Science Translational Medicine
AAAS

Supplementary Materials for

Splicing modulators impair DNA damage response and induce killing of cohesin-mutant MDS and AML

Emily C. Wheeler *et al.*

Corresponding author: Zuzana Tothova, Zuzana_Tothova@dfci.harvard.edu

Sci. Transl. Med. **16**, eade2774 (2024)
DOI: 10.1126/scitranslmed.ade2774

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:

SUPPLEMENTARY MATERIALS AND METHODS

Flow cytometry for U937 cells

U937 cells were single cell sorted with a FACSAriaII instrument (Becton Dickinson) after DAPI staining for viability (Thermo Fisher Scientific). $mCherry^+$ and GFP^+ U937 cells used for in vivo transplant studies and in vitro competition assays were sorted with a MoFlo Astrios EQ sorter (Beckman Coulter) or Sony SH800S Sony Cell Sorter. Readout of in vitro competition assays was performed using CytoFLEX (Beckman Coulter).

In vitro drug treatment and in vitro competition assays

All drug dose response assays were conducted in 96 well plates. For U937 and K562 cell lines, 10,000 cells were plated per well at $0.05x10^6$ cells/ml. Cells were split 1:4 by volume every 4 days and redosed with fresh medium supplemented with fresh drug every 4 days. For pre-treatment with DMSO, H3B-8800 (50nM), or E-7107 (1.14nM), cells were cultured in bulk for 3 days before plating in 96-well format for secondary drug dosing. Viability was measured every 4 days using the CellTiter-Glo luminescent cell viability assay (Promega G7573). A D300e drug dispenser (Tecan) was used for drug dosing in all experiments performed in 96 well at the concentrations indicated for each drug. Drug dose-response curves were fitted and IC50 concentrations were calculated using GraphPad Prism. All drugs used in this study are listed in the key resources table.

For cell counting assays at single drug dose concentrations, cells were cultured in 6-well dishes at the recommended concentration by cell type and split to equal concentration of cells for passaging. Drugs were redosed at the time of passage (3 or 4 days). Viable cell counts were measured in duplicate with a ViCell XR cell viability analyzer (Beckman Coulter).

For competition experiments, mCherry-labeled wild type and GFP-labeled STAG2-KO2 U937 cells were mixed in a 1:10 ratio and plated in 96-well plates with 3 technical replicates. Cells were passaged and split

1:4 by volume every 4 days and redosed with fresh medium supplemented with fresh H3B-8800 every 4 days. A fraction of the cells were stained for viability with DAPI, and % $mCherry^{+}DAPI^{-}$ and $GFP^{+}DAPI^{-}$ cells were determined by flow cytometry.

RNA-seq analysis

U937 RNA-seq samples were aligned to the hg19 genome, and PDX RNA-seq seq samples were aligned to a combined hg19 and mm10 genome using STAR (version 2.7.3a) (*49*). PDX reads uniquely aligning to hg19 were used to quantify gene expression and splicing. Gene counts were generated using featurecounts function of the Rsubread package (version 2.0.1) (*50*)aligning reads to the Ensembl (GRCh37.87) basic transcript annotations. Differentially expressed genes were calculated using DESeq2 (version 1.26.0)(*51*), normalizing samples by sequencing depth. Genes with a $\log_2$ foldchange in expression greater than 0.5 and an adjusted p-value less than 0.001 were called as differentially expressed.

For U937 expressed gene annotations, Ensembl annotations (GRCh37.87) were filtered for basic annotations. Transcript expression in each RNA-seq sample was quantified using kallisto (version 0.45.1)(*52*). Transcripts with expression greater than 0.5 TPM in at least two samples were retained, giving 37,389 transcripts corresponding to 20,120 genes. These expressed transcript annotations were then used for analyzing differentially expressed genes and alternative splicing. We used our get gene annotations pipeline (*47*) to call the dominant transcript isoforms for each gene, using the RNA-seq data as input. Briefly, this pipeline clusters annotated TSSs within 100bp and summed TPM values in each cluster. The cluster with the highest expression was selected as the dominant cluster, and the transcript with the highest expression was selected as the dominant TSS. Then for transcripts stemming from the dominant TSS cluster, we cluster transcript end sites (TESs), and, similar to TSSs, selected the dominant TES cluster and TES. The transcript with the highest TPM was then selected as the dominant transcript model for counting the number of exons per gene. We focused our analysis on protein coding genes, and this yielded 11,442 dominant transcript models for further analysis.

For patient samples, differential gene expression was calculated with DESeq2 using a paired analysis, comparing the pre- and post- H3B-8800 sample per patient with the design formula of ~patient + treatment.

Splicing analysis

Alternative splicing was quantified using rMATS (version 4.1.0) (*53*) considering junction-spanning reads only. All comparisons to generate ΔPSI scores are as listed in the manuscript using biological replicates for each condition. For *STAG2*-mutant AML PDX1 samples with in vivo E-7107 treatment, bam files from three replicate samples of the E-7107 treated arm were merged to obtain one replicate to increase coverage. The E-7107 treated arm had dramatic depletion of the human cells, and therefore many fewer reads were obtained in those samples, so merging was performed to increase read depth on splice junctions. The vehicle treated arm had sufficient coverage in each replicate and therefore were kept separate for analysis.

We used custom scripts available on github to process the raw output files and remove overlapping splicing calls on the same event, selecting only the event call with the most inclusion reads (https://github.com/YeoLab/rbp-maps/blob/master/preprocessing_scripts/subset_rmats_junctioncountonly.py)(*48*). Next, unique events were further filtered to only include those with a minimum of 10 reads supporting both the inclusion and skipped isoforms in at least one condition. Significantly regulated splicing events (absolute ΔPSI > 5%, FDR < 0.05) were then called for each comparison of interest.

For clustering analysis, events that were called significant (FDR < 0.05, ΔPSI > 5%) in any comparison were considered and ΔPSI was quantified from those events across all comparisons that are included in the clustering analysis. kMeans clustering was performed with the with an optimal k and the ΔPSI scores were plotted per cluster in a heatmap. For the analysis comparing all U937 cells treated with 10nM and 30nM

H3B-8800, k=4 was used for the clustering and one cluster was dropped from the final analysis for a total of 3 displayed clusters. The removed cluster contained sporadic splicing events that were only significant (FDR < 0.05, ΔPSI > 5%) in one cell type or drug concentration and were not representative of the on-target effects of H3B-8800 in cells. For patient samples, alternative splicing was quantified with rMATS for each patient individually comparing the pre- and post- H3B-8800 treatment sample. All filtering steps to call significantly regulated splicing changes (FDR < 0.05, ΔPSI > 5%) were performed as described above.

Western blotting

Cells were washed in cold 1X PBS (Corning CellGro) and lysed for 15-30min on ice using NP-40 lysis buffer (150mM NaCl, 50mM Tris pH 7.5 (Thermo Fisher Scientific 15567027), 1% NP40 (Roche 11332473001), 5% glycerol (Sigma-Aldrich G5516), and 100X Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific 78447) in water). Lysates were quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific 23225) and reduced with 6X Laemmli SDS-Sample Buffer (Boston BioProducts BP-111R) and 10X NuPAGE Sample Reducing Agent (Thermo Fisher Scientific NP0009) for 5min at 95°C.

Samples were run on NuPAGE 4-12% Bis-Tris Protein Gels (Invitrogen NP0321 and NP0316) in NuPAGE MOPS SDS Running Buffer 20X (Life Technologies NP0001) or NuPAGE MES SDS Running Buffer 20X (Life Technologies NP0002) and transferred onto Nitrocellulose Blotting Membranes (0.45µm, Invitrogen LC2001) using 10X Tris/Glycine Buffer (Bio-Rad 1610734) and 20% Methanol (VWR TXBJ349664LBRI) in water. Gels were transferred at 30V overnight at 4°C. Membranes were blocked with 5% nonfat dry milk (Bio-Rad 1706404XTU) in Tris Buffered Saline with Tween-20 (Cell Signaling Technology, 9997) for 30-60min at room temperature. Incubation with primary antibody was performed overnight at 4°C in 5% milk in TBST. Primary antibodies and concentrations include: Vinculin (V9131

sigma, 1:1000), CHK2 (MP 05-649, 1:5000), pCHK2 (CST2661, 1:1,000), y-H2AX (CST9718, 1:1000), Actin (CST8H10D10, 1:10,000), BRCA2 (OP95, 1:500), BRCA1 (MS110, 1:1000). Membranes were washed 3 times for 15min each in TBST and incubated in Anti-mouse IgG, or Anti-rabbit IgG secondary as in 5% milk in TBST for 2hrs at room temperature. Membranes were washed an additional 3 times for 15min each in TBST and visualized after a 5min incubation with SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific, 34075), or SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific, 34078).

PDX mouse models

The PDX models used in this study were generated using adult female NSGS mice (NOD-SCID; IL2Rγ null; Tg(IL3, CSF2, KITL), The Jackson Laboratory, strain 013062), which were irradiated using an gamma cell irradiator (Best Theratronics) with 200 rads ≤24 hours before transplant. The *STAG2*-mutant AML PDX1 sample was from a male patient obtained with informed consent under Dana-Farber/Harvard Cancer Center Institutional Review Board (IRB)-approved protocol #19152. Unsorted mononuclear cells were collected from the bone marrow aspirate. Dana-Farber Cancer Institute's Rapid Heme Panel Sequencing analysis revealed the following mutation and variant allele frequencies: STAG2 ENST00000371160.1:c.2096+1G>A 66.8%, BCOR ENST00000397354.3:c.4138C>T 63.3%, BCORL1 ENST00000218147.7:c.3583A>T 48.7%, RUNX1 ENST00000300305.3:c.805+1G>T 31.6%, U2AF1 ENST00000291552.4:c.101C>T 30.7%, DNMT3A ENST00000321117.5:c.2597+1delG 29.9%, ATM ENST00000278616.4:c.8269G>A 4.0%. Unsorted mononuclear cells were thawed, resuspended in Hanks' Balanced Salt Solution (Thermo Fisher Scientific 14025076), and injected via tail vein into two 31-week old NSGS mice ("P0"). For the *STAG2*-mutant AML PDX1, P0 mice were sacrificed when they became morbid, and engraftment of human cells in the bone marrow and spleen was measured with flow cytometry (CD45-PE-Cy7 antibody, BD Biosciences). Single-cell suspensions were prepared from the bone marrow, and spleen, the cells were washed, red blood cells lysed prior to injection into another cohort of 10-week old NSGS mice using the protocol described above ("P1"). Mice from this cohort were monitored daily and

sacrificed when they became morbid. Cells were collected as described above and frozen in 90% FBS, 10% DMSO freezing media. Cells from P1 mice were thawed in PBS + 1%FBS, washed and resuspended in Hanks' Balanced Salt Solution for injection of 1.3 – 2 million cells per mouse via tail vein into 8-week old NSGS mice ("P2"). Dana-Farber Cancer Institute's Rapid Heme Panel Sequencing analysis was performed on the bone marrow cells from P2 mice to confirm the mutation status of the engrafted clone as follows: BCOR c.4240C>T (p.Q1414*) 100% VAF, BCORL1 c.3583A>T (p.K1195*) 81.5% VAF, DNMT3A c.2597+1delG (splice site) 37.7% VAF, RUNX1 c.805+1G>T (splice site) 20.3% VAF, STAG2 c.2096+1G>A (splice site) 97.9% VAF, U2AF1 c.101C>T (p.S34F) 32.9% VAF. Mice from the P2 generation were used for all experiments done in this study. The *STAG2*-mutant AML PDX2 model was used as previously published, mutation status and VAFs are as follows: STAG2 p.31013* VAF 0.92, ASXL1 p.G642fs* VAF 0.513, NRAS p.G13D VAF 0.426, RUNX1 p.320* VAF 0.48. (*6*)

The cohesin wild type AML PDX 1 model was generated from an AML patient with informed consent under Dana-Farber/Harvard Cancer Center Institutional Review Board (IRB)-approved protocol #01-026. Unsorted mononuclear cells were collected from a leukapheresis sample were frozen and sequencing analysis from Dana-Farber Cancer Institute's Rapid Heme Panel Sequencing analysis revealed the following mutation and variant allele frequencies: DNMT3A c.2645G>A (p.R882H) 49.6% VAF, NPM1 c.859_860insTCTG (p.W288Cfs*12) 43.9% VAF, PTPN11 c.1504T>G (p.S502A) 1.8% VAF, FLT3-ITD. Cells were thawed in PBS + 1%FBS, washed and resuspended in Hanks' Balanced Salt Solution for injection of 1.3 – 2 million cells per mouse via tail vein injection for all mice used in this study. The cohesin wild type AML PDX 2 model (*MLL-CALM* and *NRAS(Q61L)*-mutant (64% VAF)) was generated as previously published, using a sample obtained with informed consent under Dana-Farber/Harvard Cancer Center Institutional Review Board (IRB)-approved protocol #15384(*54*). Mobilized peripheral blood adult CD34+ cells (Fred Hutchinson Hematopoietic Cell Processing and Repository) were thawed in PBS + 1%FBS, washed and resuspended in Hanks' Balanced Salt Solution for injection of 0.5 million cells per mouse via tail vein injection.

E-7107 in vivo drug treatment and RNA extraction for RNA-seq

Three weeks after transplant, *STAG2*-mutant AML PDX1 mice were bled, and a complete blood count was performed with an Element HT5 analyzer (Heska). Mice were assigned to treatment groups (n=3 per group) based on platelet counts and treatment began 4 weeks post-injection for 5 days. For the cohesin wild type AML PDX, mice were bled 5 weeks after transplant and %hCD45 cells were measured in the circulation by flow cytometry. Mice were assigned to treatment groups (n=2 per group) based on %hCD45+ cells in the peripheral blood and treatment began 8 weeks post-injection for 3 days. E-7107 (4mg/kg in 10% Ethanol, 5% Tween80 in sterile saline solution) or an equivalent volume of Vehicle, was delivered daily via tail vein injection.

Four hours after the final injection, mice were sacrificed and single-cell suspensions were made from bone marrow, cells were washed and red blood cells were lysed. Samples were stained and sorted for viable, human CD45+ cells (BD 557748) directly into RLT lysis buffer supplemented with b-Mercaptoethanol on a BD FACSAria II cell sorter. RNA extraction was performed using a Qiagen RNeasy Mini Kit. 50ng of RNA was used as input for library preparation with the KAPA mRNA HyperPrep kit. Libraries were sequenced on the Illumina NOVASeq in PE100 mode to a depth of 100M reads per sample at the Dana-Farber Cancer Institute.

E-7107 + talazoparib in vivo drug treatment

For the *STAG2*-mutant AML PDX1, mice were bled three weeks after transplant and a complete blood count was performed with an Element HT5 analyzer (Heska). Mice were assigned to treatment groups based on platelet counts and treatment began 4 weeks post-injection with a 5-days-on, 2-days-off treatment schedule. For the *STAG2*-mutant AML PDX2, mice were bled monthly to monitor %hCD45 cells in circulation by flow cytometry and to perform a complete blood count with an Element HT5 analyzer (Heska). 6 months after transplant, mice were assigned to treatment groups based on %hCD45 cells in the

peripheral blood. For the first 5 days of treatment, mice received both a tail vein injection of E-7107 (4mg/kg in 10% Ethanol, 5% Tween80 in sterile saline solution) or Vehicle, and oral gavage of talazoparib (0.25mg/kg in 0.5% methylcellulose), or methylcellulose alone depending on the treatment group. The second week of treatment continued only with talazoparib dosing until the time of sacrifice. Mice were weighed daily, and drug was withheld if the body weight dropped by ≥15% from the starting mouse weight. Mice were monitored daily and were sacrificed when morbid or when they had lost 20% of the starting body weight. Engraftment of human cells in the bone marrow, spleen, and peripheral blood was determined with flow cytometry.

E-7107 + doxorubicin and cytarabine in vivo drug treatment

STAG2-mutant AML PDX1 mice were bled three weeks after transplant and a complete blood count was performed with an Element HT5 analyzer (Heska). Mice were assigned to treatment groups based on platelet counts and treatment began 4 weeks post-injection. Treatment days 1 – 3 mice received either E-7107 (IV, 4mg/kg in 10% ethanol, 5% Tween80 in sterile saline solution) or vehicle only control. On days 4 – 8, mice in the chemotherapy arm received 5 days of cytarabine (IP 10mg/kg in sterile saline solution), and 3 days of doxorubicin (IV, 1mg/kg in PBS) given in combination with cytarabine on days 4-6. The E-7107 arm received 5+3 days of cytarabine and doxorubicin on days 6-10 following 2 days of rest after E-7107. Mice were weighed daily and sacrificed when morbid or when they had lost 20% of the starting body weight. Engraftment of human cells in the bone marrow, spleen, and peripheral blood was determined with flow cytometry.

Flow cytometry for hCD45 detection in blood samples from PDX models

Peripheral blood samples were collected with retro-orbital bleeding in EDTA coated tubes. Red blood cells were lysed twice with Lysing Buffer (BD Biosciences 555899) and remaining cells were stained for human CD45 (BD 557748). The percentage of $hCD45^+$ cells in the blood was determined with flow cytometry.

SUPPLEMENTARY FIGURES

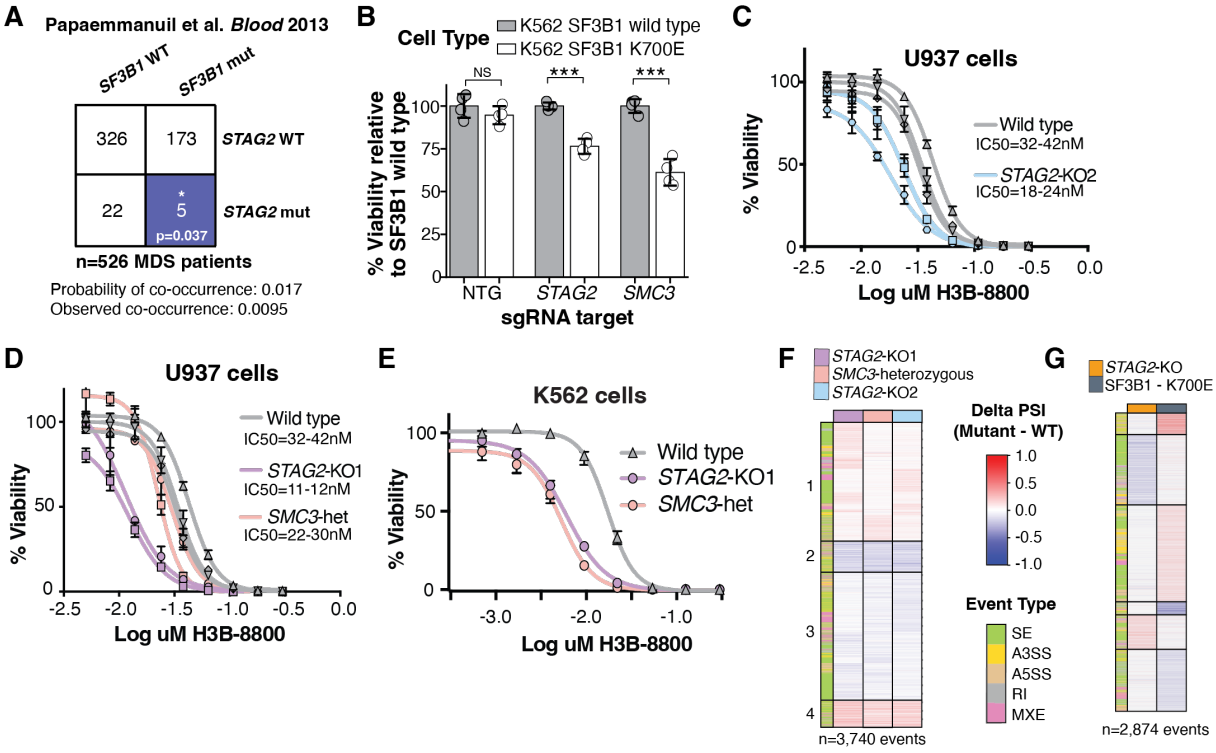

Fig. S1. Cohesin-mutant cells are sensitive to SF3B1-targeting compounds. (**A**) Co-occurrence of *SF3B1* and *STAG2* mutations in a cohort of patients with MDS (*11*). Expected and observed probably of co-occurrence are listed. Blue color indicates a significant mutually exclusive relationship between *SF3B1* and *STAG2* mutations. **P*<0.05 (Z-test). WT = wild type, mut = mutant. (**B**) Cell viability of wild type and SF3B1 K700E-mutant K562 cells following depletion of cohesin complex proteins with sgRNAs. Values are normalized to SF3B1 wild type cells for each knock-out or non-targeting guide RNA (NTG). Cell counts were taken 6 days after transduction with sgRNAs. Data shown are independent biological replicates and error bars represent +/- standard deviation. ****P*<0.001 Student's t-test. (**C** and **D**) Drug dose-response curves of WT, *STAG2*-mutant, and *SMC3*-mutant U937 clones on day 8 of treatment with H3B-8800. Error bars represent SD of three technical replicates. (**E**) Drug dose response curves of one representative example of WT and *STAG2*-KO, and *SMC3*-het K562 clones on day 12 of treatment with H3B-8800. IC50 (WT) =20nM, IC50 (*STAG2*-KO) = 6nM, IC50 (*SMC3*-het) = 5nM. Error bars represent SD of measurements of three technical replicates. (**F**) k-means clustering of delta (Δ) percent spliced in (PSI) scores for all events detected in any comparison of a cohesin-mutant genotype to wild type controls in U937 cells. Each row represents a splicing event that is significantly regulated in at least one condition (FDR < 0.05, ΔPSI > 5%), and the values plotted are ΔPSI (mutant – wild type) scores averaged from independent single-cell clones in each condition. Color bar on the left indicates the type of splicing event that was called. (**G**) k-means

clustering of ΔPSI scores for all significant splicing changes (ΔPSI > 5%, FDR < 0.05) detected in SF3B1 K700E-mutant (*26*) or *STAG2*-KO K562 cells (*28*). Heatmap values are as described in panel 1F.

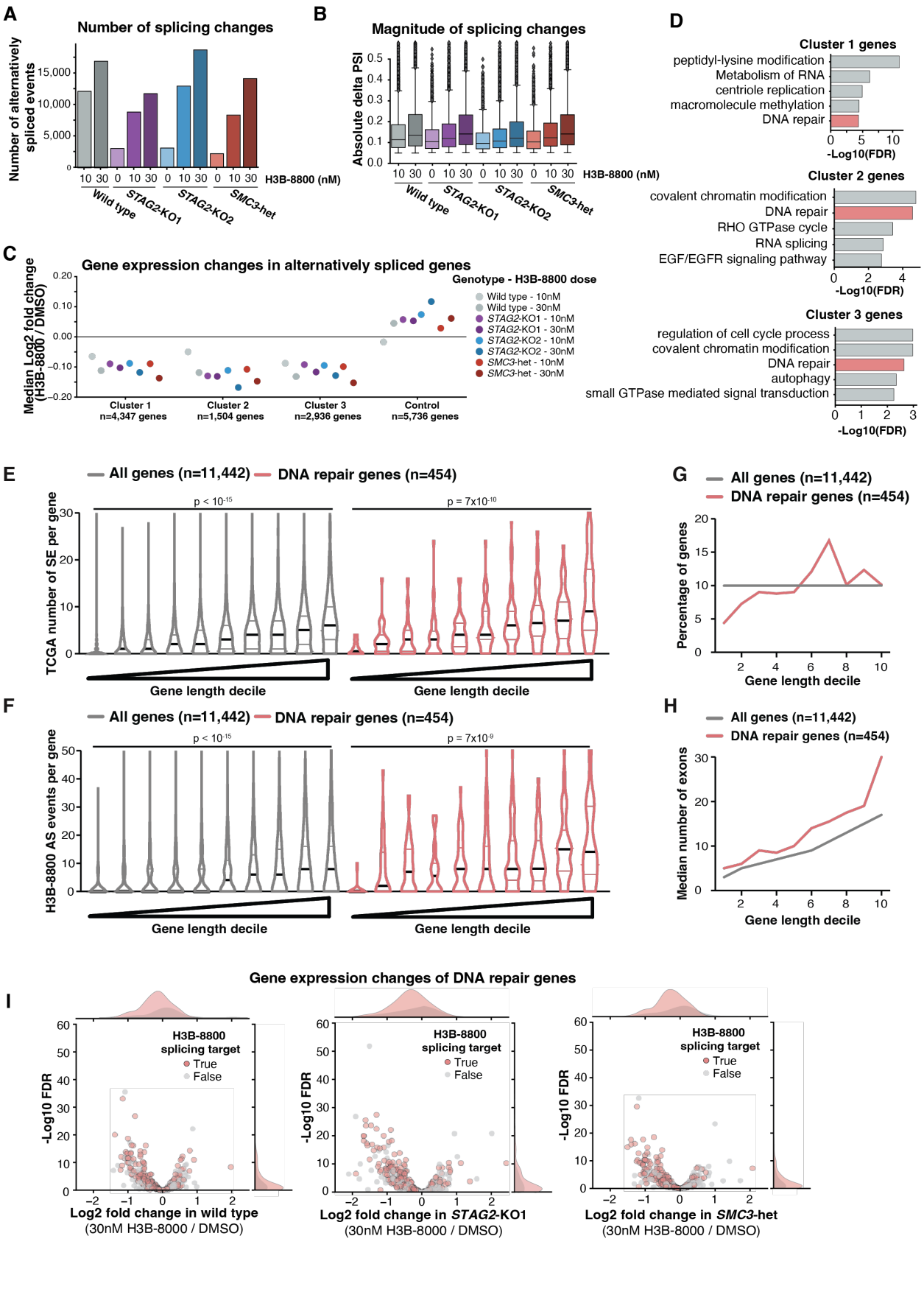
A
Number of splicing changes
Number of alternatively spliced events
15,000
10,000
5,000
0
10 30 0 10 30 0 10 30 0 10 30
H3B-8800 (nM)
Wild type
STAG2-KO1
STAG2-KO2
SMC3-het
B
Magnitude of splicing changes
Absolute delta PSI
0.5
0.4
0.3
0.2
0.1
0.0
10 30 0 10 30 0 10 30 0 10 30
H3B-8800 (nM)
Wild type
STAG2-KO1
STAG2-KO2
SMC3-het
C
Gene expression changes in alternatively spliced genes
Median Log2 fold change (H3B-8800 / DMSO)
0.20
0.10
0.00
−0.10
−0.20
Genotype - H3B-8800 dose
Wild type - 10nM
Wild type - 30nM
STAG2-KO1 - 10nM
STAG2-KO1 - 30nM
STAG2-KO2 - 10nM
STAG2-KO2 - 30nM
SMC3-het - 10nM
SMC3-het - 30nM
Cluster 1 n=4,347 genes
Cluster 2 n=1,504 genes
Cluster 3 n=2,936 genes
Control n=5,736 genes
D
Cluster 1 genes
peptidyl-lysine modification
Metabolism of RNA
centriole replication
macromolecule methylation
DNA repair
0 5 10
-Log10(FDR)
Cluster 2 genes
covalent chromatin modification
DNA repair
RHO GTPase cycle
RNA splicing
EGF/EGFR signaling pathway
0 2 4
-Log10(FDR)
Cluster 3 genes
regulation of cell cycle process
covalent chromatin modification
DNA repair
autophagy
small GTPase mediated signal transduction
0 1 2 3
-Log10(FDR)
E
All genes (n=11,442)
DNA repair genes (n=454)
p < 10-15
p = 7x10-10
TCGA number of SE per gene
30
20
10
0
Gene length decile
Gene length decile
G
All genes (n=11,442)
DNA repair genes (n=454)
Percentage of genes
20
15
10
5
0
2 4 6 8 10
Gene length decile
F
All genes (n=11,442)
DNA repair genes (n=454)
p < 10-15
p = 7x10-9
H3B-8800 AS events per gene
50
40
30
20
10
0
Gene length decile
Gene length decile
H
All genes (n=11,442)
DNA repair genes (n=454)
Median number of exons
30
20
10
0
2 4 6 8 10
Gene length decile
I
Gene expression changes of DNA repair genes
H3B-8800 splicing target
True
False
-Log10 FDR
60
50
40
30
20
10
0
−2 −1 0 1 2
Log2 fold change in wild type
(30nM H3B-8000 / DMSO)
Log2 fold change in STAG2-KO1
(30nM H3B-8000 / DMSO)
Log2 fold change in SMC3-het
(30nM H3B-8000 / DMSO)

Fig. S2. H3B-8800 treatment induces mis-splicing and downregulation of DNA damage repair genes. (**A**) Total number of significant splicing changes (FDR < 0.05, ΔPSI > 5%, 2 or 3 biological replicates included per condition) in each mutant genotype and drug treatment. Cohesin-mutant cell lines that were treated with DMSO (0nM H3B-8800) are compared to DMSO-treated wild type cells. For each cell line treated with H3B-8800, the number of events is calculated compared to the DMSO-treated condition of the same genotype. (**B**) Absolute value of ΔPSI (percent spliced in) found in all significant H3B-8800–regulated splicing changes (FDR < 0.05, ΔPSI > 5%, 2 or 3 biological replicates included per condition). Cohesin-mutant cell lines that were treated with DMSO (0nM H3B-8800) are compared to DMSO-treated wild type cells. For each cell line treated with H3B-8800, the number of events is calculated compared to the DMSO-treated condition of the same genotype. Boxes represent the median value and the interquartile range. Whiskers extend to the rest of the distribution except outlier points that are plotted as diamonds. (**C**) Median log2 fold change of genes in each cluster of H3B-8800–regulated splicing changes. A control set was generated from genes that are not alternatively spliced in H3B-8800–treated cells. Fold change in gene expression was calculated for each concentration of H3B-8800 relative to DMSO-treated cells of the same genotype (n=2 or 3 biological replicates per genotype). (**D**) Top 5 gene ontology terms ranked on significance among alternatively spliced genes in each cluster of H3B-8800–regulated splicing events (Figure 2B). Ontology enrichment was performed with Metascape (*55*). (**E and F**) Violin plots comparing the number of skipped exons in TCGA. Pan-cancer TCGA exon-skipping events were collated from ExonSkipDB (*29*). (**E**) or the number of alternatively spliced (AS) events detected following H3B-8800 treatment (**F**) per gene across gene length deciles, comparing DNA repair genes (n=454) to all expressed protein-coding genes (n=11,442). Horizontal lines in violin plots depict the median (black) and 1st and third quartiles (gray or red). p-values were calculated using Welch's t-test. (**G**) The percentage of DNA repair and all expressed genes are compared across gene length deciles. (**H**) The median number of exons for DNA repair genes and all expressed protein-coding genes are compared across gene length deciles. (**I**) Volcano plot of gene expression changes in DNA repair genes in wild type, *STAG2*-KO1, and *SMC3*-heterozygous cells treated with 30nM H3B-8800, relative to DMSO-treated controls. DNA repair genes that contain H3B-8800–regulated splicing changes are highlighted in red.

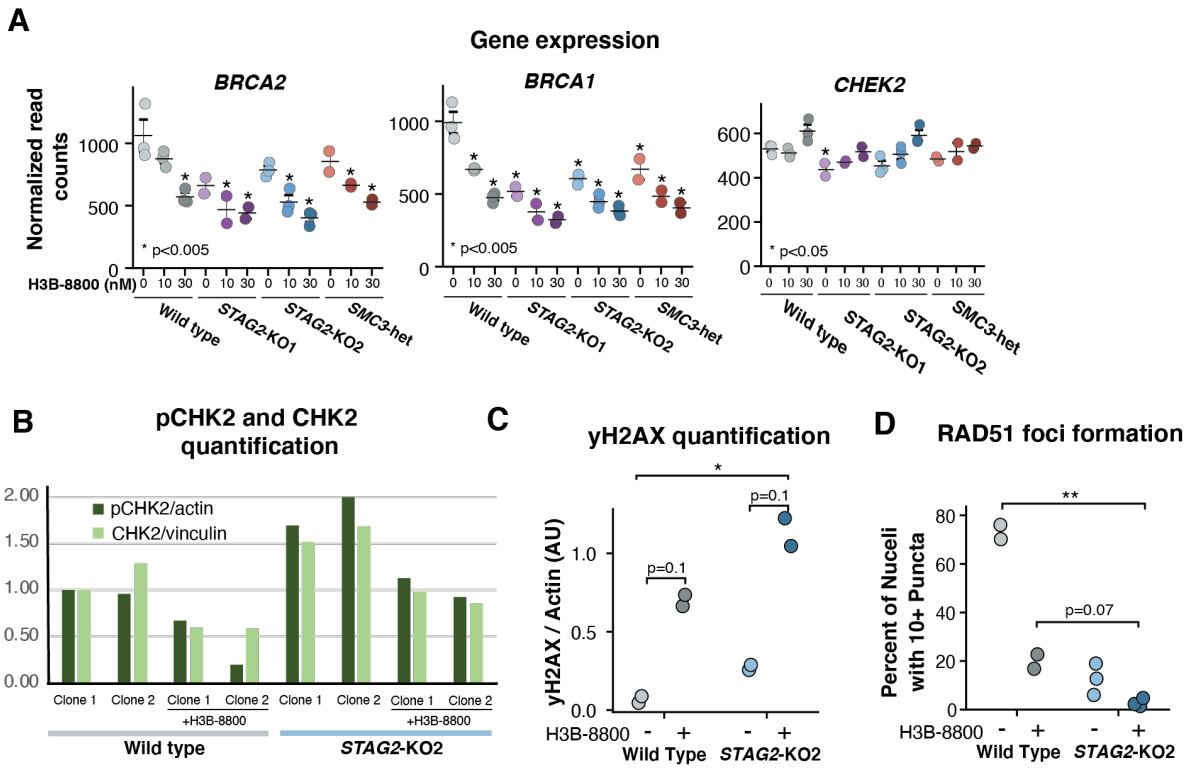

Fig. S3. Mis-splicing of DNA repair genes alters protein function and results in accumulation of DNA damage. (**A**) DESeq2 normalized read counts measuring gene expression of *BRCA2, BRCA1,* and *CHEK2*. Error bars represent mean +/- SEM of biological replicates (n=2 or 3 per genotype). **P*<0.005 relative to DMSO-treated wild type cells (FDR-corrected Wald test). (**B**) Quantification of Western blot band intensity normalized to actin and vinculin in panel from Fig.3D using ImageJ/Fiji. (**C**) Quantification of Western blot band intensity normalized to actin in panel from Fig.3E performed using Adobe Photoshop. *P*=0.08, Kruskal-Wallis test. (**D**) Quantification of percent of cell nuclei with 10 or more RAD51^{+} puncta as determined with immunofluorescence in U937 cells treated with 50nM H3B-8800 or DMSO control for 24 hours. (See example images in Figure S3B). Minimum of 50 cell nuclei were imaged per datapoint. Datapoints represent biological replicates. **P*<0.05, Kruskal-Wallis test. ***P*<0.01 in post-hoc analysis using the Dunn's test.

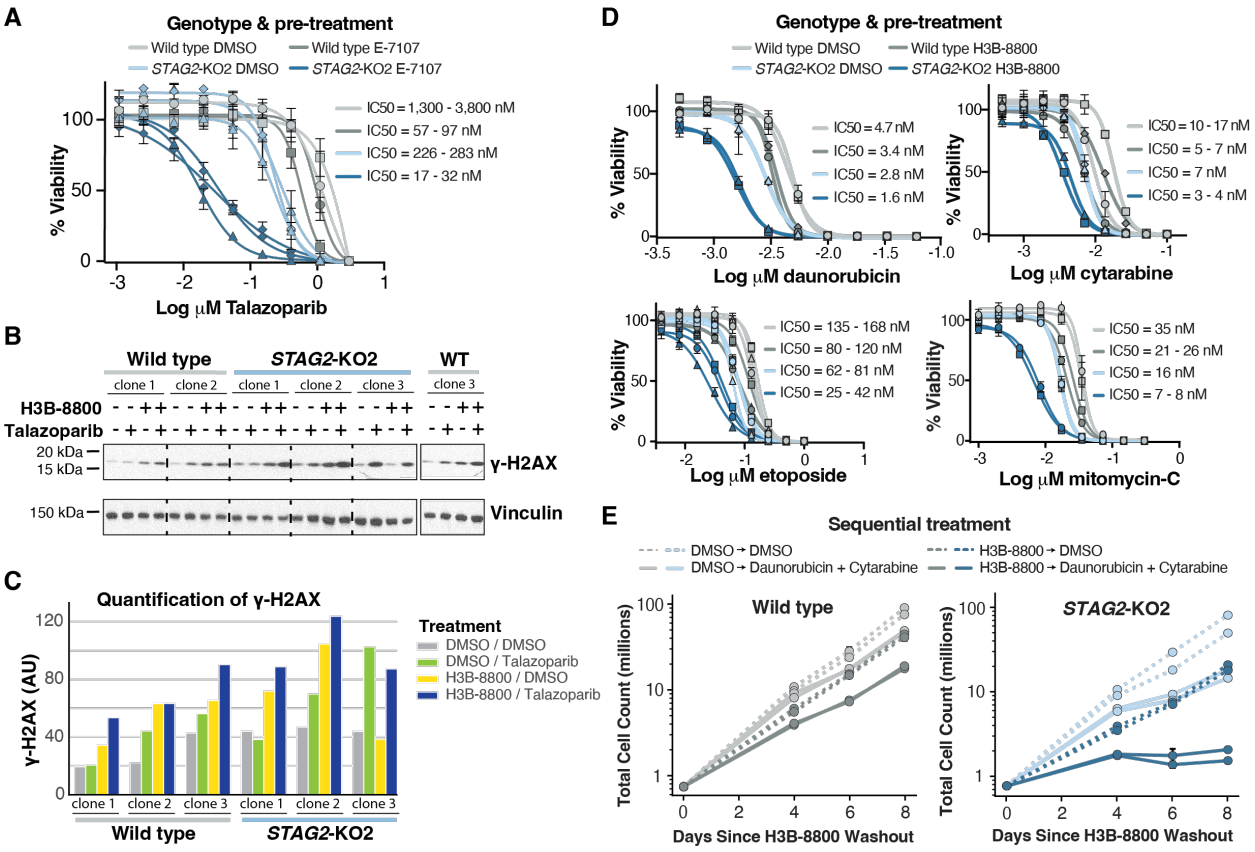

Fig. S4. Splicing modulation sensitizes cohesin-mutant AML cell lines to killing by talazoparib and chemotherapy. (**A**) Drug dose-response curves of wild type and *STAG2*-KO2 cells pre-treated for 3 days with 1.14 nM E-7107 or DMSO, followed by 8 days of treatment with talazoparib. Error bars represent SD of technical triplicate measurements for each biological replicate sample (n=2 per genotype and condition). (**B**) Western blot analysis of γH2AX in individual U937 single-cell clones after sequential treatment with H3B-8800 or DMSO and talazoparib or DMSO. Cells were treated with either H3B-8800 or DMSO for 3 days (top row), when the drug was washed out and replaced with either talazoparib or DMSO for 4 days (bottom row). Cells were collected for Western blot on day 7. (**C**) Quantification of γH2AX Western blot (panel B) summarizing independent single-cell clones of each genotype. (**D**) Drug dose-response curves of wild type and *STAG2*-KO2 cells pre-treated with 50nM H3B-8800 for 3 days followed by 8 days of treatment with daunorubicin, cytarabine, etoposide, or mitomycin-C. Error bars represent SD of measurements in technical triplicates measurements for each biological replicate sample (n=2 per genotype

and condition). (**E**) Growth curves depicting total number of wild type or *STAG2*-KO2 cells pre-treated with DMSO or H3B-8800 (50nM) into 8 days of DMSO or daunorubicin (5nM) + cytarabine (10nM) combination treatment. Error bars represent SD of technical duplicate measurements for each biological replicate sample (n=2 per genotype and condition).

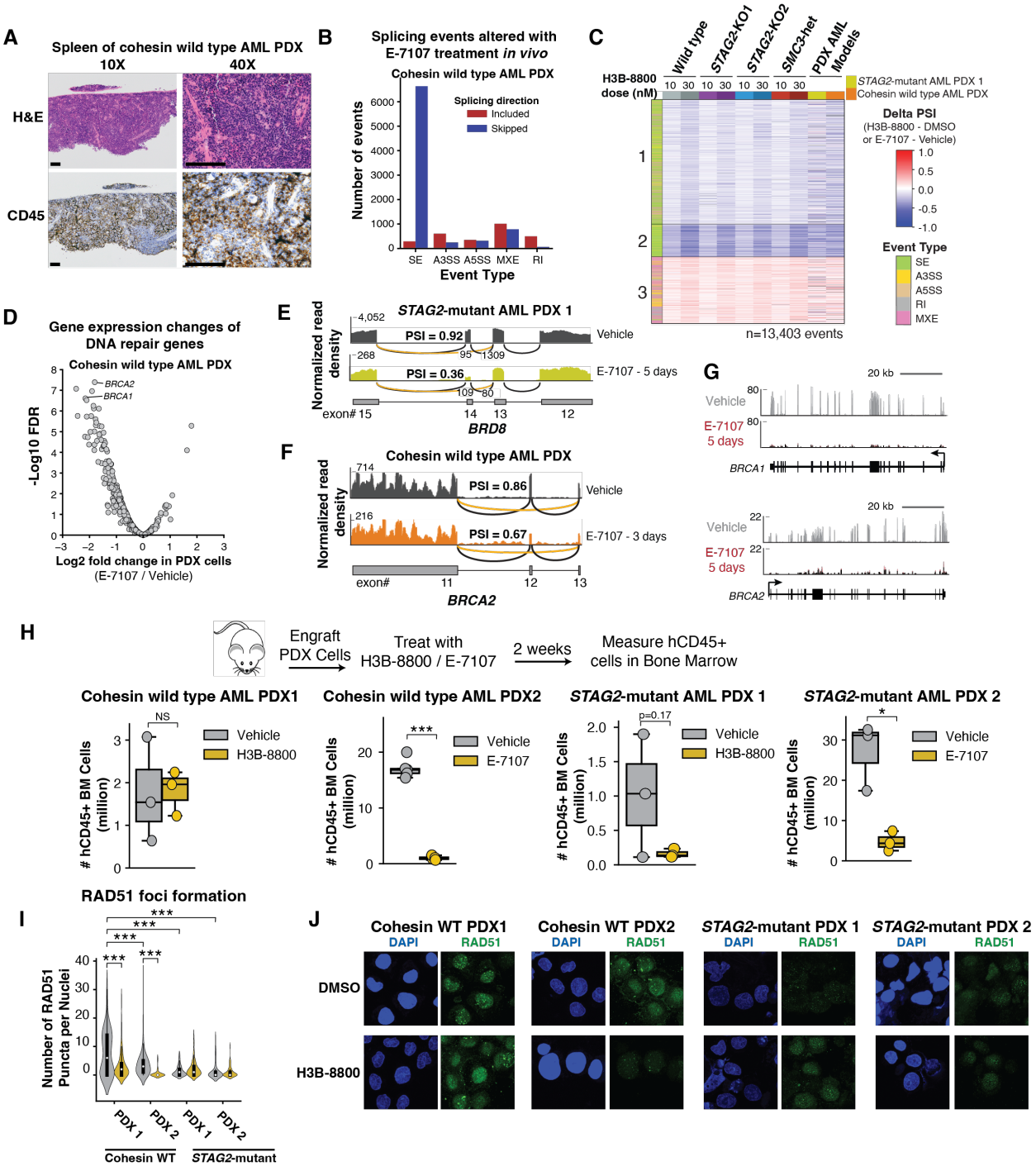

Fig. S5. Splicing modulation targets DNA repair genes and cohesin-mutant cells in vivo.

(**A**) Morphologic evaluation of spleen of cohesin wild type AML patient derived xenograft (PDX) shows infiltration with human leukemia blasts. Images were stained using H&E and hCD45-targeting antibody and imaged at 10x and 40x (scale bar: 0.125mm) original magnification. (**B**) Total number and directionality of significant splicing alterations (FDR < 0.05, ΔPSI > 5%) differentially called in cohesin

wild-type human AML cells isolated from bone marrow of NSGS mice treated with E-7107 compared to vehicle for 3 days in vivo. Splicing events are categorized by event type and direction of regulation in E-7107 versus vehicle treated mice. (SE = skipped exon, A3SS = alternative 3' splice site, A5SS = alternative 5' splice site, MXE = mutually exclusive exon, RI = retained intron). n=2 mice per condition. (**C**) Heatmap of ΔPSI scores for H3B-8800–regulated exons called from U937 cells (Figure 3A) including two human AML PDX models that were sequenced. ΔPSI scores from PDX models compare E-7107 to vehicle treated control mice for 3 days (cohesin wild type AML) or 5 days (*STAG2*-mutant AML). n=3 mice per condition for *STAG2*-mutant AML and n=2 mice for cohesin wild type AML. Color bar on the left indicates the type of splicing event that was called. (**D**) Volcano plot depicting differential gene expression of DNA repair genes in cohesin wild type AML PDX cells isolated from the bone marrow of NSGS mice treated with E-7107 versus vehicle for 3 days in vivo. n=2 mice per condition. (**E**) RNA-seq normalized read density and splice junction track of exon skipping in *BRD8* exon 14 from one representative replicate of vehicle treated *STAG2*-mutant PDX AML cells and the combined coverage of 3 replicates of E-7107 treated cells. Average PSI scores of exon14 are shown. Average number of reads supporting exon skipping (orange line) and exon inclusion (black line) are reported. PSI (percent spliced in). (**F**) RNA-seq normalized read density and splice junction track of exon skipping in *BRCA2* exon 14 from one representative replicate of vehicle and E-7107-treated cohesin wild type AML PDX model. Average PSI scores of exon12 are shown. Constitutive junctions are shown with black lines, exon skipping junctions are shown in orange. PSI (percent spliced in). (**G**) The integrative genomics viewer (IGV) browser tracks of RNA-seq coverage on *BRCA1* and *BRCA2* genes from *STAG2*-mutant human AML1 PDX cells isolated from the bone marrow of NSGS mice treated with E-7107 or vehicle for 5 days in vivo. 3 mice per condition are merged for visualization. (**H**) Total number of human cells per mouse bone marrow following two weeks of treatment with H3B-8800 (8mg/kg, oral gavage), E-7107 (4mg/kg tail vein injection), or vehicle in four different AML PDX models. Animals were treated for 5 days, followed by 2 days of rest, and an additional 5 days of treatment. Mice were sacrificed at the end of the second week and bone marrow was collected. P-values are calculated with student's t-test (**P*<0.05, ****P*<0.0005). (**I**) Quantification of percent of cell nuclei with 10 or more $RAD51^{+}$ puncta as determined with immunofluorescence in PDX cells treated with 50nM H3B-8800 or DMSO control for 24 hours. (see example images in panel J). Minimum of 50 cell nuclei were imaged per datapoint. (**J**) Representative images from nuclei (DAPI, blue) and RAD51 (green) stained PDX cells treated with 50nM H3B-8800 or DMSO control for 24 hours.

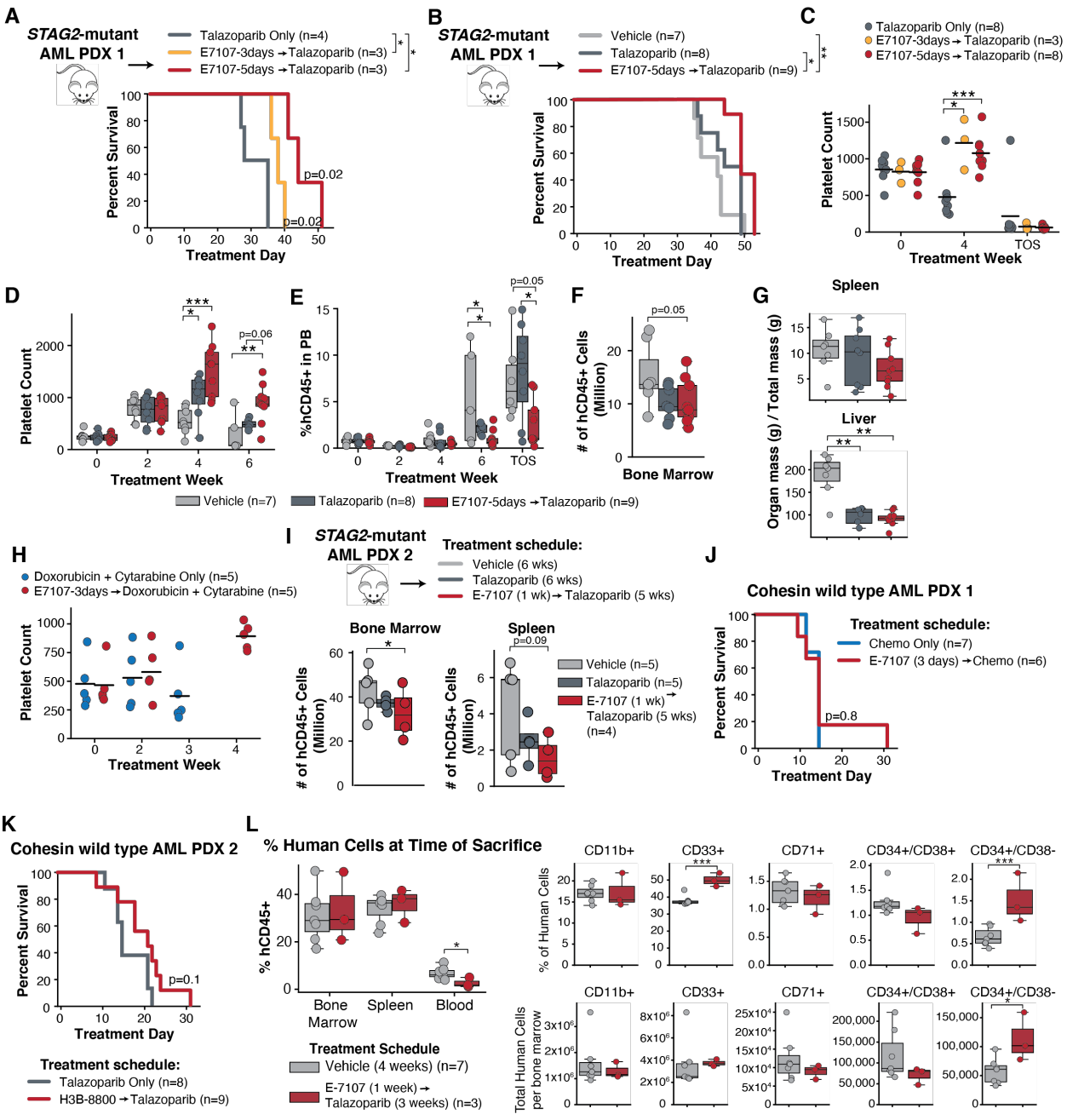

Fig. S6. Low-dose splicing modulation combined with talazoparib targets cohesin-mutant AML in vivo. (**A**) Survival analysis from first in vivo experiment with E-7107 and talazoparib for *STAG2*-mutant AML PDX 1 model (other mutations include *BCOR/RUNX1/U2AF1/DNMT3A*). Treatment of mice assigned to three treatment arms was initiated 3 weeks after bone marrow transplantation: talazoparib only (n=4), E-7107 for 3 days followed by talazoparib (n=3), or E-7107 for 5 days followed by talazoparib (n=3). P-values shown are compared to the talazoparib only arm (log-rank test). (**B**) Survival analysis from an independent in vivo experiment with E-7107 and talazoparib for *STAG2*-mutant AML PDX 1 model. Treatment of mice was initiated 3 weeks after transplant for 3 arms: vehicle (n=7), talazoparib only (n=8),

and E-7107 for 5 days followed by talazoparib (n=9). (**C**) Platelet count from peripheral blood of *STAG2*-Mutant AML1 PDX mice before treatment (T0), 4 weeks after start of treatment (4wk), and at the time of sacrifice (TOS). Students t-test comparing each treatment group to talazoparib-only group for each timepoint. (**D**) Platelet count from peripheral blood of *STAG2*-mutant AML PDX 1 mice before treatment (Week 0) and every other week while on treatment. p=0.001 (4wk), p=0.02 (6wk) (Kruskal-Wallis), **P*<0.05, ***P*<0.01, ****P*<0.001 in post-hoc analysis using the Dunn's test. (**E**) Percent of human $CD45^+$ cells measured in peripheral blood (PB) in *STAG2*-mutant AML PDX1 mice before treatment (Week 0) and every other week on treatment. TOS = time of sacrifice corresponding to survival curve shown in panel C. *P*=0.057 (6wk), *P*=0.03 (TOS) (Kruskal-Wallis), **P*<0.05 in post-hoc analysis using the Dunn's test. (**F**) Total human $CD45^+$ cells at the time of sacrifice from the bone marrow of mice from *STAG2*-mutant AML PDX1. These data correspond to the mice presented in panel C. *P*=0.1 (Kruskal-Wallis) (**G**) Spleen (top) and liver (bottom) weights normalized to mouse weight at the time of sacrifice from *STAG2*-mutant AML PDX1. These data correspond to the mice presented in panel C. *P*=0.003 (liver) (Kruskal-Wallis), ***P*<0.01 in post-hoc analysis using the Dunn's test. (**H**) Platelet count from peripheral blood of *STAG2*-mutant AML PDX 1 mice before treatment (Week 0) and at each week while on treatment. These mice correspond to the cohort shown in Figure 5F. (**I**) Total human cells at end of experiment per organ indicated in *STAG2*-mutant AML PDX 2. Mice were assigned to three treatment groups: vehicle (n=5), talazoparib only (n=5), or E-7107 followed by talazoparib (n=4). All animals were sacrificed following 6 weeks of treatment as indicated. *P*=0.1 (bone marrow), *P*=0.2 (liver) (Kruskal-Wallis), **P*<0.05 in post-hoc analysis using the Dunn's test. (**J**) Survival analysis from cohesin wild type AML PDX 1 (other mutations include *FLT3-ITD/DNMT3A/NPM1/PTPN11*). Treatment began 3 weeks after bone marrow transplantation: combination chemotherapy (5+3 doxorubicin and cytarabine) (n=7), or E-7107 for 3 days, followed by 2 days of rest, and combination chemotherapy (5+3 doxorubicin and cytarabine) (n=6). *P* value was calculated with a log-rank test. (**K**) Survival analysis from cohesin wild type AML PDX 2 (*MLL-CALM* and *NRAS(Q61L)*). Treatment began 1 week after bone marrow transplantation: talazoparib only (n=8), or H3B-8800 for 5 days followed by talazoparib (n=9). *P*-value was calculated with a log-rank test. (**L**) Human cell analysis at the time of sacrifice from mice transplanted with healthy $CD34^+$ cells. Treatment began 6 weeks after bone marrow transplantation: Vehicle only (n=7), or E-7107 for 5 days, followed by talazoparib for 3 weeks (n=3). The percentage of human cells (CD34%) was measured in bone marrow, spleen, and peripheral blood. Bone marrow cells were stained with CD11B, CD33, CD71, CD34, and CD38 to quantify different cell populations. *P*-value was calculated with student's t-test (**P*<0.05, ****P*<0.0005).

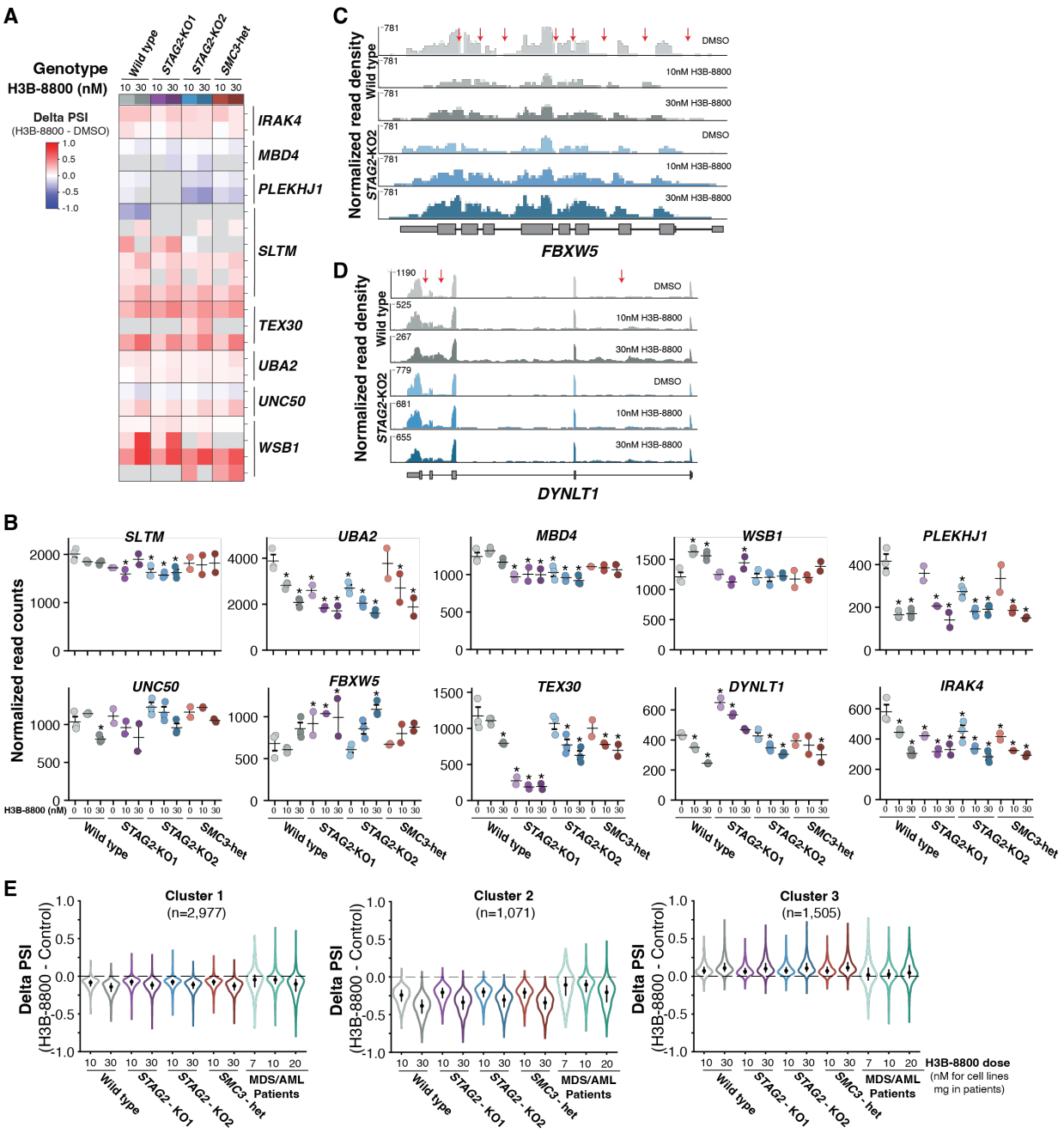

Fig. S7. Splicing changes and downregulation of DNA repair genes are conserved in MDS/AML patients. (A) ΔPSI of all H3B-8800–regulated splicing events detected in genes that were used as biomarkers of on-target activity in the Phase 1 clinical trial (*25*). ΔPSI is the average difference of 2 or 3 biological replicates per drug concentration relative to DMSO-treated controls of the same genotype. Events that were detected in at least one comparison are included in the heatmap. Grey boxes indicate the event was not detected in the represented sample. (**B**) DESeq2 normalized read counts of the genes used as biomarkers of on-target activity in the Phase 1 clinical trial of H3B-8800 that are expressed in U937 cells.

Error bars represent mean ± SEM of biological replicates (n=2 or 3 per genotype). *FDR < 0.05 compared to DMSO-treated wild type cells (FDR-corrected Wald test). (**C**) RNA-seq normalized read density from one representative example gene (*FBXW5*) in H3B-8800–treated wild type and *STAG2*-KO2 cells. Red arrows indicate regions with a dose-dependent effect on intron retention. (**D**) RNA-seq normalized read density from one representative example gene (*DYNLT1*) in H3B-8800–treated wild type and *STAG2*-KO2 cells. Red arrows indicate regions with a dose-dependent effect on intron retention. (**E**) Violin plots depicting the ΔPSI scores for splicing events in each cluster from Figure 5C. Dot represents the median, and bars extend from the first to third quartile range.

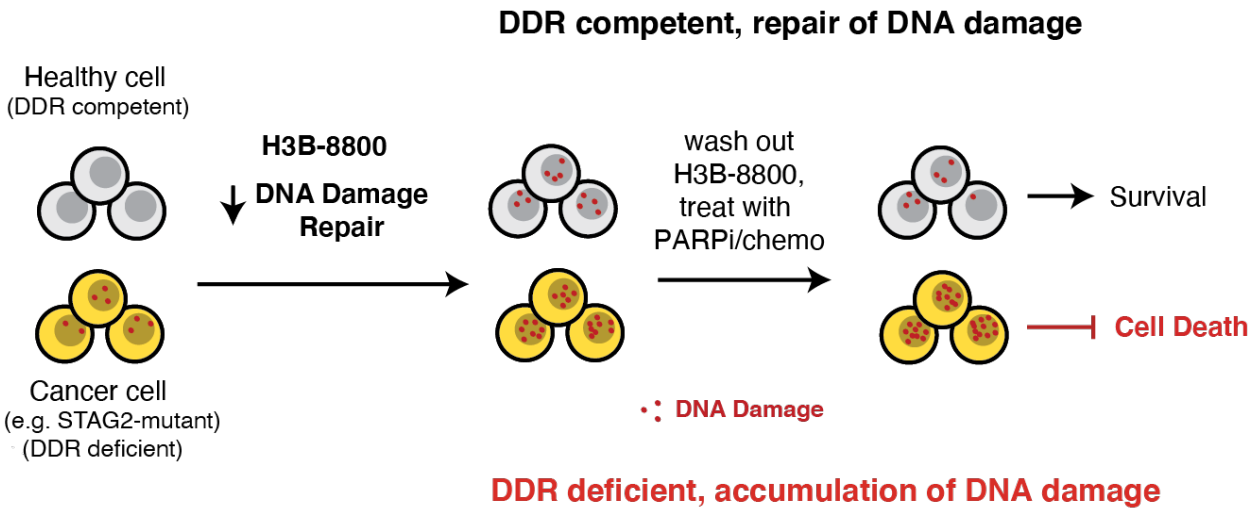

Fig. S8. Proposed model. Cartoon depicting the mechanism of combining splicing-modulation and PARP inhibition (PARPi) or chemotherapy on targeting DNA damage repair (DDR) deficient cancer cells. SF3B1 splicing modulators induce selective cell killing of DDR-deficient cancer cells (such as *STAG2*-mutant AML cells) by impairing proper splicing and expression of DDR genes, which leads to DNA damage accumulation and subsequent sensitization to treatment with PARPi or chemotherapy.

Supplementary Tables

Table S1. **Gene Ontology of H3B-8800 Target Genes.** Metascape summary for enrichment of gene ontology terms among genes that are mis-spliced in each cluster (Figure 2B) in cells treated with H3B-8800. The top 20 enriched categories are reported for each cluster of genes, ranked on significance value. In cases where more than 500 genes were in a cluster, the top 500 ranked on magnitude of ΔPSI were used to calculate enrichment of gene ontology categories. (Excel file)

Table S2. **Patient Sample Information.** Information on disease type and mutation status of patients for whom sequencing was performed on peripheral blood samples pre- and post- H3B-8800 treatment.

Patient	Sex	Disease Type	Karyotype & Mutation	H3B-8800 drug dose
#1	Male	MDS EB2	46XY, IDH1 132C, SRSF2 P95H	7mg po daily
#2	Male	AML MRC transformed from RAEB-2	47 XY, +19[7], SRSF2 P95H, DNMT3A R882H, BCOR R1661*, SETD1B E1146delinsV, NRAS G12A	10mg po daily
#3	Male	AML transformed from CMML-2	46 XY, +8 [20], ASXL1 L823*, RUNX1 W108fs*33, SRSF2 P95L	20mg po daily

Data file S1. Raw data for all experiments where $n<20$.