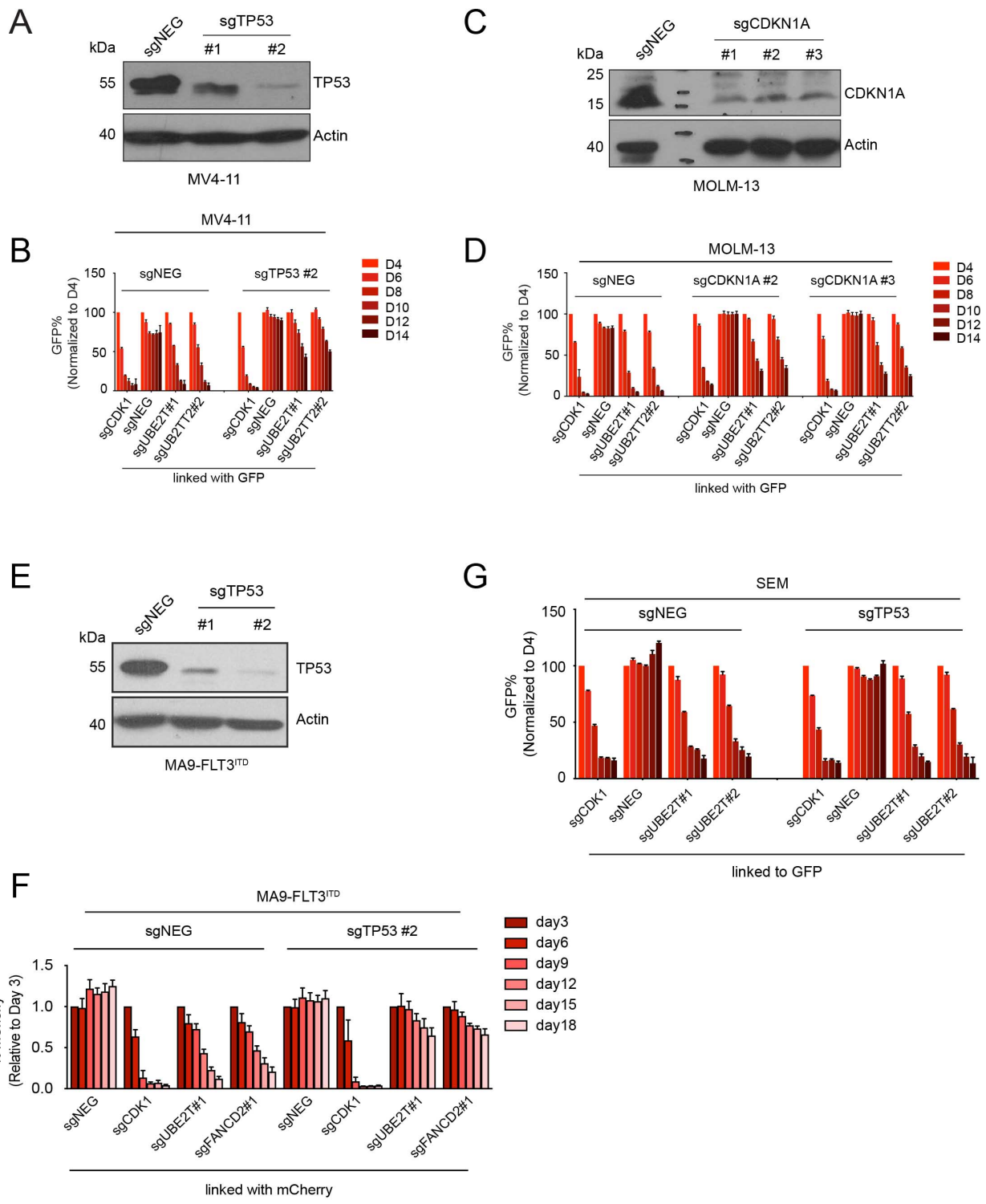
B

Cell line	Pertubation	HALLMARK_p53_Pathway		LSC_Somervaille	
		NES	FWER p-value	NES	FWER p-value
MOLM-13 (p53 wild type)	sgUBE2T	2.17	0.000	-2.86	0.000
	sgFANCD2	2.27	0.000	-2.59	0.000
SEM (p53 mut)	sgUBE2T	-1.30	0.391	-0.88	0.471
	sgFANCD2	-1.34	0.293	-0.80	0.488

C

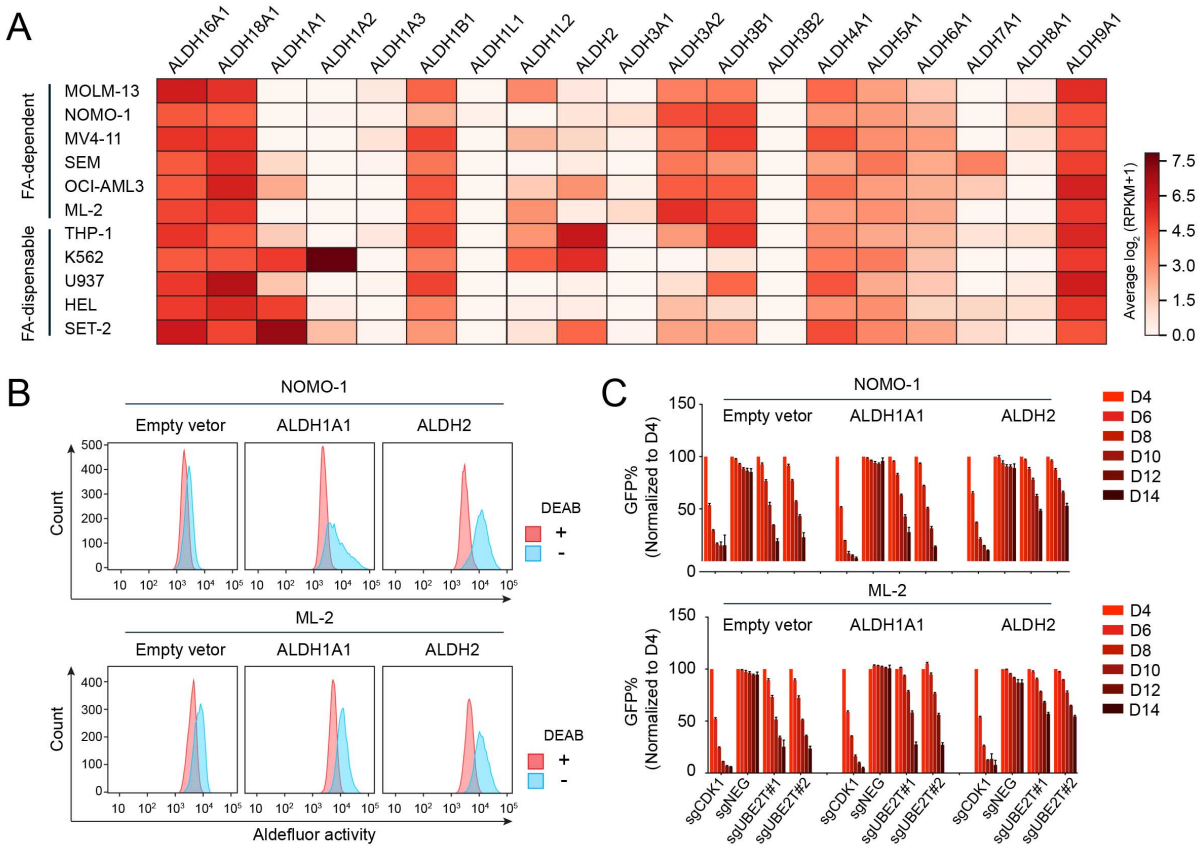


Supplementary Figure S5. **Additional experiments evaluating connection between FA dependency and TP53. (A)** TP53 mutation status of leukemia cell lines used in this study. The TP53 mutation status was extracted from the indicated databases and was confirmed by Sanger sequencing. ALDH2 expression was based on RNA-seq data. FA dependency column summarizes results from Figure 2A, 1D, and 4H. **(B)** Gene set enrichment analysis (GSEA) of RNA-seq data obtained from MOLM-13 or SEM cells lentivirally transduced with the indicated sgRNAs (Subramanian et al., 2005). Normalized enrichment score (NES) and family-wise error rate (FWER) *p*-value are shown. **(C)** Gene Ontology analysis of RNA-seq obtained from FA-deficient MOLM-13 and SEM cells. Significantly down regulated genes (p-value <= 0.1, $\log_2$FoldChange <= -0.5) were analyzed using Metascape. GO terms are ranked by their significance (p value).



Supplementary Figure S6. TP53 and CDKN1A inactivation alleviates FA dependency. (A) Western blot performed on whole cell lysates prepared from MV4-11 cells on day 6 following transduction with the indicated sgRNAs. **(B)** Competition-based proliferation assays in MV4-11 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. **(C)** Western blot analysis of whole cell lysates prepared from MOLM-13 cells on day 6 following transduction with the indicated sgRNAs. **(D)** Competition-based proliferation assays in MOLM-13 cells following sequential sgRNA transduction. Negative sgRNA or CDKN1A 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. **(E)** Western blot analysis performed on lysates obtained from MA9-FLT3^{ITD} cells on day 6 following TP53 sgRNA transduction. **(F)** Competition-based proliferation assays in MA9-FLT3^{ITD} 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 mCherry). n=3. All bar graphs represent the mean ± SEM. All sgRNA experiments were performed in Cas9-expressing cell lines. **(G)** Competition-based proliferation assays in SEM 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 mCherry). n=3. All bar graphs represent the mean ± SEM. All sgRNA experiments were performed in Cas9-expressing cell lines.



F

Pearson correlation of ALDH2 expression and Fanconi anemia pathway dependency

	FANCD2	FANCI	FANCL	UBE2T
liver	0.506	0.710	0.349	0.478
esophagus	0.317	0.493	0.372	0.535
cervix	0.733	0.648	0.183	0.086
urinary_tract	0.347	0.485	0.413	0.299
skin	0.080	0.178	0.306	0.186
ovary	0.138	0.227	0.153	0.223
soft_tissue	-0.162	0.254	0.250	0.352
bile_duct	-0.047	0.254	-0.144	0.275
gastric	0.034	0.265	0.071	-0.091
uterus	0.051	0.241	-0.088	-0.127

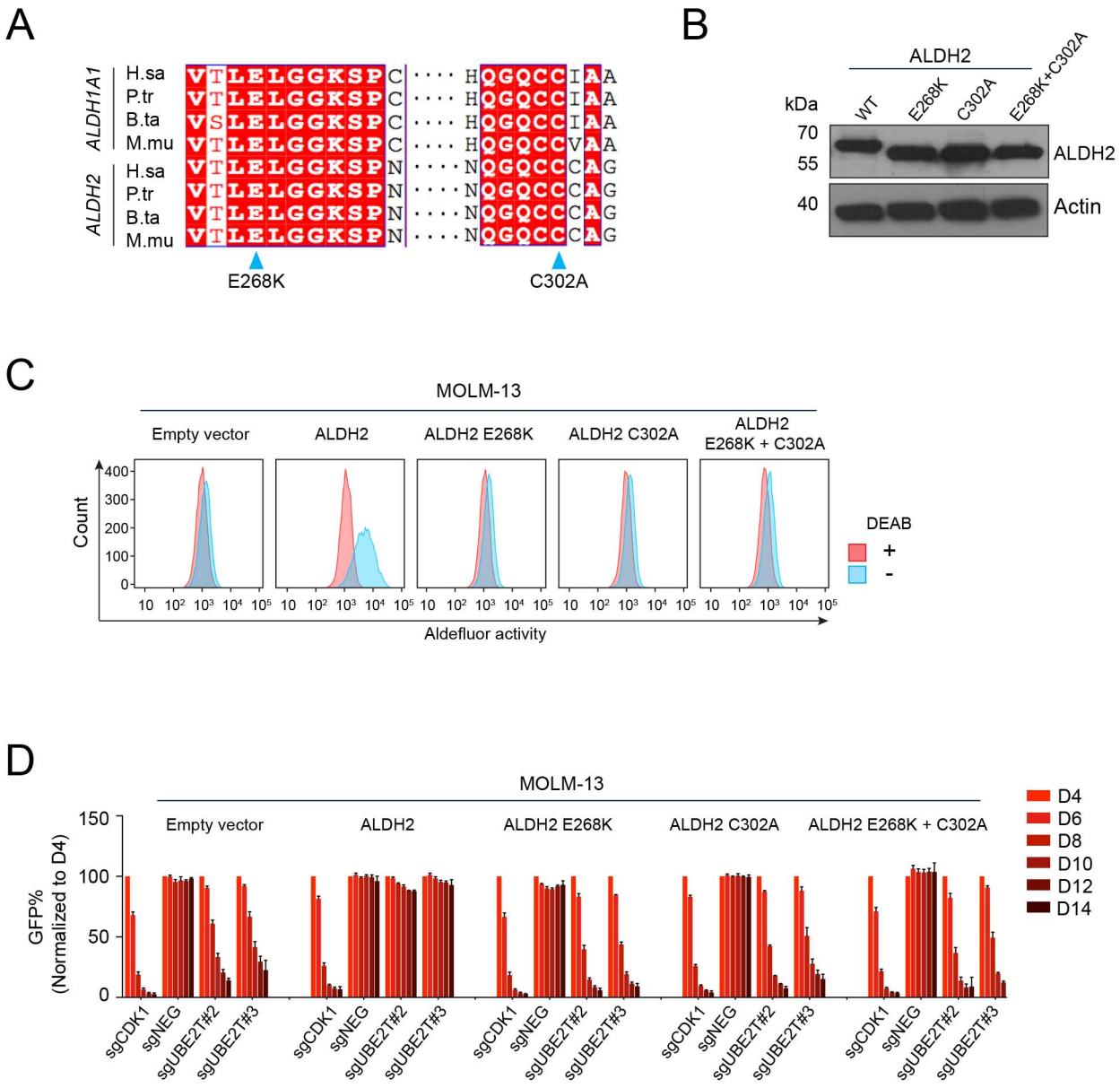
E

Two class comparison of Fanconi anemia pathway dependency in ALDH2 expression high/low cancer cell lines (solid tumors)

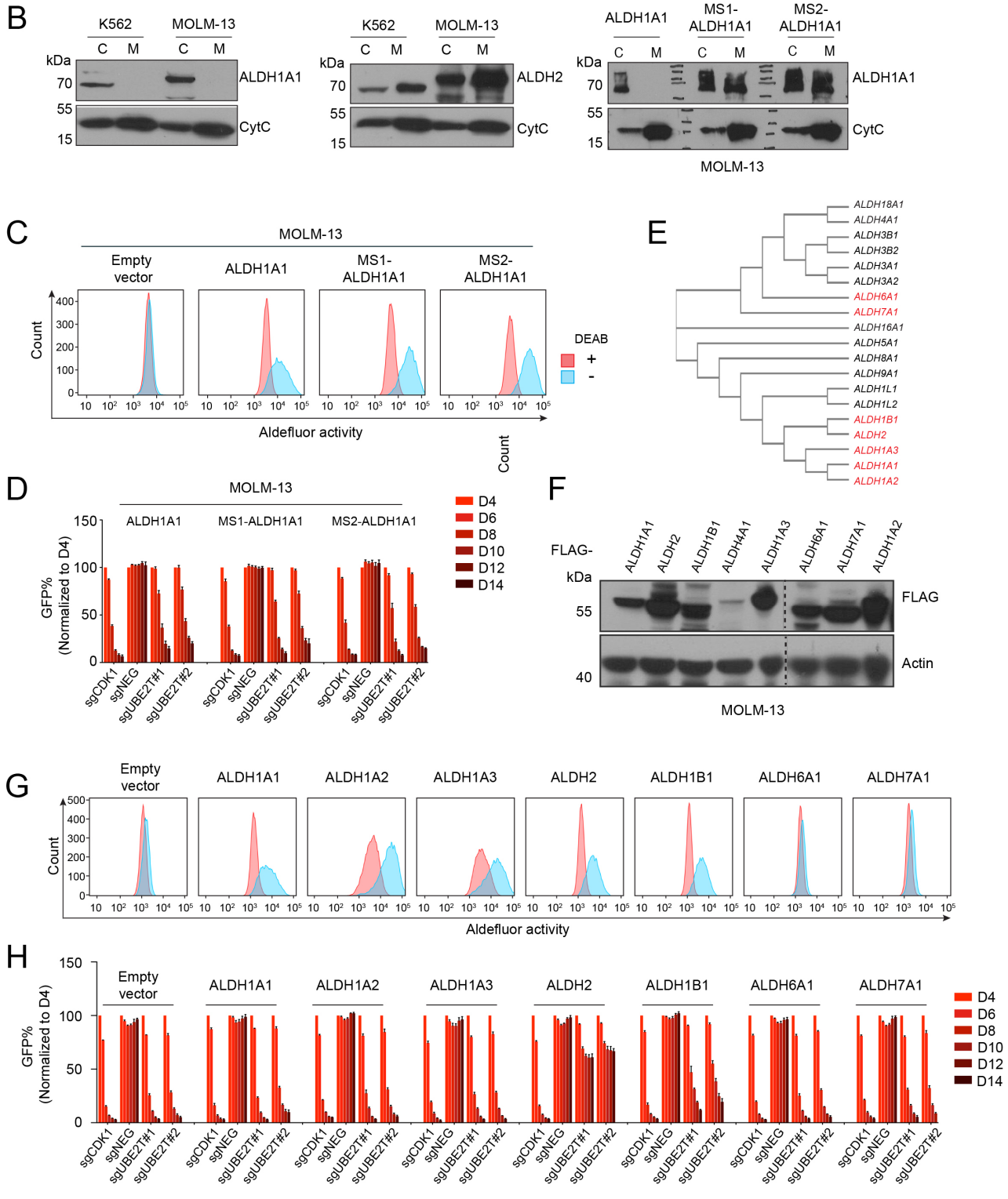
Ranking	Gene	Effect Size	p-Value	q-Value
5	FANCI	0.086	0.000	0.000
10	UBE2T	0.066	0.000	0.001
12	FANCB	0.053	0.000	0.003
29	FANCG	0.029	0.000	0.149
35	FANCL	0.026	0.000	0.103
67	FANCD2	0.015	0.001	0.208
570	FANCF	0.005	0.015	0.476

Supplementary Figure S7. Additional analysis linking ALDH2 silencing with FA dependency in AML. (A) RNA-seq analysis of mRNA levels for 19 ALDH genes in 11 leukemia cell lines. **(B)** Aldefluor analysis of NOMO-1 and ML-2 cells transduced with FLAG-ALDH1A1 or FLAG-ALDH2. Lines were treated with 3.75 μM BAAA, with or without the ALDH inhibitor DEAB (15 μM). **(C)** Competition-based proliferation assay in Cas9+ NOMO-1 and ML-2 cells expressing the

indicated cDNAs, followed by transduction with GFP-linked sgRNA. (n=3). **(D)** Western blot analysis of ALDH1A1 or ALDH2 proteins in the indicated leukemia cell lysates. **E)** For this analysis blood, engineered blood, lymphocytes, and plasma cell categories of cell lines were excluded. All remaining cell lines were then ranked based on their ALDH2 mRNA expression value and cells with the highest ALDH2 expression (n = 275) and lowest ALDH2 expression (n = 205) were categorized as ALDH2-high and ALDH2-low groups. The two classes comparison was performed using the DepMap portal (21Q1) to query genes that are differentially dependent in ALDH2-high and -low tumors. The effect size describes the correlation between gene dependency and ALDH2 mRNA level, with a higher value indicating a stronger dependency in ALDH2-low cancer cell lines. **F)** Cancer cell lines were first grouped based on lineage (using categories having at least 10 cancer cell lines). The Pearson correlation compares the dependency of a given gene (CERES essentiality scores) with ALDH2 mRNA level (TPM). A higher correlation suggests that the indicated gene is a dependency when ALDH2 mRNA is low.

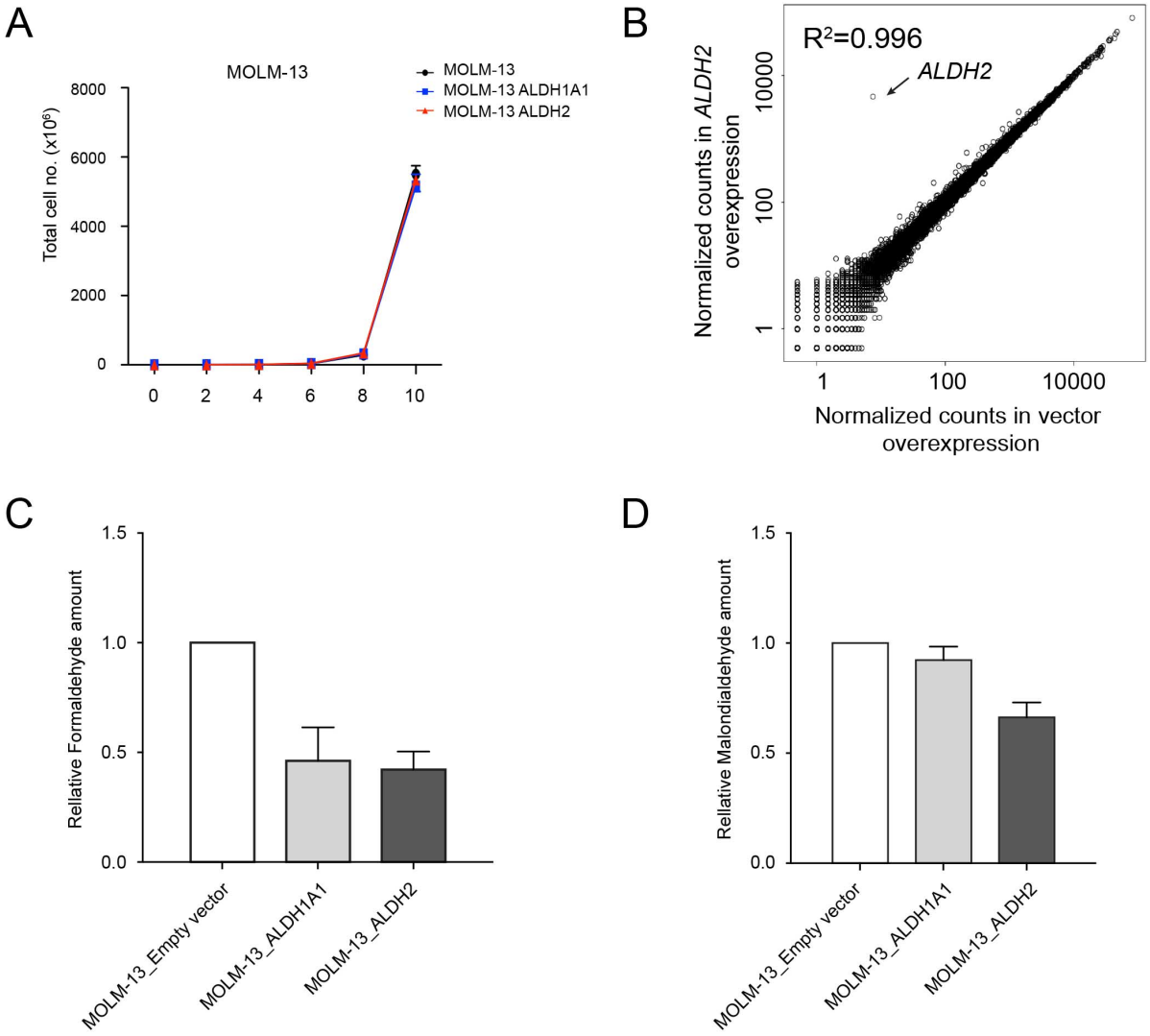


Supplementary Figure S8. The catalytic activity of ALDH2 is required to bypass FA protein dependency in AML. (A) Multiple sequence alignment of ALDH1A1 and ALDH2. The catalytic residues are highlighted by blue triangles. **(B)** Western blot analysis of wild-type and catalytic mutants FLAG-ALDH2 in MOLM-13 lysates. **(C)** Aldefluor analysis of MOLM-13 cells transduced with FLAG-ALDH1A1 or FLAG-ALDH2. Lines were treated with 3.75 μM BAAA, with or without the ALDH inhibitor DEAB (15 μM). **(D)** Competition-based proliferation assay in MOLM-13 cells expressing wild-type or catalytic mutant ALDH2 using the indicated sgRNAs linked to GFP.

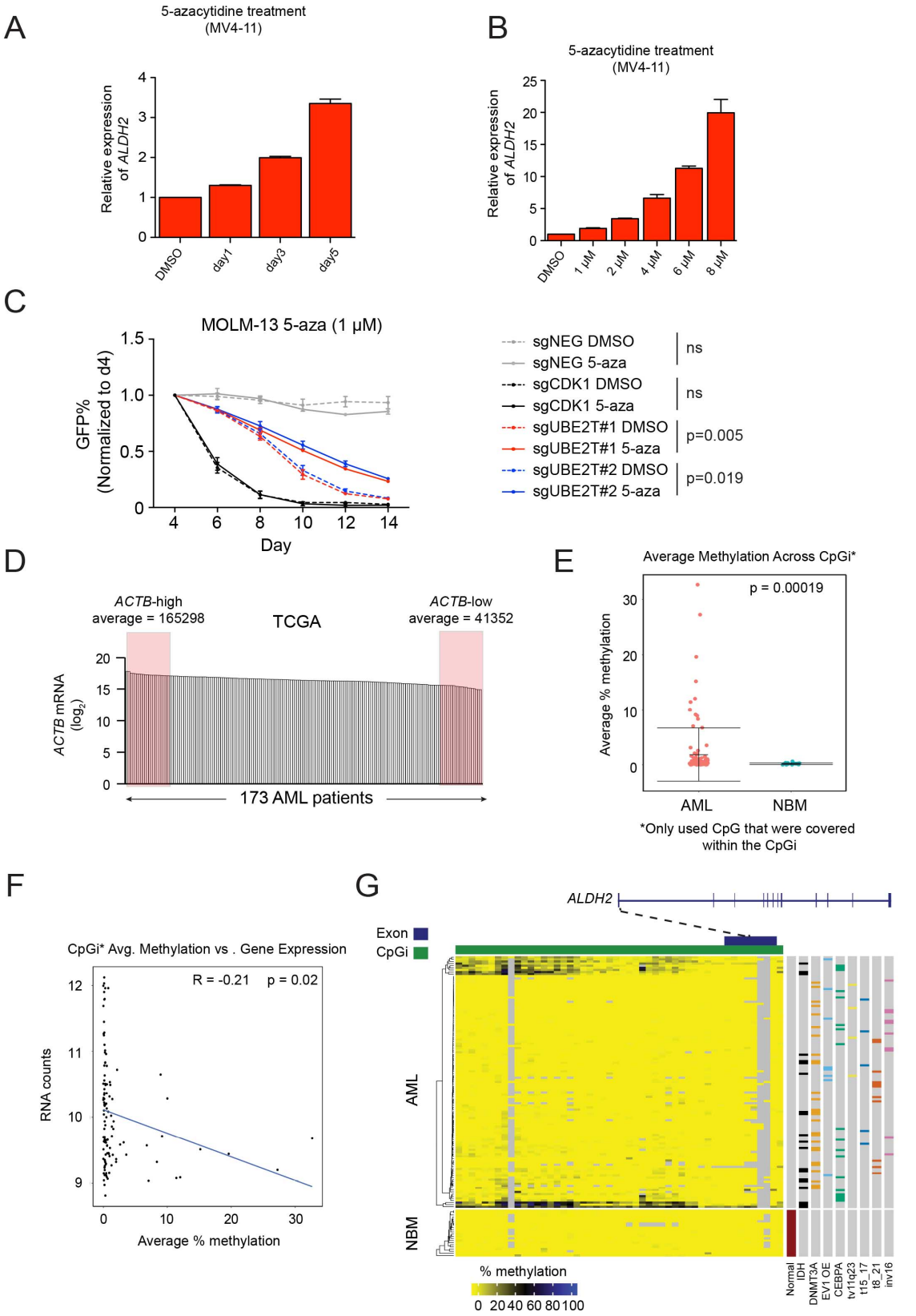


Supplementary Figure S9. ALDH2 is unique among ALDH enzymes in showing synthetic lethality with the Fanconi anemia complex. (A) Two sequences of mitochondrial localization signal peptides fused to ALDH1A1. **(B)** Western blot analysis of cytoplasm (C) or mitochondrial (M) fractions of indicated cell lines. **(C)** Aldefluor analysis of MOLM-13 cells

transduced with FLAG-ALDH1A1 or fusions with mitochondrial localization signal peptides (MS1 and MS2). Cells were treated with 3.75 μM BAAA, with or without the ALDH inhibitor DEAB (15 μM). **(D)** Competition-based proliferation assay in MOLM-13 cells expressing wild-type or mitochondrial localization signal peptide fused ALDH1A1 using the indicated sgRNAs linked to GFP. **(E)** Phylogenetic analysis of ALDH gene family. **(F)** Western blot analysis of whole cell lysates prepared from MOLM-13 cells transduced with the indicated FLAG-ALDH cDNAs. **(G)** Aldefluor assay of MOLM-13 cells transduced with the indicated ALDH genes. Lines were treated with 3.75 μM BAAA, with or without the ALDH inhibitor DEAB (15 μM). **(H)** Competition-based proliferation assay in MOLM-13 cells expressing the indicated ALDH genes. All bar graphs represent the mean ± SEM (n=3).



Supplementary Figure S10. Re-expression of ALDH2 in MOLM-13 leads to minimal changes in cell proliferation and transcription, but leads to suppression of endogenous aldehydes. (A) Cell proliferation *in vitro* over the course of 10 days of the indicated cell lines. (n=3). **(B)** RNA-seq scatter plot analysis comparing fold-change of mRNA levels following ALDH2 overexpression versus empty vector in MOLM-13 cells. **(C)** Fluorometric measurements of formaldehyde level measurements cell extracts. (n=3). **(D)** Fluorometric measurements of malondialdehyde levels in cell extracts. (n=3). All bar graphs represent the mean ± SEM (n=3).



Supplementary Figure S11. ALDH2 expression is silenced by aberrant DNA methylation in AML. (A-B) RT-qPCR analysis of ALDH2 mRNA levels in MV4-11 cells following treatment with 5-azacytidine. In (A), 1 μM concentration was used. In (B), a 36 hour timepoint was used. **(C)** Competition-based proliferation assays following sgRNA transduction, in the presence or absence of 1 μM 5-azacytidine. Each sgRNA is linked with GFP. n=3. All bar graphs represent the mean ± SEM.

All sgRNA experiments were performed in Cas9-expressing MOLM-13 cells lines. A two-way ANOVA statistical test was performed to calculate p-values. **(D)** ACTB mRNA levels in 173 AML patient samples from TCGA. Shown are the samples classified as ACTB-high and ACTB-low. **(E)** Barplot of the average percent methylation of the CpGi. Each dot is representative of one patient. p-value from two tailed Student's t-test. **(F)** Scatterplot of the average percent methylation of the CpGi versus the RNA expression level of ALDH2. Each dot is representative of one patient. Line of best fit was fit using a linear model. Correlation was calculated using Pearson's method in R. **(G)** Heatmap of the percent methylation of the covered CG within the *ALDH2* promoter. Each column is one CG and each row is one patient. Rows are clustered using Euclidian distances and the complex clustering method and missing CGs are depicted in gray. Select patient mutational and cytogenic data is plotted on the right of the heatmap.