Supporting Information

Profiling the Landscape of Drug Resistance Mutations in Neosubstrates to Molecular Glue Degraders

Pallavi M. Gosavi,[1,2]‡ Kevin C. Ngan,[1,2]‡ Megan J. R. Yeo,[1,2] Cindy Su,[1,2] Jiaming Li, [1,2] Nicholas Z. Lue,[1,2] Samuel M. Hoenig, [1,2] Brian B. Liau[1,2,]*

[1]Department of Chemistry and Chemical Biology, Cambridge, MA 02138, USA
[2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

*E-mail: liau@chemistry.harvard.edu

Index:

MATERIALS AND METHODS

Cell culture and lentiviral production. MOLM-13 cells were a gift from M.D. Shair (Harvard University). HEK293T cells were a gift from B.E. Bernstein (Massachusetts General Hospital). 293FT and K562 cells were obtained from Thermo Fisher Scientific and ATCC, respectively. All cell lines were cultured in a humidified 5% CO_2 incubator at 37 °C and routinely tested for mycoplasma (Sigma-Aldrich). All media were supplemented with 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin (Gibco) and FBS (Peak Serum). MOLM-13 and K562 cells were cultured in RPMI-1640 (Life Technologies) supplemented with 10% FBS. HEK293T and 293FT cells were cultured in DMEM (Gibco) supplemented with 10% FBS, 6 mM L-glutamine (Gibco), 0.1 mM MEM non-essential amino acids (Gibco), and 1 mM sodium pyruvate. For lentivirus production, the transfer plasmids of interest were co-transfected with *GAG/POL* and *VSVG* plasmids into 293FT cells using FuGENE HD (Promega) according to manufacturer's protocol. Media was exchanged after 6-8 h and the viral supernatant was collected 48 h after transfection and sterile-filtered (0.45 μm). MOLM-13 and K562 cells were transduced by spinfection at 1,800 × *g* for 1.5 h at 37 °C with 5 μg ml^{-1} and 1 μg ml^{-1} polybrene (Santa Cruz Biotechnology), respectively. 48 h post-transduction, cells were selected with 1 μg ml^{-1} and 2 μg ml^{-1} puromycin (Thermo Fisher Scientific) respectively for 10 days.

Chemical reagents. Compounds were stored at −80 °C in 100% DMSO. The vehicle condition represents 0.1% (v/v) DMSO treatment. Indisulam and E7820 were purchased from Sigma-Aldrich (≥98% purity by HPLC) and Cayman Chemical (≥98% purity by HPLC), respectively. MLN-4924 was purchased from MedChemExpress (≥98% purity by NMR). CC-885 was purchased from MedKoo Biosciences (98% purity by HPLC) and ZXH-1-161 was a gift from N. S. Gray (Stanford University).

Cell growth assays. MOLM-13 cells were seeded at a cell density of 3×10^5 cells/mL in a 96-well plate in triplicate with drug or vehicle treatments. Cell viability was assessed by flow cytometry after 3 days of drug or vehicle treatment using Helix NP NIR viability dye (BioLegend). Dose-response curves were fitted through interpolation using GraphPad Prism v.7 nonlinear regression fit ([inhibitor] versus normalized response–variable slope).

Pooled sgRNA library cloning and CRISPR-suppressor scanning experiments. The *RBM39* and *GSPT1* tiling libraries contained all sgRNAs targeting the coding sequence (*NP_909122.1* and *NP_002085.3*) with MIT specificity scores above 25 (CRISPOR) and were ordered as an oligo pool from Twist Biosciences. These sgRNA sequences are listed in the Supplementary Table 1-2. The oligo pool was cloned into pLentiCRISPR.v2 as previously described.[40] pLentiCRISPR.v2 was a gift from F. Zhang (Addgene 52961, Broad Institute). Lentiviral particles carrying the resulting sgRNA library were generated as described above and titered according to the published procedure.[40] Cells ($13\text{-}15 \times 10^7$) were transduced at a multiplicity of infection < 0.3 and subsequently selected with 1 μg ml^{-1} puromycin for 10 days. The cells were then split into pools and treated with the drug(s) or vehicle in triplicate. The cells were passaged every 3-4 days and seeded at a density of $0.3\text{-}0.4 \times 10^6$ cells/mL. The cells transduced with the *RBM39* sgRNA library were treated with 1 μM of sulfonamide analogs (indisulam or E7820) while the cells transduced with the *GSPT1* sgRNA

library were treated with increasing concentrations of drugs (CC-885 and ZXH-1-161) according to the growth curve to produce a consistent impact on cell proliferation. Briefly, for CC-885, the cells were initially treated at 1 nM for a week, with doses escalating each subsequent week to 2.5 nM, 5 nM, and 10 nM over 4 weeks. For ZXH-1-161, the cells were initially treated at 85 nM for a week, with doses escalating each subsequent week to 500 nM, 750 nM, and 1000 nM over 4 weeks. Genomic DNA was isolated from the drug- and vehicle-treated cells at the specified time points using the QIAamp DNA Blood Mini Kit (Qiagen). The sgRNA composition of the population from each replicate was PCR amplified and barcoded followed by next-generation sequencing on a MiSeq (Illumina) using 150-cycle, single-end reads as previously described.[41]

CRISPR-suppressor scanning data analysis. All data processing and analysis were performed using Python v.3.8.3 (www.python.org). Raw sequencing data were processed as previously described.[32,42] In brief, reads were counted by identifying the 20-nt sequence downstream of the 'CGAAACACCG' prefix and mapped against a reference file containing all library sgRNA sequences with 0 mismatch allowance. sgRNAs with 0 reads in the plasmid library were excluded from the analysis. Read counts were then converted to reads per million, $\log_2$-transformed after adding a pseudocount of 1 to each sgRNA, and then normalized by subtracting the $\log_2$-transformed sgRNA counts in the plasmid library. sgRNA enrichment scores were calculated by averaging across replicates for each condition and normalized by subtracting the mean enrichment scores of the negative control sgRNAs. This value as calculated in the vehicle-treated conditions is referred to as the "fitness score." sgRNA "resistance scores" were calculated using sgRNA enrichment scores in the drug-treated conditions. sgRNAs were classified as "enriched" if their resistance scores were greater than the mean + 2 standard deviations of the negative control sgRNAs. sgRNAs were mapped to their respective protein amino acid positions using the genomic coordinates of the predicted cut site in the *RBM39* (*NP_909122.1*) and *GSPT1* (*NP_002085.3*) coding sequences. sgRNAs were assigned to a single amino acid if the cut site fell within a codon or assigned to the two flanking amino acids if the cut site fell between codons.

Individual sgRNA validation experiments and genotyping data analysis. Raw sequencing data were processed and aligned to *RBM39* and *GSPT1* using CRISPResso2 (v.2.0.40) to identify genomic variants and quantify allele frequencies.[43] An in-house Python pipeline was used to classify and characterize variants at the protein level. Variants were classified as 'in-frame' if the indel size was a multiple of three and did not span an intron-exon junction. In-frame variants were then globally re-aligned to the reference coding sequence at the nucleotide level with a custom codon-aware implementation of the Needleman-Wunsch algorithm using the 'PairwiseAligner' module of Biopython (v.1.7.8), trimmed, and translated into their corresponding protein variants.[44] Variants were classified as 'loss-of-function' if the indel size was not a multiple of three (i.e., frameshift), led to a premature stop codon (i.e., nonsense), or disrupted canonical splice site positions (the 2 nt immediately flanking each exon). Editing outcome predictions for individual sgRNAs were obtained using the inDelphi web server (https://indelphi.giffordlab.mit.edu) in single mode with K562 as the cell-type.[45]

Plasmids. Open reading frames (ORFs) for RBM39 (wt and variants), GSPT1 (wt and variants), CRBN, and DCAF15 were cloned into pcDNA3.1 vectors using Gibson cloning (New England Biolabs). HA and FLAG tags were introduced through the primers used for Gibson cloning. RBM39 and GSPT1 proteins with N-terminal HiBiT tags were cloned into an expression vector with the herpes simplex virus thymidine kinase (TK) promoter using Gibson cloning. RBM39 wt and variants were cloned into the Artichoke reporter plasmid (a gift from B. Ebert, Addgene 73320) using Golden Gate cloning.

Lytic HiBiT detection assay. HEK293T cells (0.8×10^6) were plated in a 6-well plate. After 24 h, cells were transfected with 50 ng of plasmids encoding the HiBiT-tagged proteins using FuGENE HD (Promega). Cells were trypsinized (Gibco) after 24 h and 20,000 cells were plated per well in triplicate in white, opaque 96-well plates (Corning). Cells were allowed to attach for 24 h and then treated with various concentrations of compounds or 0.1% vehicle for 6 h (CC-885) or 24 h (E7820). An equal volume of Nano-Glo HiBiT reagent (Promega) containing the lytic buffer, substrate and the LgBiT protein was then added according to the manufacturer's protocol. The plate was incubated at room temperature for 10 min with shaking at 350 rpm before measuring end point luminescence on a SpectraMax i3x microplate reader (Molecular Devices).

Fluorescent degradation reporter assay. Lentiviral particles carrying the respective constructs in the Artichoke EGFP-IRES-mCherry reporter vector were produced and used to transduce MOLM-13 and K562 cells as described above. 48 h after transduction, cells were selected with appropriate puromycin concentration for 3-5 days. The selected cells were then treated with various concentrations of E7820 or 0.1% vehicle for 24 h. EGFP and mCherry fluorescence were measured on a NovoCyte 3000RYB flow cytometer (Agilent) after drug or vehicle treatment. The geometric mean of the ratio of EGFP to mCherry fluorescence was calculated for each sample using the NovoExpress software (v. 1.5.0, Agilent). The ratios for the individual drug-treated samples were normalized to the ratios of the vehicle-treated samples.

Immunoblotting. MOLM-13 cells were plated in triplicate at 30,000 cells per well in a 96-well plate. After treatment with various concentrations of E7820 or 0.1% vehicle for 24 h, cells were harvested and washed with PBS (Corning) once. Cells were then lysed in RIPA buffer (Boston BioProducts) supplemented with 1× Halt Protease Inhibitor (Thermo Fisher Scientific). Total protein concentration in the clarified lysates was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific). Samples were prepared for SDS-PAGE followed by immunoblot analysis according to standard procedures. For immunoblotting, anti-Caper (Bethyl laboratories, A300-291A, 1:15,000), anti-HA (Cell Signaling Technology, 3724S, 1:1,000), anti-FLAG (Cell Signaling Technology, 2368S, 1:1000), anti-CRBN (Novus Biologicals, NBP1-91810, 1:500), anti-CRBN (Cell Signaling Technology, 71810, 1:1,000), anti-GAPDH (Santa Cruz Biotechnology, sc-47724), and anti-β-actin antibody (Sigma-Aldrich, A1978, 1:5,000) were used.

Co-immunoprecipitation. HEK293T (3×10^5) cells were plated into a 6-well plate and transfected 24 h post-plating with 400 ng pcDNA3.1-HA-RBM39 and 800 ng pcDNA3.1-DCAF15 expression

vectors using FuGENE HD (Promega). 48 h post-transfection, cells were pre-treated with 1 μM MLN-4924 for 2 h. After 2 h, cells were treated with either vehicle or 1 μM E7820 for 4 h. Cells were then harvested and washed with PBS. Cells were lysed in lysis buffer (50 mM NaCl, 50 mM NaH$_2$PO$_4$, 50 mM sodium citrate, 20 mM HEPES pH 7.4, 1% NP-40, 5% glycerol) supplemented with vehicle or E7820 (1 μM) and the protein concentration of the cell lysate was quantified as described above. Cell lysates containing 500 μg of protein were then incubated with Protein G Dynabeads (Thermo Fisher Scientific) and 2 μg anti-FLAG M2 antibody (Sigma-Aldrich) overnight with rotation at 4 °C followed by washing with the lysis buffer three times. Samples were then prepared for SDS-PAGE and analyzed by immunoblotting as described above. For co-immunoprecipitation of CRBN with GSPT1 variants, HEK293T (3×10^6) cells were seeded in a 10 cm plate a day prior to transfection. The cells were transiently transfected with plasmids expressing GSPT1-HA wild type and variants (6 μg) using lipofectamine 3000 according to manufacturer's protocol (Thermo Fisher Scientific). After 48 h, the cells were treated with 1 μM MLN-4924 for 3 h followed by treatment with either vehicle or 10 μM CC-885 for 2h. Cells were then harvested, washed with PBS, snap frozen and stored at –80 °C until further use. Cells were then lysed with 500 μL lysis buffer (50 mM Tris pH 7.5, 50 mM NaCl, 0.5% NP-40, 10% glycerol, 1X Halt protease inhibitor cocktail (Thermo Fisher Scientific), 1X EDTA and 1X PhosphoSTOP (Roche)) supplemented with 1 μM MLN-4924 and vehicle or 10 μM CC-885 and the protein concentration was quantified as described above. Cell lysates containing 1 mg of protein were incubated with anti-HA antibody (1:100) for 1 h at 4 °C followed by incubation with Protein G Dynabeads (Thermo Fisher Scientific) for 1 h at 4 °C. The beads were then washed with lysis buffer supplemented with with 1 μM MLN-4924 and vehicle or 10 μM CC-885, three times. Samples were then prepared for SDS-PAGE and analyzed by immunoblotting as described above.

Sequence conservation analysis. Sequence conservation scores were obtained for RBM39 and GSPT1 using the ConSurf web server (https://consurf.tau.ac.il).[46] ConSurf scores were computed using the ConSeq method.[47] In brief, ConSurf homologues were collected from the UniRef90 database using HMMER and up to 250 homologues with 70% to 99% sequence identity were sampled from the list of unique hits for multiple sequence alignment, phylogenetic tree construction, and conservation scoring. ConSurf scores are normalized such that the mean score across all input residues is zero and the standard deviation is one. ConSurf scores are relative, with the lowest scores representing the most conserved positions in the input sequence. As sequence conservation can exhibit significant variation between adjacent residues, LOESS regression was performed on the ConSurf scores to assess the overall conservation profile of each residue with respect to neighboring residues. LOESS regression was performed using the 'lowess' function of the statsmodels package (v.0.12.1) in Python with frac = (10 AA/L), it = 0.

Statistical methods. All statistical tests described were performed as two-sided tests. Pearson coefficients and significance values were calculated using the 'stats.pearsonr' function of the SciPy package (v.1.6.0) in Python. Other statistical parameters including the exact value and definition of *n*, the definition of center, dispersion, precision measures (e.g., mean ± s.d. or s.e.m.), and statistical significance are reported in figures and figure legends. Data for protein degradation assays and cell proliferation assays were graphed and fit to sigmoidal curves by non-

linear regression (GraphPad Prism). For degradation assays, the DC_{50} (IC_{50}), D_{max} (100 − bottom asymptote), and corresponding s.e.m. values were determined from the sigmoidal fit.

Protein computational modeling. Modeling of select GSPT1 mutants was performed using Molecular Operating Environment (Chemical Computing Group). Using the structure of wt GSPT1 complexed to CC-885, DDB1, and CRBN (PDB: 5HXB), each 5-7 residue stretch containing the desired mutation was modeled using the Loop Modeler application (RMSD limit = 1, Loop limit = 100, Energy window = 10). Where possible, PDB loop searching was preferred over de novo loop searching in identifying candidate loops. For each mutant, the candidate obtained with the lowest MM/GBVI energy was selected as the final model.

Data Availability. Data are provided in the main text and figures, supplementary figures (Fig S1-S5) and tables (Supplementary Table).

Code Availability. Custom python scripts for data analysis of CRISPR-suppressor scanning and single sgRNA-mutagenesis experiments are available upon request.

Safety Statement. No unexpected or unusually high safety hazards were encountered in this study.
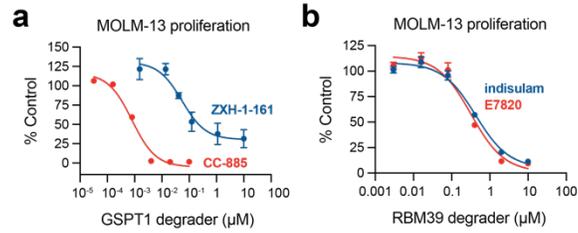
Supplementary Figure S1. CRISPR-suppressor scanning of GSPT1 and RBM39

a) Dose-response curves for MOLM-13 cell proliferation relative to vehicle-treated cells (*y* axis, % control) after E7820 or indisulam treatment for 72 h. Data represent mean ± s.e.m. across three technical replicates. Experiment performed once.

b) Dose-response curves for MOLM-13 cell proliferation relative to vehicle-treated cells (*y* axis, % control) after CC-885 treatment (72 h) or ZXH-1-161 treatment (48 h). Data represent mean ± s.e.m. across three technical replicates. Experiment performed once.
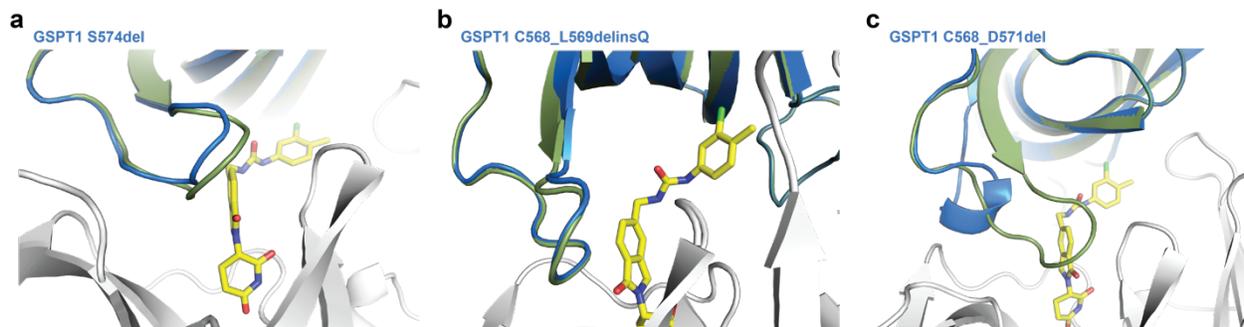
Supplementary Figure S2. Protein computational modeling of selected CC-885 resistant GSPT1 mutants

Structural views of wt GSPT1 (green, PDB: 5HXB) overlaid with modeled GSPT1 resistant mutants (blue) in complex with CC-885 (yellow) and CRBN (white), showing changes in loop structure caused by the indicated indel mutations. GSPT1 S574del is shown in **a**, GSPT1 C568_L569delinsQ is shown in **b**, and GSPT1 C568_D571del is shown in **c**.

a Degrader dose-response

b RBM39 variant degradation with DCAF15 co-transfection

wt D_{max} = 91.60 ± 0.95%
RDAdel D_{max} = 90.6 ± 1.3% P = 0.56 (ns)
R267/G268R D_{max} = 14.3 ± 2.3%
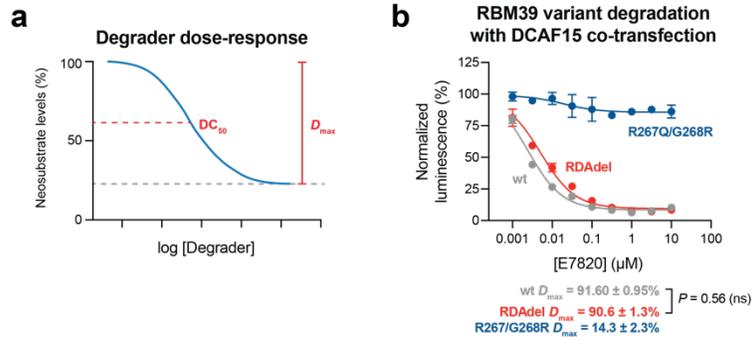
Supplementary Figure S3. Degradation assay data for RBM39 variants

a) Schematic of a generic sigmoidal dose-response curve illustrating neosubstrate degradation (*y* axis, % protein) as a function of degrader concentration (*x* axis), with key parameters (DC_{50} and D_{max}) denoted.

b) Dose-response curves for wt and mutant HiBiT-RBM39-HA cellular protein levels, as indicated by vehicle-normalized luminescence (*y* axis, %), in HEK293T cells co-transfected with a DCAF15 expression plasmid (pcDNA3.1-DCAF15, 100 ng) and treated with E7820 for 24 h. Data represent mean ± s.e.m. across three technical replicates. The D_{max} ± s.e.m. and *P* values (two-sided Student's *t*-test, ns: not significant) are shown below. One of two independent experiments is shown.
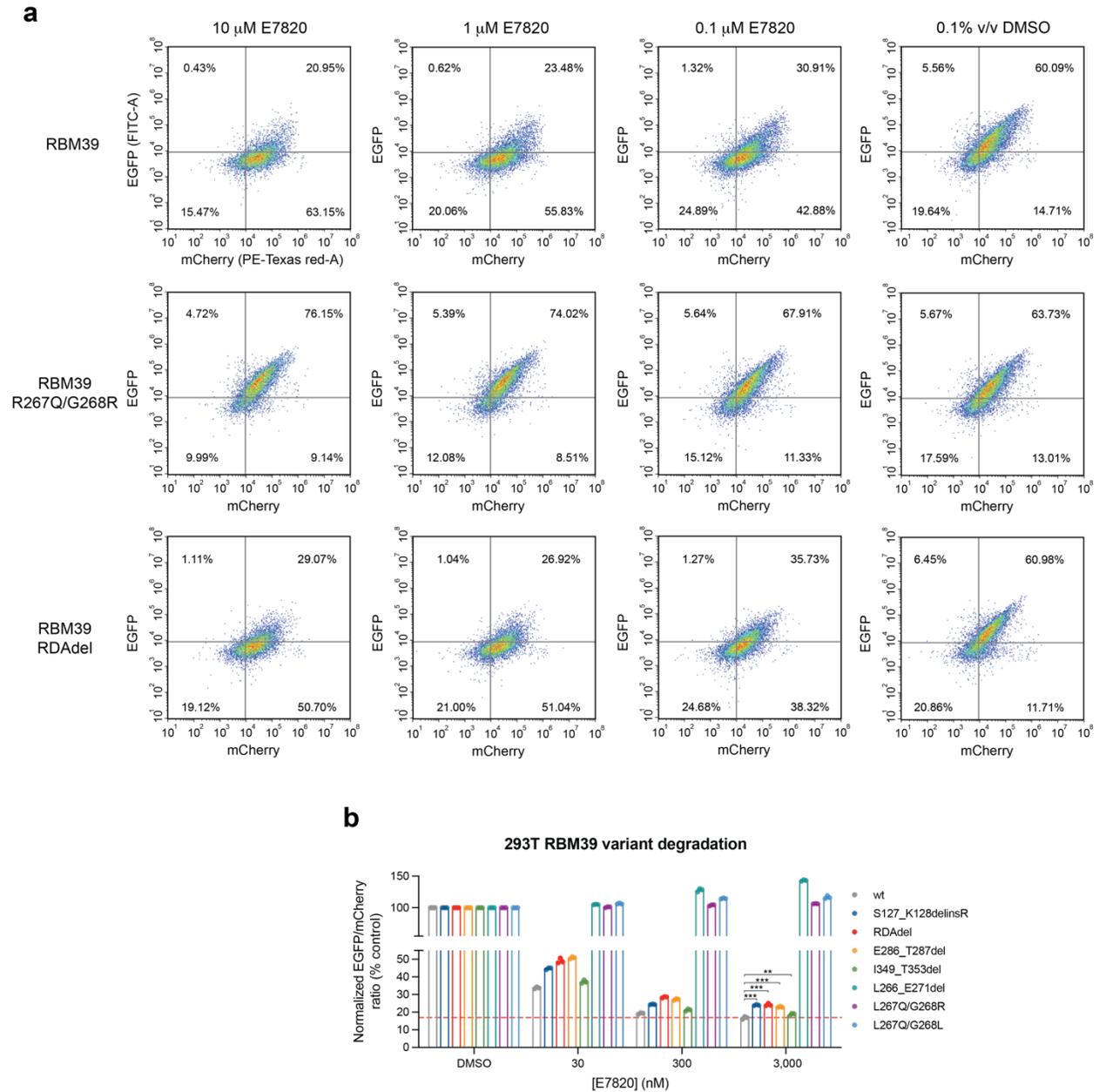
Supplementary Figure S4. EGFP-IRES-mCherry fluorescent reporter assay for monitoring RBM39 variant degradation

a) Flow cytometry pseudocolor density plots showing EGFP (*y* axis) and mCherry (*x* axis) fluorescence for wt RBM39 (top), RBM39 R267Q/G268R (middle), and RBM39 RDAdel (bottom) after vehicle or E7820 treatment for 24 h. Populations were gated on mCherry-positive cells using the gating strategy in **Supplementary Figure S4c**. One of two independent experiments shown.

b) Bar plots showing wt and mutant RBM39 cellular protein levels, as indicated by vehicle-normalized EGFP to mCherry ratio (*y* axis, %), after treatment with E7820 for 24 h in HEK293T cells stably expressing the RBM39-EGFP-IRES-mCherry fluorescent reporter. Data represent

mean ± s.e.m. across three technical replicates. Dotted red line indicates the mean signal of wt 293T treated with 3 µM E7820. Significance levels comparing the degradation levels of the indicated RBM39 mutants to wt RBM39 at 10 µM E7820 are shown ($P < 10^{-2}$: **; $P < 10^{-3}$: ***, two-sided Student's t-test). One of two independent experiments is shown.
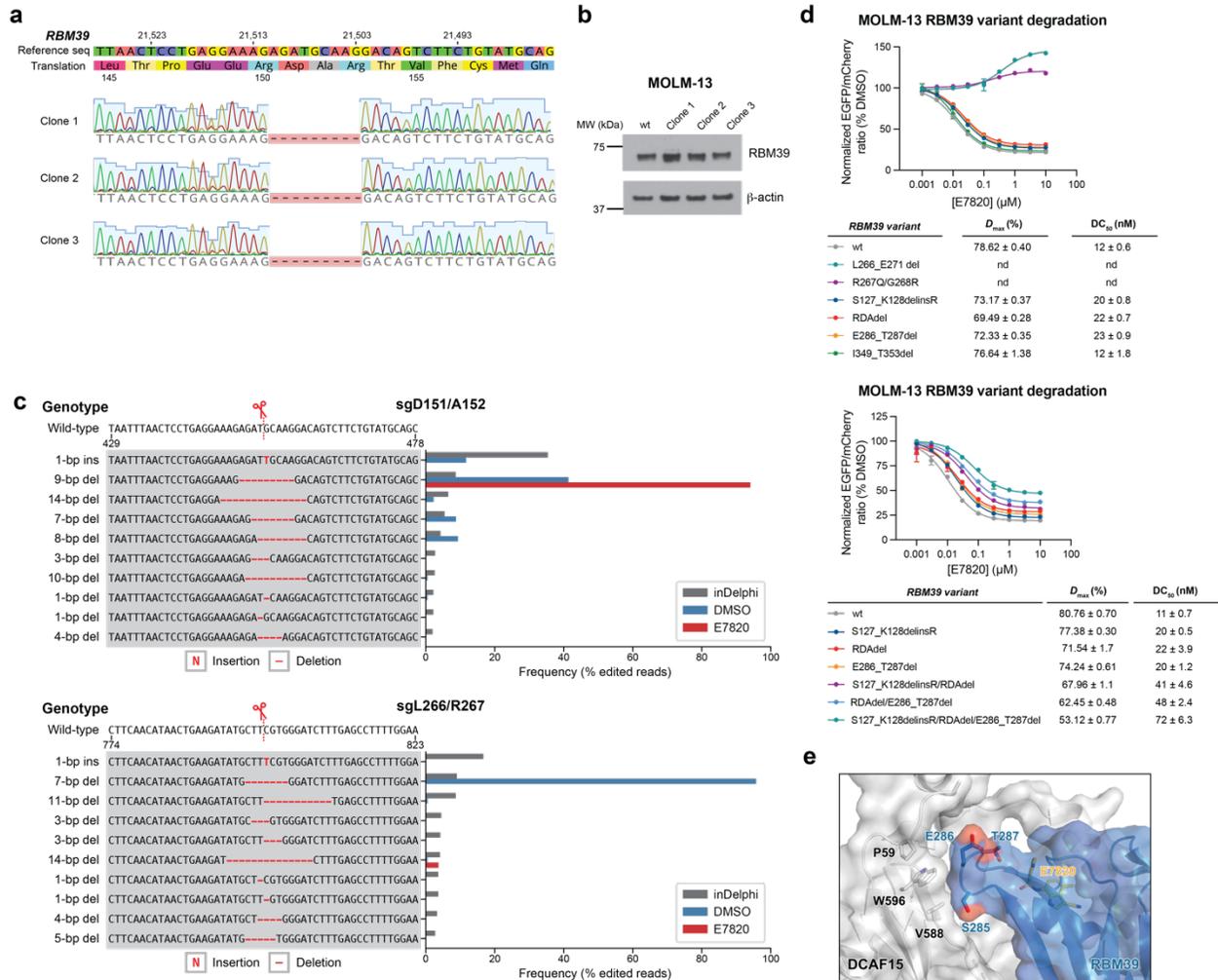
Supplementary Figure S5. Characterization of MOLM-13^{RDAdel} and RBM39 variants identified by single sgRNA CRISPR mutagenesis

a) Schematic depicting the coding mutations and genotypes identified in the MOLM-13^{RDAdel} clonal cell lines.

b) Immunoblots showing that RBM39 are expressed at comparable levels in wt MOLM-13 and MOLM-13^{RDAdel} clonal cell lines. One of two independent experiments is shown.

c) Left: Schematic showing the top 10 editing outcomes predicted by inDelphi for gD151/A152 (top) and gL266/R267 (bottom). Right: Bar plot showing frequency (%, x axis) of each editing outcome (y axis) as predicted by inDelphi (gray) or as observed in MOLM-13 cells after vehicle (blue) or E7820 (red) treatment for 4 weeks.

d) Dose-response curves for wt and mutant RBM39 cellular protein levels, as indicated by vehicle-normalized EGFP to mCherry ratio (y axis, %), in MOLM-13 cells treated with E7820 for 24 h. Top and bottom panels represent different experiments. Data represent mean ± s.e.m. across three technical replicates. Values for D_{max} ± s.e.m. and DC_{50} ± s.e.m. are tabulated (nd: not determined). One of two independent experiments is shown.

S12

e) Structural view of the E7820-DCAF15-RBM39(RRM2) ternary complex depicting the interaction between the RBM39(RRM2) (blue) D284–R289 hairpin and DCAF15 (grey), with E7820 shown in yellow (PDB: 6UE5).
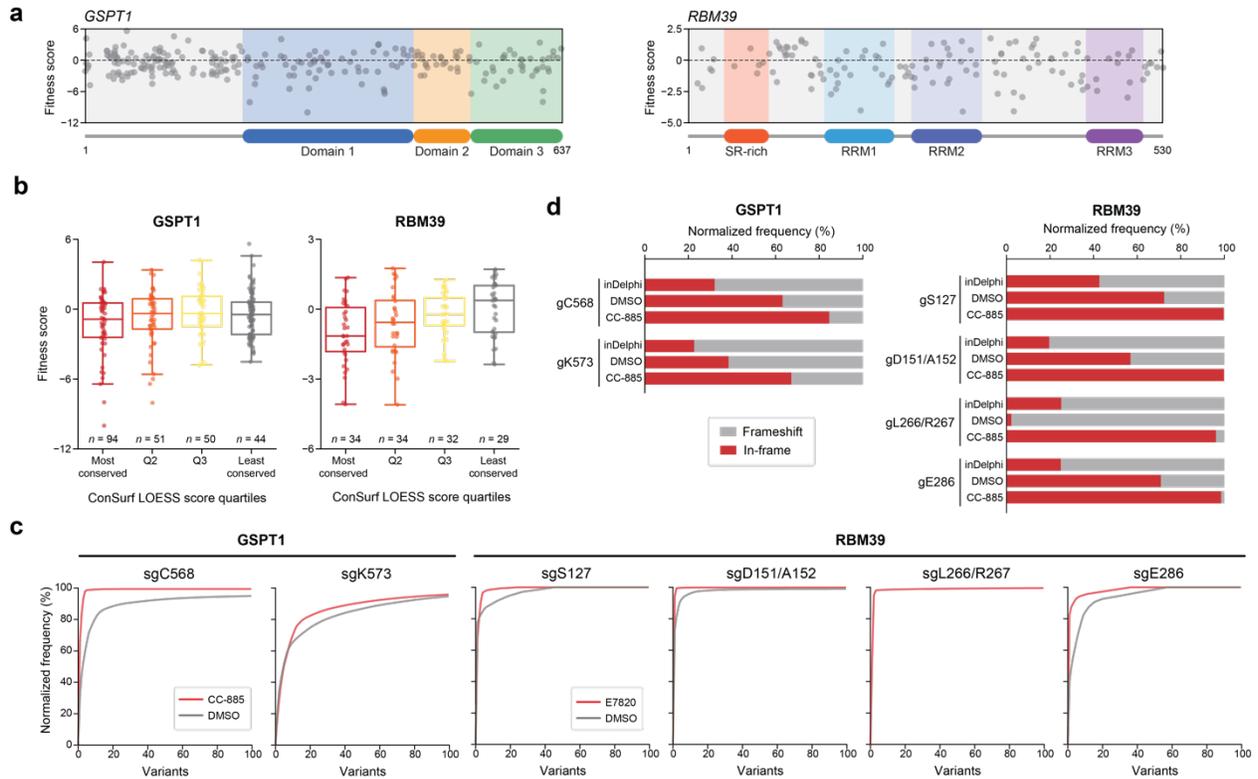
Supplementary Figure S6. Analysis of GSPT1 and RBM39 fitness and CRISPR mutagenesis

a) Scatter plot showing sgRNA fitness scores (*y* axis) in MOLM-13 for CRISPR-scanning of *GSPT1* (left, *n* = 239) or *RBM39* (right, *n* = 129). Fitness scores were calculated as the $\log_2$(fold-change sgRNA enrichment at week 4 under vehicle treatment versus the plasmid library) normalized to the mean of the negative control sgRNAs (*n* = 22 for GSPT1 and 80 for RBM39). The sgRNAs are arrayed by amino acid position in the respective CDS on *x* axis corresponding to the position of the predicted cut site. When the sgRNA cut site falls between two amino acids, both amino acids are denoted. Data points represents mean value across three replicate treatments. Protein domains are demarcated by the colored panels.

b) Box plots with jitter showing fitness scores for sgRNAs targeting *GSPT1* (left) and *RBM39* (right) grouped by the ConSurf LOESS score quartile. sgRNAs were assigned ConSurf LOESS scores based on the amino acid corresponding to their predicted cut site positions; sgRNAs cutting between amino acids were assigned the mean of the flanking amino acids' scores. Dots represent individual sgRNAs. The box shows the median, 25th, and 75th percentiles with whiskers denoting 1.5 × the interquartile range.

c) Cumulative plots showing the normalized variant frequency (*y* axis, %) for the 100 most abundant in-frame edited variants (*x* axis) for each indicated sgRNA after vehicle or drug treatment (see Methods) for 4 weeks. Variants are rank ordered on the *x* axis by decreasing normalized frequency for each respective sgRNA condition. Variant frequency was normalized to the total frequency of all in-frame edited variants. The plot for RBM39 gL266/R267 under vehicle treatment is not shown since in-frame edited variants comprise <1% of the total frequency.

S14

d) Stacked bar plots showing the normalized frequency distribution of frameshift (gray) and in-frame (red) variant types (*y* axis, % of all editing outcomes) for the indicated sgRNAs as predicted by inDelphi or as observed in MOLM-13 after vehicle or drug treatment for 4 weeks. Variant type frequency was normalized to the total frequency of all edited variants. Note that inDelphi does not predict editing outcomes containing substitutions or insertions greater than 1-nt.