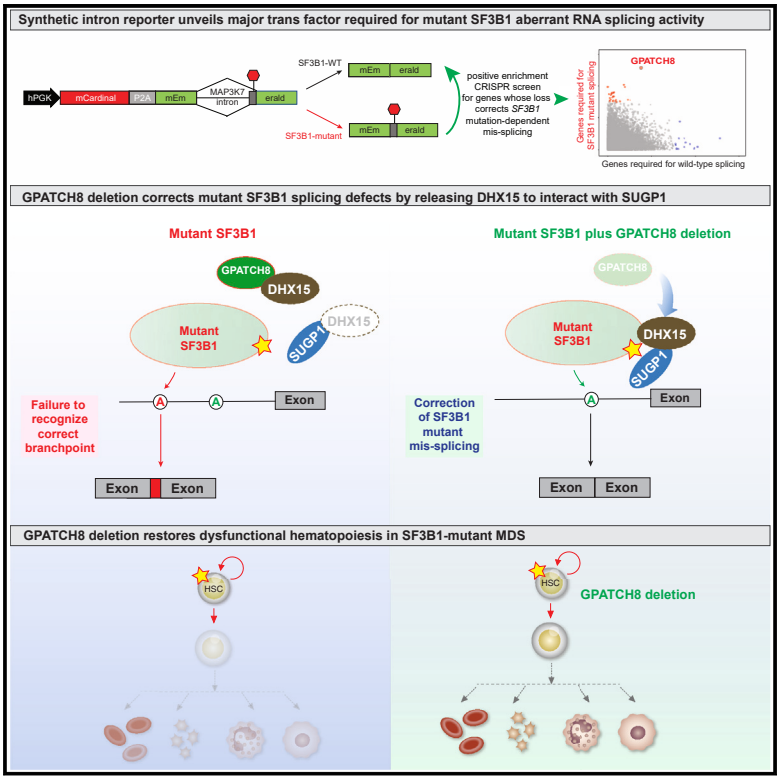
Molecular Cell

Article

GPATCH8 modulates mutant SF3B1 mis-splicing and pathogenicity in hematologic malignancies

Graphical abstract



Highlights

- GPATCH8 is required for mutant SF3B1-dependent splicing alterations
- GPATCH8 is involved in quality control of intronic branchpoint selection
- GPATCH8 opposes activity of SUGP1, and each competes for interaction with DHX15
- Deletion of GPATCH8 corrects aberrant differentiation of SF3B1-mutant MDS cells

Authors

Salima Benbarche, Jose Mario Bello Pineda, Laura Baquero Galvis, ..., Sergei Doulatov, Robert K. Bradley, Omar Abdel-Wahab

Correspondence

rbradley@fredhutch.org (R.K.B.), abdelwao@mskcc.org (O.A.-W.)

In brief

Mutations in the RNA splicing factor *SF3B1* are common in cancer and result in widespread alterations in splicing. Here, Benbarche et al. identify GPATCH8 as a novel splicing factor required for mis-splicing by mutant SF3B1 and demonstrate the therapeutic impact of correcting aberrant splicing in *SF3B1*-mutant cancers by targeting GPATCH8.

Benbarche et al., 2024, Molecular Cell *84*, 1886–1903
May 16, 2024
https://doi.org/10.1016/j.molcel.2024.04.006

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

GPATCH8 modulates mutant SF3B1 mis-splicing and pathogenicity in hematologic malignancies

Salima Benbarche,[1,9] Jose Mario Bello Pineda,[2,3,4,9] Laura Baquero Galvis,[5,9] Jeetayu Biswas,[1] Bo Liu,[1] Eric Wang,[6] Qian Zhang,[1] Simon J. Hogg,[1] Kadeen Lyttle,[1] Ariana Dahi,[1] Alexander M. Lewis,[1] Martina Sarchi,[5] Jahan Rahman,[1] Nina Fox,[1] Yuxi Ai,[1] Sanjoy Mehta,[7] Ralph Garippa,[7] Juliana Ortiz-Pacheco,[8] Zhuoning Li,[8] Mara Monetti,[8] Robert F. Stanley,[1] Sergei Doulatov,[5] Robert K. Bradley,[2,3,*] and Omar Abdel-Wahab[1,10,*]

[1]Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA
[2]Public Health Sciences and Basic Sciences Divisions, Fred Hutchinson Cancer Center, Seattle, WA, USA
[3]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[4]Medical Scientist Training Program, University of Washington, Seattle, WA, USA
[5]Division of Hematology/Oncology, Department of Medicine, Department of Genome Sciences, Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA, USA
[6]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
[7]Gene Editing and Screening Core Facility, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Institute and Cancer Center, New York, NY, USA
[8]Proteomics Innovation Laboratory, Memorial Sloan Kettering Institute and Cancer Center, New York, NY, USA
[9]These authors contributed equally
[10]Lead contact
*Correspondence: rbradley@fredhutch.org (R.K.B.), abdelwao@mskcc.org (O.A.-W.)
https://doi.org/10.1016/j.molcel.2024.04.006

SUMMARY

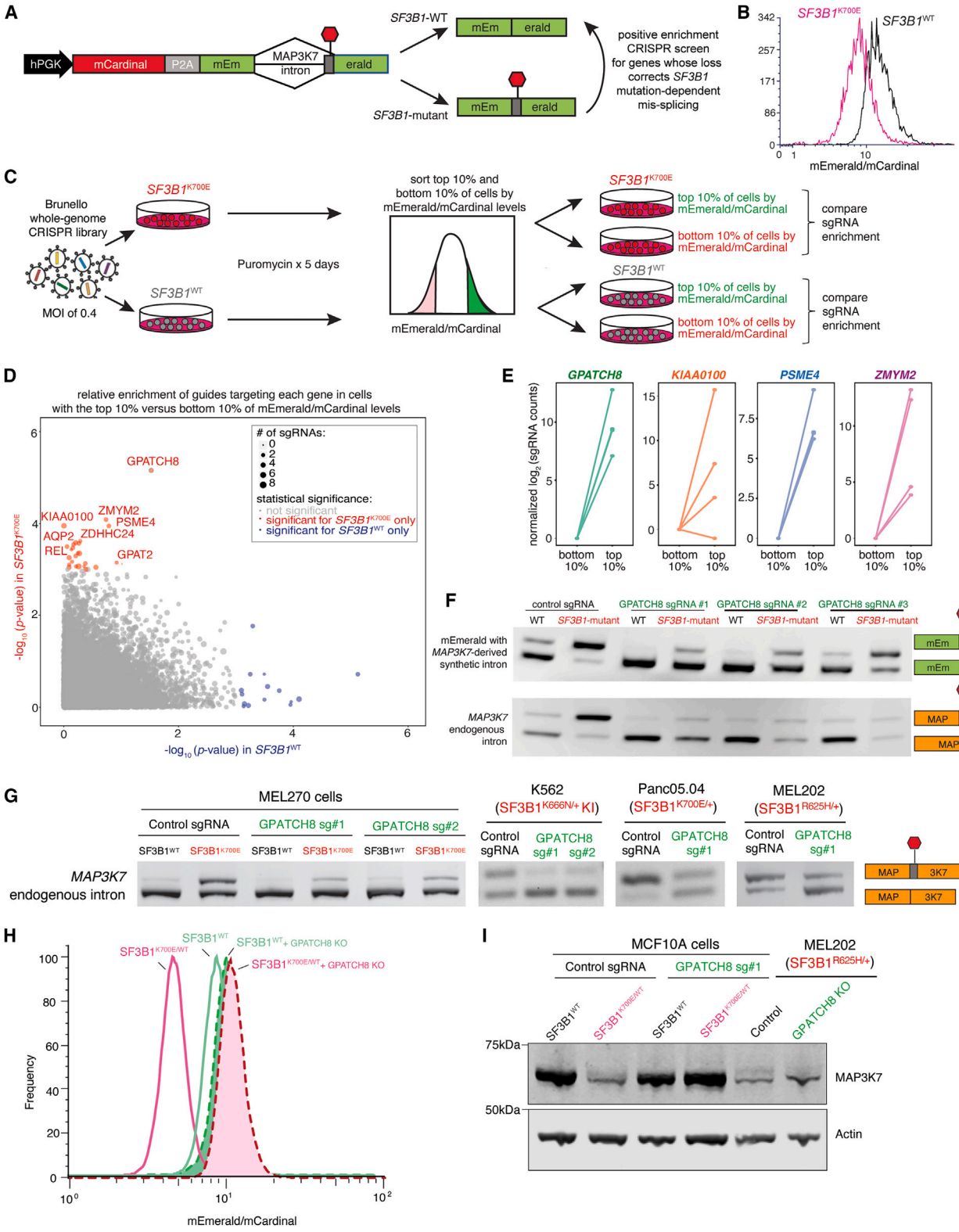
Mutations in the RNA splicing factor gene *SF3B1* are common across hematologic and solid cancers and result in widespread alterations in splicing, yet there is currently no therapeutic means to correct this mis-splicing. Here, we utilize synthetic introns uniquely responsive to mutant SF3B1 to identify *trans* factors required for aberrant mutant SF3B1 splicing activity. This revealed the G-patch domain-containing protein GPATCH8 as required for mutant SF3B1-induced splicing alterations and impaired hematopoiesis. GPATCH8 is involved in quality control of branchpoint selection, interacts with the RNA helicase DHX15, and functionally opposes SURP and G-patch domain containing 1 (SUGP1), a G-patch protein recently implicated in *SF3B1*-mutant diseases. Silencing of GPATCH8 corrected one-third of mutant SF3B1-dependent splicing defects and was sufficient to improve dysfunctional hematopoiesis in *SF3B1*-mutant mice and primary human progenitors. These data identify GPATCH8 as a novel splicing factor required for mis-splicing by mutant SF3B1 and highlight the therapeutic impact of correcting aberrant splicing in *SF3B1*-mutant cancers.

INTRODUCTION

The discovery of frequent mutations in genes encoding components of the RNA splicing machinery was among the most unexpected findings from genomic sequencing efforts. Mutations in RNA splicing factors are the single most common class of mutations in patients with myelodysplastic syndromes (MDS) and also highly recurrent in acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and non-mucosal forms of melanoma.[1–6] These same mutations are also seen in 2%–5% of patients with breast cancer and non-small cell lung cancer as well as other common solid cancers.[7,8] Of the RNA splicing factors recurrently mutated in cancer, mutations altering the core RNA splicing factor SF3B1 are the most prevalent.[9]

Mutations in *SF3B1* occur as heterozygous gain of function point mutations affecting specific amino acid residues concentrated in the HEAT repeat domains 4–7 of SF3B1.[1–6] A number of prior studies delineated that cancer-associated *SF3B1* mutations globally alter RNA splicing in a manner distinct from SF3B1 loss and result in widespread use of aberrant branchpoint nucleotides.[10,11] This change in RNA splicing is consistent with SF3B1's role as part of the U2 small nuclear ribonucleoprotein particle (snRNP) complex, which is responsible for branchsite recognition in early spliceosome assembly. During the early stages of RNA splicing, SF1 binds to the canonical branchsite with the help of the U2AF heterodimeric complex, which recognizes the canonical 3′ splice site (3′ss) and its adjacent polypyrimidine tract.[12,13] The U2 snRNP complex is then recruited to the branchsite, where the U2 snRNA base pairs with the sequences around the branchsite, thereby replacing SF1 and U2AF. Three RNA-dependent ATPase helicases (DDX42, DDX46, and DHX15) associate with the U2 snRNP complex and perform

CellPress
OPEN ACCESS



(legend on next page)

critical roles facilitating assembly and disassembly of proteins at the branchsite during this early stage of RNA splicing.[14]

Cancer-associated mutations in *SF3B1* have repeatedly been shown to alter the selection of branchpoints (BPs) by the U2 snRNP complex and, as a consequence, result in the use of cryptic pre-mRNA 3′ss and generation of aberrantly spliced mRNA.[10,11,15] Despite the consistency of this result, it is not entirely clear how mutations in *SF3B1* result in mis-splicing. One prevailing hypothesis is that mutations in *SF3B1* disrupt physical interactions between SF3B1 and auxiliary splicing factors required for normal branchsite recognition. For example, several studies have suggested that mutations in *SF3B1* weaken the interaction of SF3B1 with a variety of candidate RNA-splicing-related proteins, including SURP and G-patch domain containing 1 (SUGP1),[15] Prp5,[16] and DDX42.[14,17] Of these, a body of data has implicated the G-patch-domain-containing protein SUGP1 in the pathogenic changes induced by mutant SF3B1.[15,18–20] SUGP1 is involved in splicing fidelity during early spliceosome assembly and physically associates the SF3b and U2AF complexes with the ATP-dependent RNA helicase DHX15.[18,21,22]

The G-patch domain is a short (~45 amino acid), flexible, glycine-rich motif that binds and stabilizes the conformations of DHX helicases in a manner that increases their RNA affinities, helicase, and ATPase activities.[23,24] The G-patch domain of SUGP1 activates the helicase activity of DHX15, thereby allowing it to displace SF1 from the branchsite and permitting U2 snRNA to recognize the branchsite.[15,21,22] Zhang et al. identified that cancer-associated mutations in *SF3B1* reduce SF3B1's affinity to SUGP1, thereby reducing eviction of SF1 from the branchsite and prohibiting accessibility of the SF3b complex to the correct branchsite.[15] This impairment in SF1 removal from pre-mRNA forces the U2 snRNP complex to utilize aberrant, cryptic BPs characteristic of *SF3B1*-mutant cells. In support of this model, cancers with naturally occurring somatic mutations in *SUGP1* as well as *DHX15* mimic some of the pathogenic splicing changes unique to *SF3B1*-mutant cancer cells.[19,20]

Although the above studies reveal how mutations in *SF3B1* may induce mis-splicing, here, we set out to identify *trans*-acting proteins required for mis-splicing by mutant SF3B1. This revealed the previously poorly characterized protein GPATCH8 as selectively required for mis-splicing across diverse *SF3B1*-mutant models and the multitude of *SF3B1* hotspot mutations.

RESULTS

Synthetic introns reveal *trans*-acting factors required for mutant SF3B1-dependent mis-splicing

We recently harnessed the aberrant RNA splicing of SF3B1 mutant cells to develop synthetic introns that were efficiently spliced in cancer cells bearing *SF3B1* mutations but unspliced in otherwise isogenic wild-type (WT) cells.[25] This approach allows for quantitative assessment of protein production when the synthetic intron is used to regulate the expression of a fluorescent protein.

We initially hypothesized that synthetic intron-regulated expression of fluorescent protein production could serve as a powerful tool to discover *trans*-acting proteins required for mutant SF3B1 aberrant splicing activity. To accomplish this, we utilized a synthetic intron based on endogenous mis-splicing of the kinase gene *MAP3K7*. *MAP3K7* has consistently been shown to be mis-spliced across *SF3B1* hotspot mutations, cancer cell backgrounds, and human and mouse cells.[26,27] *SF3B1* mutation-dependent mis-splicing of *MAP3K7* results in use of an aberrant intron-proximal 3′ss and open-reading frame disruption in a manner that triggers nonsense-mediated mRNA decay (NMD) and reduces MAP3K7 protein production. Although the endogenous MAP3K7 mis-spliced intron is 1,407 nt, the synthetic version is 250 nt in length and consists of only the first 100 nt and last 150 nt of the mis-spliced intron. When introduced into the coding sequence of mEmerald within a vector that allows for constitutive expression of mCardinal, *SF3B1*-mutant cells stably expressing this construct produce ~1.9-fold less fluorescent protein (as read out by the mEmerald:mCardinal ratio via flow cytometry) than isogenic *SF3B1*-WT cells (Figures 1A, 1B, and S1A).

We next sought to use this particular synthetic intron construct for positive enrichment whole-genome CRISPR screens to identify proteins whose loss resulted in correction of *SF3B1* mutation-dependent mis-splicing of the MAP3K7 synthetic intron, as read out by single guide RNAs (sgRNAs) enriched in the top 10% of mEmerald:mCardinal expressing *SF3B1*-mutant cells

Figure 1. Synthetic introns reveal *trans*-acting factors required for mutant SF3B1-dependent mis-splicing

(A) Schema of synthetic intron fluorescent reporter utilizing a *MAP3K7*-derived synthetic intron to report mutant SF3B1 splicing activity.

(B) Histogram of mEmerald/mCardinal fluorescent protein expression in isogenic MCF10A cells with or without knockin of heterozygous *SF3B1*^{K700E} mutation and stable expression of the MAP3K7 synthetic intron reporter from (A).

(C) Schema of whole-genome CRISPR screen from cells in (B) evaluating enrichment of sgRNAs in top 10% and bottom 10% mEmerald/mCardinal-expressing cells.

(D) Relative per-gene enrichment in SF3B1-mutant versus wild-type (WT) cells from a whole-genome CRISPR knockout screen to identify genes whose loss corrects mis-splicing in SF3B1-mutant cells using the reporter in (A). Genes shown in red and blue reached statistical significance (defined as p value < 0.001).

(E) Normalized $\log_2$ sgRNA read counts in bottom versus top 10% mEmerald/mCardinal expressing cells for key genes from (D).

(F) RT-PCR validating that GPATCH8 knockout antagonizes *SF3B1*^{K700E} mutant-dependent mis-splicing of both the synthetic intron (top) and endogenous MAP3K7 (bottom).

(G) As in (F) for endogenous *MAP3K7* utilizing uveal melanoma (MEL270) cells with or without *SF3B1*^{K700E} mutation, AML (K562) cells with or without knockin of heterozygous *SF3B1*^{K666N} mutation, pancreatic cancer (Panc05.04) cells with endogenous *SF3B1*$^{K700E/WT}$ mutation, and uveal melanoma (MEL202) cells with endogenous *SF3B1*$^{R625H/WT}$ mutation.

(H) As in (B) but with or without GPATCH8 knockout.

(I) Western blot of MAP3K7 protein in MCF10A cells from (H) and MEL202 cells from (G).

See also Figure S1.

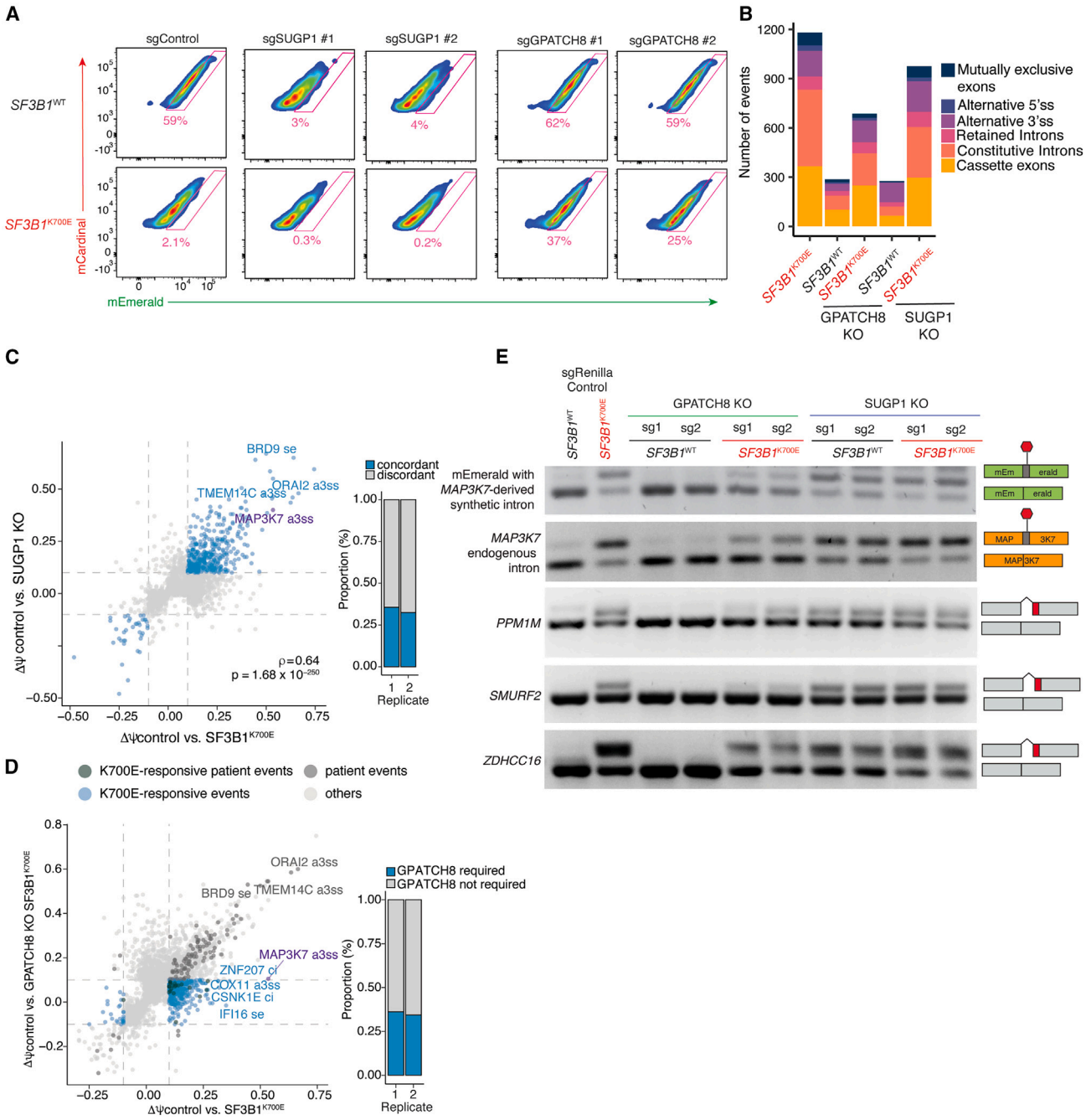


Figure 2. Antagonistic effects of SUGP1 versus GPATCH8 deletion on SF3B1 mutation-induced aberrant RNA splicing

(A) Representative flow cytometry analysis evaluating impact of SUGP1 versus GPATCH8 deletion on splicing of the MAP3K7-derived synthetic intron reporter. Experiment performed in K562 cells with or without knockin of heterozygous *SF3B1*^{K700E} mutation and stable CRISPR-Cas9 knockout (KO) of GPATCH8 or SUGP1 as indicated.

(B) Number and types of splicing events altered in K562 cells from (A) by genome-wide RNA-seq analysis. Each mutant condition was compared with sgControl.

(C) Scatter plot of isoform expression values in *SF3B1*^{K700E} mutant cells compared with SUGP1 KO cells (left) and bar plot indicating proportion of splicing events in *SF3B1*^{K700E} mutant cells that are concordant with those in SUGP1 KO cells across two biological replicates. a3ss, alternative 3′ splice site; ci, constitutive intron; se, cassette exon. Correlation coefficient (ρ) and *p* value estimated via Pearson's correlation.

(legend continued on next page)

(Figure 1C). This screen was performed in MCF10A breast cancer cells with or without endogenous heterozygous knockin of the *SF3B1*^{K700E} mutation and stable expression of the MAP3K7 synthetic intron construct in biological duplicate. These *SF3B1*$^{K700E/WT}$ MCF10A cells contain a heterozygous point mutation in the endogenous *SF3B1* locus and recapitulate the molecular alterations in RNA splicing seen in SF3B1 mutant breast cancer.[7] Following transduction and stable selection of the Brunello whole-genome CRISPR sgRNA library in these cells, we sorted the top and bottom 10% of mEmerald:mCardinal expressing cells and compared sgRNAs enriched in these populations across *SF3B1*-mutant and WT cells (Figure 1C). We generated sgRNAs against mEmerald and mCardinal and delineated a gating strategy for mEmerald and mCardinal expressing cells, as shown in Figure S1A.

The positive enrichment screen revealed a number of genes where >3 sgRNAs were statistically enriched (defined as p value < 0.001) in the mEmerald:mCardinal high population in *SF3B1*-mutant cells relative to WT, including *GPATCH8*, *KIAA0100*, *PSME4*, and *ZMYM2* (Figures 1D and 1E). Given the association of G-patch-domain-containing proteins with RNA metabolism[24] and recent studies linking SUGP1, another G-patch-domain-containing protein, to mutant SF3B1 splicing,[15,18–20] we next evaluated the potential role of GPATCH8 in *SF3B1*-mutation-dependent mis-splicing more carefully. In validation studies using a set of individual sgRNAs against GPATCH8, knockout (KO) of GPATCH8 corrected MAP3K7 synthetic intron mis-splicing as revealed by RT-qPCR of the reporter transcript, RNA sequencing (RNA-seq) of the endogenous transcript, and a ~2.2-fold increase in the mEmerald:mCardinal fluorescent protein reporter (Figures 1F–1H and S1B–S1E).

Beyond impacting the synthetic intronic sequence, GPATCH8 KO also corrected mis-splicing of the endogenous MAP3K7 transcript (Figures 1F and 1G). The correction of mis-splicing of endogenous MAP3K7 upon GPATCH8 KO was consistent across breast cancer (MCF10A), AML (K562), pancreatic cancer (Panc05.04), and uveal melanoma (MEL270 and MEL202) cell lines with either genetically engineered or naturally occurring *SF3B1* mutations (Figures 1F–1H, S1D, and S1E). Moreover, GPATCH8 KO also corrected mis-splicing across all recurrent *SF3B1* missense mutations, including K700E, K666N, and R625H mutations (Figures 1F–1H, S1D, and S1E). As noted earlier, *SF3B1* mutation-dependent mis-splicing of MAP3K7 resulted in reduced MAP3K7 protein expression due to NMD.[26,27] Consistent with the correction of MAP3K7 mis-splicing by GPATCH8 KO, stable GPATCH8 deletion also rescued MAP3K7 protein levels in a variety of *SF3B1*-mutant cancer cell lines (Figure 1I). Of note, reduced GPATCH8 mRNA expression was confirmed by qRT-PCR in each of the above cell lines (Figures S1F–S1J), and there was minimal effect on cell viability or proliferation following GPATCH8 suppression in these immortalized cell lines in basal growth conditions (Figures S1K–S1N).

Antagonistic effects of SUGP1 versus GPATCH8 deletion on *SF3B1* mutation-induced aberrant RNA splicing

We next determined the impact of GPATCH8 deletion on RNA splicing genome-wide. We performed RNA-seq analyses of both AML (K562) and breast cancer (MCF10A) cells with or without knockin of the *SF3B1*^{K700E} mutation and with or without GPATCH8 deletion.

SUGP1 is one of >20 human G-patch-domain-containing proteins, and several recent studies have suggested that disruption of SUGP1 recapitulates some of the RNA splicing defects caused by mutant SF3B1. Interestingly, using the MAP3K7 synthetic intron bichromatic reporter, SUGP1 deletion phenocopied *SF3B1* mutations with both alterations resulting in a very similar reduction in mEmerald:mCardinal fluorescent protein compared with parental WT cells. By contrast, GPATCH8 deletion massively increased mEmerald:mCardinal fluorescent protein in *SF3B1*-mutant cells (as seen earlier) (Figures 2A and S1O). We therefore utilized RNA-seq studies to globally compare the impact of SUGP1 deletion with GPATCH8 deletion or SF3B1 mutation.

We found that deletion of either GPATCH8 or SUGP1 alone resulted in global alterations of RNA splicing compared with parental cells (Figures S1O and S2A–S2C). Importantly, however, the changes in RNA splicing induced by deletion of either G-patch domain protein were less than the magnitude of changes seen with the *SF3B1*^{K700E} mutation (Figures 2B and S2C). We next evaluated the impact of each RNA splicing factor perturbation on categories of alternative RNA splicing events and the directionality of each type of splicing event. Cassette exon splicing, followed by intron retention, and then alternative 3′ss selection were the most impacted types of splicing events seen with *SF3B1* mutation or deletion of either GPATCH8 or SUGP1 (Figures 2B and S2D). Consistent with prior studies,[15,18] mutation of *SF3B1* or deletion of SUGP1 promoted usage of a more intron-proximal alternative 3′ss. Overall, ~35% of differentially spliced events showed concordant directionality when comparing treatment responses with the *SF3B1*^{K700E} mutation and the SUGP1 deletion, inclusive of splicing events observed in *SF3B1*-mutant AML[28] (Figure 2C). The conformity between responses represents a reproducible, statistically significant positive correlation between these two RNA splicing perturbants (Pearson's correlation = 0.64, $p < 1.68 \times 10^{-250}$). Deletion of GPATCH8 corrects ~30% of *SF3B1*^{K700E} splicing events, likely corresponding to mutant-induced alternative splicing dependent on intact GPATCH8 (Figure 2D). The GPATCH8 deletion splicing response was marginally correlated to that of the SF3B1 K700E mutation and minimally overlapped with splicing alterations

(D) Scatter plot (left) of splicing events in *SF3B1*^{K700E} cells compared with *SF3B1*^{K700E} cells with deletion of GPATCH8 inclusive of events observed in MDS patients (green, dark gray), and bar plot (right) indicating proportion of *SF3B1*^{K700E} splicing events where GPATCH8 is required (blue) or not required (gray), across two biological replicates.

(E) RT-PCR of splicing events regulated by mutant SF3B1 and/or influenced by deletion of SUGP1 or GPATCH8. Experiment performed in K562 cells with or without knockin of heterozygous *SF3B1*^{K700E} mutation and stable KO of GPATCH8 or SUGP1.

See also Figure S2.

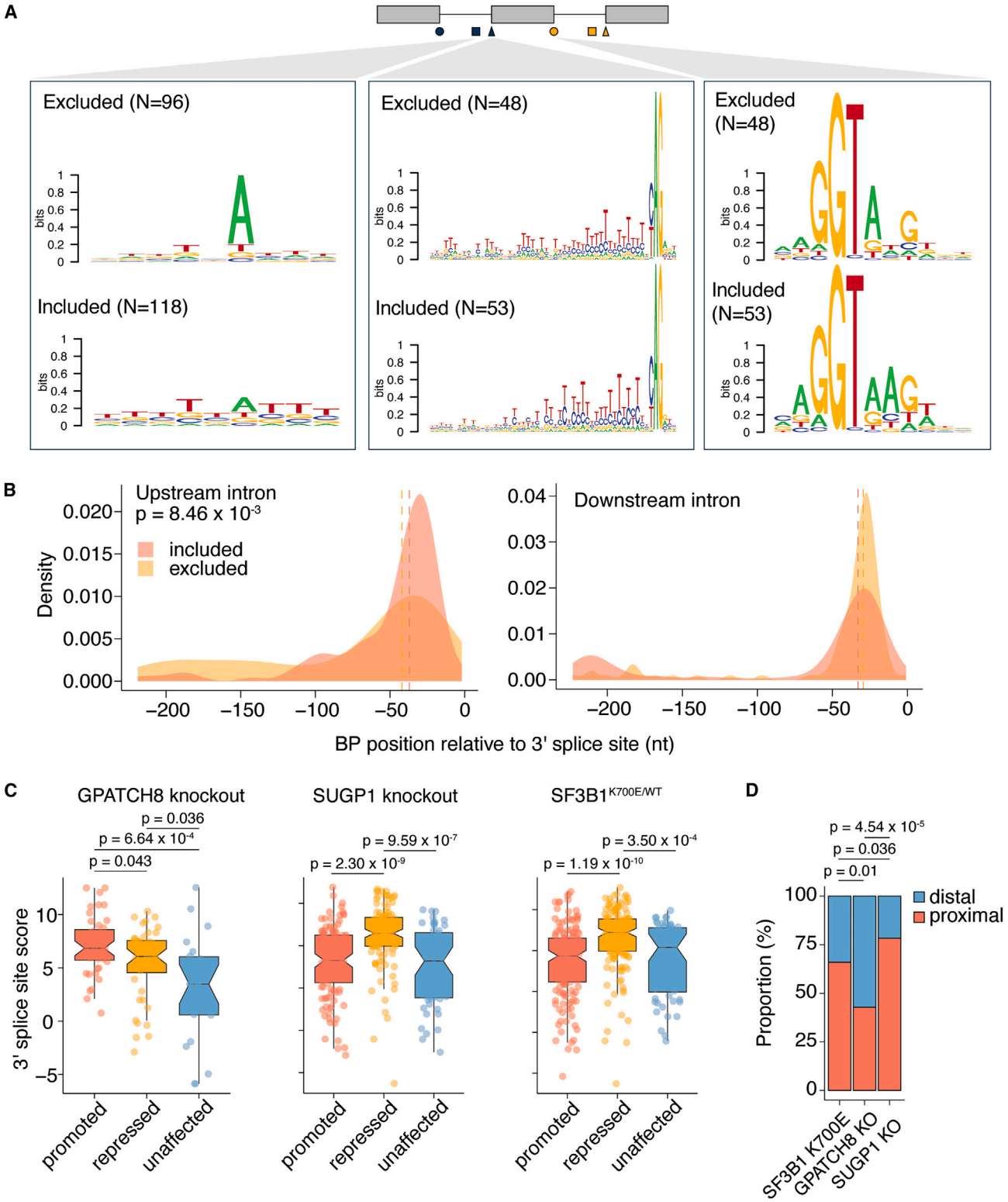


Figure 3. Opposing effects of GPATCH8 and SUGP1 on 3′ splice site usage

(A) Sequence logos of the 5′ splice site (circle), branchpoint sequence (square), and 3′ splice site (triangle) surrounding cassette exons skipped or included upon GPATCH8 deletion in SF3B1 wild-type (WT) K562 cells.

(B) Branchpoint (BP) positions relative to the 3′ splice site in introns flanking cassette exons in skipped or included upon GPATCH8 deletion.

(legend continued on next page)

associated with SUGP1 deletion (Figures S2E and S3D). These results suggest that GPATCH8 deletion opposes a portion of the RNA splicing defects created by mutant SF3B1, while SUGP1 deletion phenocopies a portion of *SF3B1*-mutant aberrations. The contrasting activity of GPATCH8 and SUGP1 on mutant SF3B1 is evidently demonstrated by RNA-seq and RT-PCR of splicing events within *MAP3K7*, *PPM1M*, *SMURF2*, and *ZDHHC16*, as a few examples (Figure 2E). Gene set enrichment analysis (GSEA) revealed that splicing events perturbed by mutant SF3B1 and corrected by GPATCH8 deletion were preferentially associated with genes encoding proteins involved in signal transduction ($p = 0.042$, adjusted for multiple-hypothesis testing via Benjamini-Hochberg false discovery rate [FDR] correction) (Figure S2F). Importantly, the impact of GPATCH8 deletion on RNA splicing alterations caused by mutant SF3B1 was seen across different cancer cell lines (K562 and MCF10A cells) as well as distinct *SF3B1* hotspot mutations (Figures S1F, S2G, and S3A–S3C). Notably, we observed that a number of SF3B1-mutant splicing changes were dependent on the presence of SUGP1, exclusive of the MAP3K7 aberrant isoform (Figure S3E).

Given that cassette exon splicing events were the most altered type of splicing change with *SF3B1* mutation, GPATCH8 deletion, or SUGP1 deletion, we next evaluated splicing signals in the introns flanking cassette exons to gain insight into how GPATCH8 and SUGP1 regulate splicing. Cassette exons excluded upon GPATCH8 deletion, corresponding to GPATCH8-promoted events, are associated with weaker 3′ss containing shorter polypyrimidine tracts. Although the spatial distribution of BPs differs based on response to GPATCH8 KO, with excluded exons linked to upstream introns containing more intron-proximal BPs, the majority are restricted to a small region within the 3′ intron. The BPs upstream of these GPATCH8-promoted exons more frequently utilize the canonical adenosine BP nucleotide and have intron-proximal uridine enrichment, both features characteristic of human BPs.[29] By contrast, GPATCH8-repressed exons, those included upon GPATCH8 KO, are associated with degenerate BP sequence motifs (Figures 3A and 3B). Each of these findings suggests that GPATCH8 may normally function in the selection of suboptimal 3′ss, possibly through a BP-dependent mechanism. GPATCH8 may promote usage of a strong BP, which could lead to concomitant usage of suboptimal 3′ss, possibly by suppressing competing weaker BPs. The upstream intron may be particularly important, as it represents the first site of splicing regulation in cassette exons in the context of co-transcriptional splicing. GPATCH8-repressed exons may have upstream introns enriched for these suboptimal BPs, which could explain the observed directionality. Lastly, GPATCH8-responsive exons are associated with differences in 5′ splice sites (5′ss), with promoted exons harboring weaker 5′ss whose sequences diverge from the canonical GTRAGT motif: increased preference for A at the +3 position and decreased preference for A at the +4 position (defined relative to the intron-exon boundary; Figure 3A). These data suggest that GPATCH8 typically participates in the selection of weak 5′ss, possibly coupling to the upstream 3′ss through an exon-spanning mechanism such as in exon definition. Importantly, the sequence features characteristic of GPATCH8 KO-responsive cassette exons are recapitulated in experiments with both MCF10A and K562 cells (Figures S3A and S3B).

We performed a similar series of analyses of the splicing signals surrounding differentially utilized cassette exons upon SUGP1 deletion. A recent study suggested that SUGP1 participates in the quality control function of DHX15 for repressing suboptimal introns.[22] We similarly observed that the introns upstream of exons excluded after SUGP1 KO (SUGP1-promoted cassette exons) harbor stronger 3′ss relative to included exons (Figures S3F and S3G). Although the polypyrimidine tract is less defined, we observed a preserved preference for G at the +1-position relative to the terminal intronic nucleotide. This nucleotide is an important feature of the canonical yAG|r 3′ss motif.[30] We similarly see differences in branchpoint distribution and sequence preferences in differentially regulated cassette exons. Of note, SUGP1-promoted exons have stronger upstream BPs, those that better conform to the YUNAY consensus motif,[31] compared with SUGP1-repressed exons. These results suggest that branchpoint selection influences SUGP1-mediated repression of weak 3′ss.

Given that cancer-associated mutations in SF3B1 are associated with shifts in usage toward more intron-proximal 3′ss relative to WT, we also evaluated how 3′ss selection changes with deletion of GPATCH8, deletion of SUGP1, or the SF3B1$^{K700E/WT}$ mutation. We examined the alternative 3′ss events in these three treatment conditions and identified 3′ss that were utilized more (termed promoted) or less (termed repressed) frequently relative to the control context. We also analyzed the set of 3′ss in these perturbation-responsive introns that were completely unused or did not change in usage frequency by more than 10% relative to control (termed unaffected). Loss of GPATCH8 resulted in usage of more canonical 3′ss, contrasting with the loss of SUGP1 or the SF3B1$^{K700E/WT}$ mutation, which resulted in the increased usage of suboptimal 3′ss (Figure S4A). These results suggest that GPATCH8 and SUGP1 function in the selection and repression of suboptimal 3′ss, respectively, concordant with the findings associated with analysis of GPATCH8- or SUGP1-responsive cassette exons.

We verified these results by estimating the associated splice site strengths in these 3′ss categories using MaxEntScan.[32] 3′ss promoted by GPATCH8 loss had stronger splice sites compared with repressed sites. By contrast, repressed 3′ss were associated with higher splice site scores compared with promoted sites in the context of either the SUGP1 deletion or the SF3B1$^{K700E/WT}$ mutation (Figure 3C).

Finally, we evaluated the direction of splice site preference associated with deletion of GPATCH8, deletion of SUGP1, or

(C) Box-and-whisker plots of MaxEntScan 3′ splice site scores associated with GPATCH deletion (left), SUGP1 deletion (middle), or *SF3B1*$^{K700E/WT}$ (right). p values estimated via a two-sided Mann-Whitney U test.

(D) Proportion of alternative 3′ splice site events showing intron-distal (blue) versus intron-proximal (red) preference shifts relative to WT associated with deletion of GPATCH8, deletion of SUGP1, or *SF3B1*$^{K700E/WT}$ mutation. p values estimated via a two-sided proportion test.

See also Figures S3 and S4.

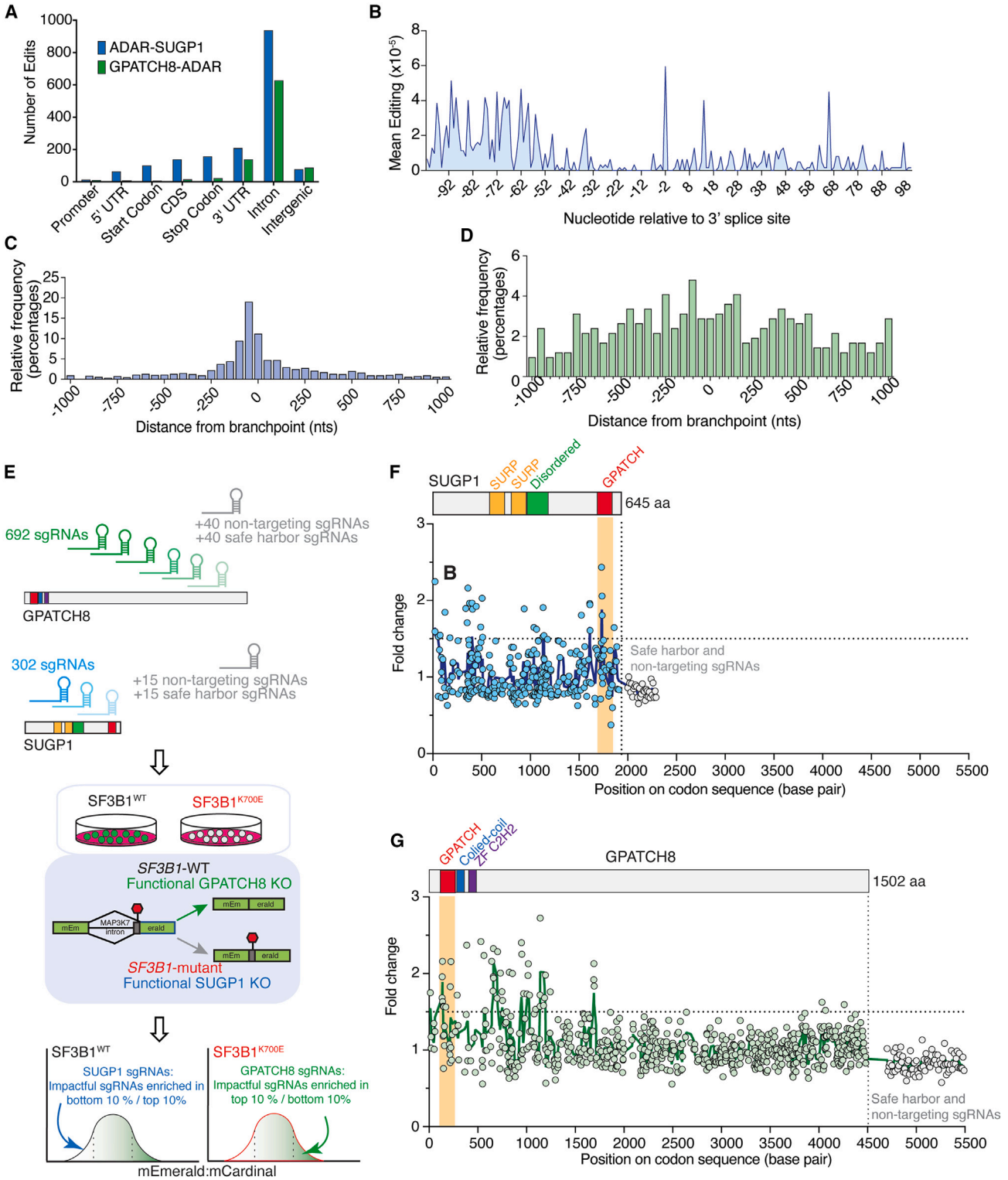


Figure 4. GPATCH8 and SUGP1 bind intronic sequences in pre-mRNA, and the G-patch domains of each protein are required for splicing regulation

(A and B) TRIBE-seq analysis of SUGP1 (blue) and GPATCH8 (green) guided RNA-editing sites in 293T cells showing (A) number of edit sites by location in transcript and (B) edited sites of SUGP1 relative to 3′ splice site.

(C) Relative frequency of RNA editing sites for ADAR-SUGP1 centered around predicted branchpoints.

(legend continued on next page)

SF3B1$^{K700E/WT}$ mutation. SF3B1$^{K700E/WT}$ mutation and SUGP1 KO both promoted usage of an upstream 3′ss relative to the WT-selected site. By contrast, GPATCH8 deletion caused preferential usage of more intron-distal 3′ss (Figures 3D and S4B).

Overall, there is remarkable similarity in splice site sequence preferences and shifts in 3′ss usage frequency between SUGP1 deletion and SF3B1$^{K700E/WT}$. By contrast, the direction of alternative 3′ss usage with SUGP1 deletion and SF3B1$^{K700E/WT}$ mutation is opposite to that seen with GPATCH8 deletion.

Given that mutations in the RNA splicing factors SRSF2 and U2AF1 are also highly recurrent in patients with MDS, we also sought to evaluate if GPATCH8 was required for the pathologic splicing events driven by *SRSF2*^{P95H} or *U2AF1*^{S34F} mutations. To address this, we performed RNA-seq analyses of K562 cells with or without expression of *SRSF2*^{P95H} or *U2AF1*^{S34F} mutations and treated each cell line with GPATCH8 sgRNA or control sgRNA. This effort revealed that GPATCH8 is not required for SRSF2 or U2AF1 mutant mis-splicing (Figures S4C and S4D). GPATCH8 KO failed to correct the hallmark global sequence-specific change in cassette exons usage driven by mutant SRSF2[33] or the change in 3′ss usage driven by mutations in U2AF1.[34]

GPATCH8 and SUGP1 bind intronic sequences in pre-mRNA, and the G-Patch domain of each protein is required for splicing regulation

Given the roles of GPATCH8 and SUGP1 in RNA splicing regulation, we next sought to evaluate RNAs bound by each protein. To accomplish this, we adapted the HyperTRIBE method[35] and fused GPATCH8 or SUGP1 to the catalytic domain of the *Drosophila* ADAR (adenosine deaminase acting on RNA) enzyme. When expressed in cells, each fusion protein marks its binding sites with nearby A-to-I editing events, thereby allowing for global mapping of the mRNA targets of each protein. Editing sites for both SUGP1 and GPATCH8 were most heavily enriched within introns (Figures 4A and S5A). Interestingly, for SUGP1, editing events clustered upstream of canonical 3′ss (Figure 4B) and were enriched at the intronic branchpoint (Figure 4C). By contrast, GPATCH8 editing events were more evenly distributed throughout the intron (Figures 4D and S5B). Of note, we evaluated the overlap in differential splicing events induced by GPATCH8 or SUGP1 KO compared with their RNA-binding targets by TRIBE-seq (all of which were performed in the same cell type). This revealed a statistically significant overlap in RNA-binding and splicing targets for each protein (Figures S5C and S5D).

In addition to a G-patch domain, GPATCH8 also contains coiled coil, C2H2 zinc finger, and SR domains, and the functional importance of each to GPATCH8 function is not yet known. To evaluate the functional importance of each domain of GPATCH8 and SUGP1 in regulating SF3B1 mutant mis-splicing events in a high-throughput manner, we performed CRISPR sgRNA domain scanning[36] of the coding region of GPATCH8 and SUGP1.

We generated a pool of 692 anti-GPATCH8 sgRNAs (with 40 non-targeting and 40 safe harbor negative control sgRNAs) as well as 302 anti-SUGP1 sgRNAs (with 15 non-targeting and 15 safe harbor negative control sgRNAs) (Figure 4E). This pooled series of sgRNAs were introduced into SF3B1 WT and mutant K562 cells bearing the MAP3K7 mEmerald split synthetic intron reporter, and we sequenced sgRNAs enriched in the top 10% and bottom 10% of mEmerald:mCardinal cells by FACS after 2 days of culture. As such, sgRNAs with functional impact on SUGP1 RNA splicing activity would be expected to be enriched in the bottom 10% of fluorescent cells, whereas sgRNAs with functional impact on GPATCH8 RNA splicing activity would be expected to be enriched in the top 10%. This experiment was performed in biological duplicate with good concordance (Figure S5E) and highlighted the functional importance of G-patch domains in both SUGP1 and GPATCH8 (Figures 4F and 4G). Moreover, the entire N-terminal portion of GPATCH8, consisting of the G-patch, coiled coil, and C2H2 zinc finger, was critical for GPATCH8 regulation of SF3B1 mutant mis-splicing.

GPATCH8's G-patch domain is required for splicing regulation

Given the homology of the G-patch domains between GPATCH8 and SUGP1 (Figure 5A), we therefore next sought to delineate if specific residues of GPATCH8's G-patch domain were essential for its splicing activity. To accomplish this, we expressed full-length GPATCH8 or a C terminus truncated version containing only its coiled coil, C2H2 zinc finger, and G-patch domains into the MAP3K7 synthetic intron reporter expressing K562 cells WT or mutant for SF3B1 (Figure 5B). As shown earlier, *SF3B1*-mutant cells had virtually no mEmerald expression, while KO of GPATCH8 rescued mEmerald fluorescence (Figures 5C and 5D). Reintroduction of full-length GPATCH8 into the *SF3B1*-mutant GPATCH8 KO cells suppressed mEmerald expression, underscoring the requirement of GPATCH8 for *SF3B1* mutation-dependent mis-splicing. Interestingly, the N terminus alone of GPATCH8 was sufficient to carry out this activity of GPATCH8. Given that the N terminus of GPATCH8 contains its G-patch domain, we next sought to mutate critical residues in the G-patch domain to determine whether this domain is critical for GPATCH8's role in suppressing aberrant SF3B1 mutant mis-splicing events. To accomplish this, we mutated two conserved glycine residues (GPATCH8 Glycine 52 and Glycine 60), which are conserved across G-patch-domain-containing proteins and previously shown to be critical for the function of other G-patch-domain-containing proteins.[18,24] This revealed

(D) As in (C) but for GPATCH8-ADAR.
(E) Schema of the experimental design for CRISPR-Cas9-tiling scan of the protein coding domains of GPATCH8 and SUGP1.
(F) Normalized CRISPR score of each sgRNA construct (dots) and the smoothed score (line) of the SUGP1 sgRNA tiling screen. Data represent the average of a duplicate experiment.
(G) As in (F) but for GPATCH8.
See also Figure S5.

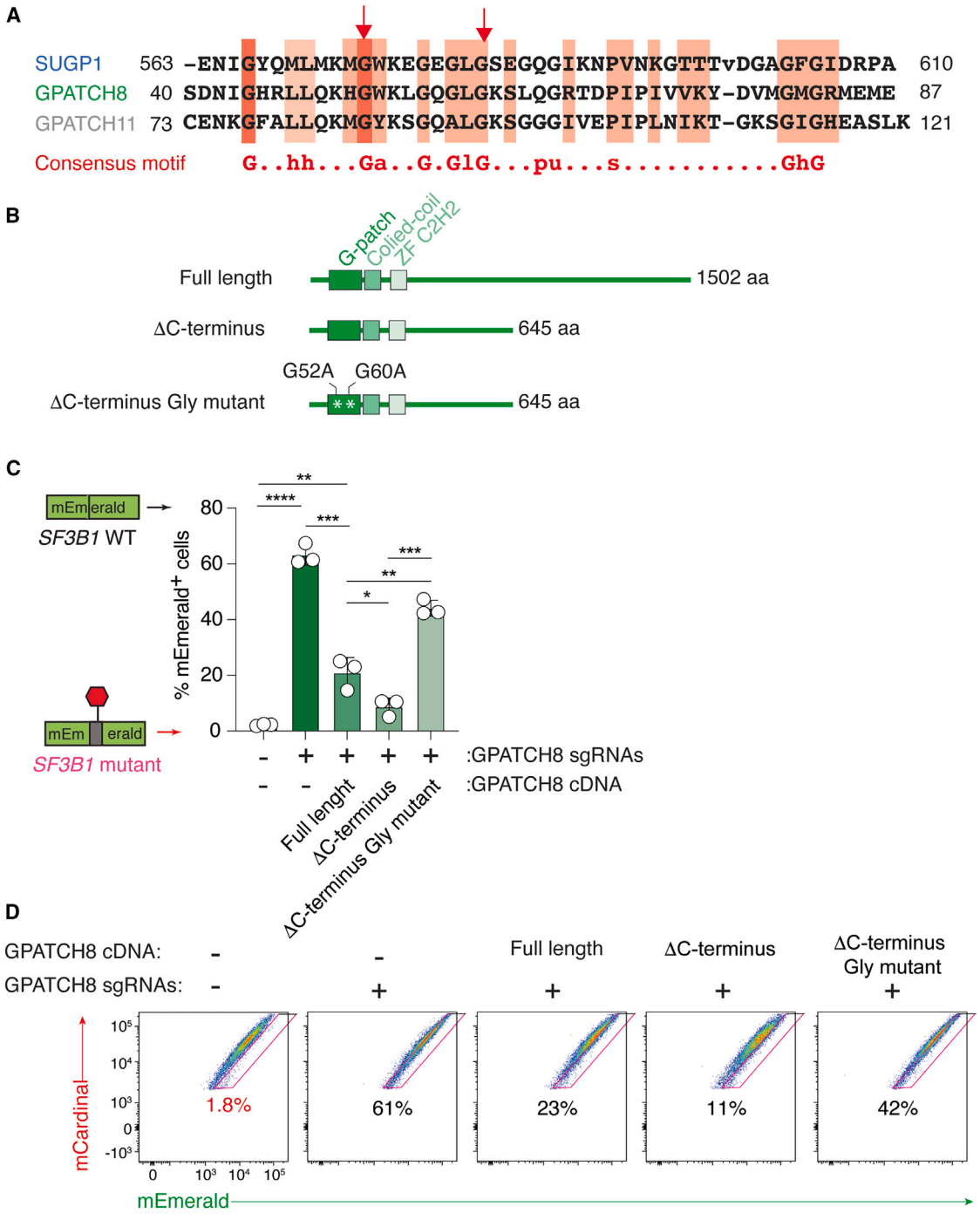


Figure 5. GPATCH8's G-patch domain is required for splicing regulation and is not interchangeable with SUGP1's G-patch domain

(A) Sequence alignment of the G-patch domains of SUGP1, GPATCH8, and GPATCH11. Invariable residues and positions with 70% similarity are highlighted in dark and light pink, respectively. The consensus sequence is shown in red below the alignment with small letters denoting the following: a, aromatic; h, hydrophobic; l, aliphatic; s, small; and u, tiny amino acids. Red arrows indicate the glycine residues mutated in the indicated cDNAs in (B).

(B) Diagram of the full-length GPATCH8 and mutations made, including deletion of the C terminus of GPATCH8 and mutation of the G-patch domain.

(C) Bar graph of % mEmerald^{+} cells in SF3B1$^{K700E/WT}$ knockin K562 cells containing the MAP3K7 synthetic intron interrupting mEmerald's coding sequence. Measurements are from three independent transductions (mean ± SD, unpaired two-sided student t test, * $p < 0.0332$, ** $p < 0.0021$, *** $p < 0.0002$, and **** $p < 0.0001$).

(D) FACS plots representation of mCardinal/mEmerald fluorescence from (C).

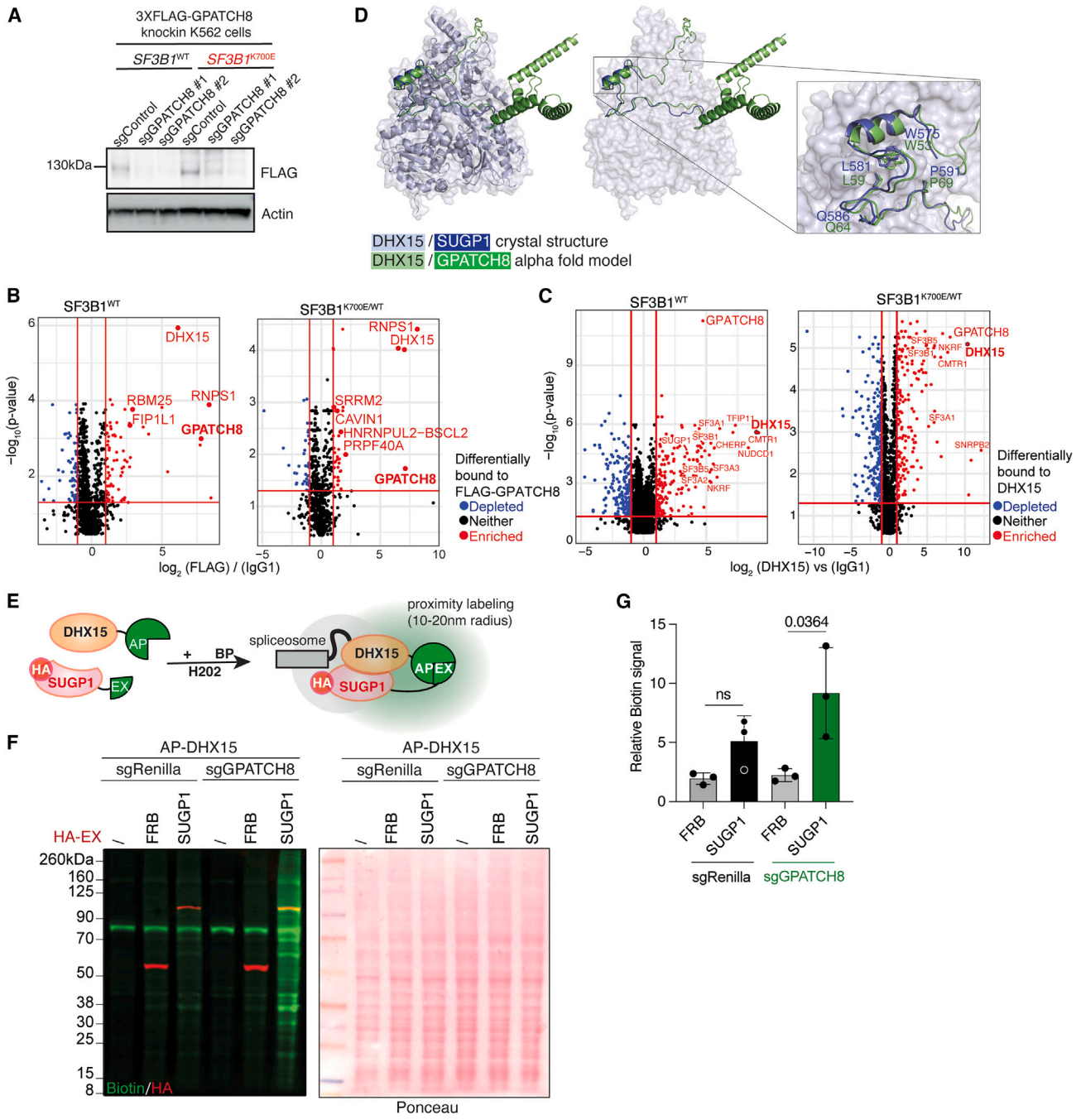


Figure 6. GPATCH8 interacts with DHX15 through its G-Patch domain, and GPATCH8 loss promotes SUGP1/DHX15 interaction

(A) Western blot of N-terminally 3×-FLAG-tagged endogenous GPATCH8 in K562 cells with or without SF3B1^{K700E} heterozygous knockin and with or without GPATCH8 knockout.

(B) Volcano plot of proteins enriched with anti-FLAG versus mouse IgG1 isotype control immunoprecipitation in cells from (A) followed by mass spectrometry. Measurements are from three independent immunoprecipitations. Unpaired t test was used to calculate *p* value in differential analysis, and a volcano plot was generated based on $\log_2$ fold change and q value (multiple testing corrected *p* value). A q value of ≤ 0.05 was considered the statistically significant cut-off.

(C) As in (B) but for immunoprecipitation of endogenous DHX15.

(D) Overlay of published crystal structure of SUGP1 in complex with DHX15 (dark blue and light blue, respectively) with alpha-fold predicted structure of GPATCH8 and DHX15 interaction (dark green and light green, respectively). Inset shows G-patch domains of SUGP1 and GPATCH8.

(E) Diagram of split APEX experiment used to evaluate physical interaction of AP-tagged DHX15 and HA-EX-tagged SUGP1 expressed in 293T cells in the presence versus absence of GPATCH8.

(legend continued on next page)

that mutating two conserved Glycine residues (GPATCH8 Glycine 52 and Glycine 60) within GPATCH8's G-patch domain greatly diminished GPATCH8's splicing activity (Figures 5C and 5D).

DHX15 is the ATPase ligand for GPATCH8

G-patch domains have been shown to enhance the affinity, helicase, and/or ATPase activity of RNA helicases.[24] However, the target helicase remains to be identified for many of the >20 human G-patch proteins. To identify helicase-binding partners of GPATCH8, we next sought to purify endogenous GPATCH8 from *SF3B1*-WT and -mutant cells. Given the lack of commercial antibodies that correctly recognize GPATCH8, we knocked in the sequence encoding the FLAG epitope into the N terminus of the locus encoding *GPATCH8* in K562 cells WT for *SF3B1* or with the *SF3B1*^{K700E} mutation knocked in (Figures 6A and S5F–S5I). Immunoprecipitation of FLAG in these cells followed by mass spectrometry in biological triplicate revealed GPATCH8 as a top protein, as expected, followed by robust interactions with the ATP-dependent helicase DHX15 in both *SF3B1*-mutant and -WT cells (Figure 6B).

Reciprocal unbiased proteomic assessment of DHX15-binding partners identifies GPATCH8 as a top DHX15-interacting protein in SF3B1 WT and mutant cells (Figure 6C). In SF3B1 WT K562 cells, GPATCH8 enrichment with DHX15 was far more significant than SUGP1 enrichment. This was also true for SF3B1 mutant cells, where GPATCH8 was among the most significantly enriched DHX15 interaction partners, and there was no significant enrichment of SUGP1 (Figure 6C).

Recently, Zhang et al. published the co-crystal structure of the human DHX15-SUGP1 G-patch complex.[18] Mapping the predicted structure of the G-patch domain of GPATCH8 onto DHX15 identified that GPATCH8's G-patch domain interacts with DHX15 at the exact same location that DHX15 interacts with SUGP1 (Figure 6D). Consistent with these findings, immunoprecipitation of distinct variants of FLAG-tagged GPATCH8 cDNA constructs followed by western blotting for DHX15 revealed that the interaction between GPATCH8 and DHX15 occurs at the N terminus of GPATCH8 (Figures S6A and S6B). Moreover, the GPATCH8/DHX15 interaction is lost with mutations of the conserved glycine residues (Glycine 52 and Glycine 60) within the G-patch domain of GPATCH8 (Figure S6B).

Our identification of opposing effects of GPATCH8 and SUGP1 in regulating SF3B1 mutant aberrant splicing events suggests that there may be competition between GPATCH8 and SUGP1 in their interactions with DHX15. To test the hypothesis that GPATCH8 deletion promotes DHX15-SUGP1 interaction, we took advantage of a recently published proximity labeling method whereby physical interaction of DHX15 and SUGP1 generates ascorbate peroxidase (APEX).[37] In this system, portions of the cDNA encoding the APEX enzyme are split between DHX15 and SUGP1 and DHX15-SUGP1 physical interaction is thereby readout via the production of APEX-catalyzed biotinylation in the presence of hydrogen peroxide (Figure 6E). Reconstitution of APEX biotinylates proteins within the APEX radius of 10–20 nM.[38]

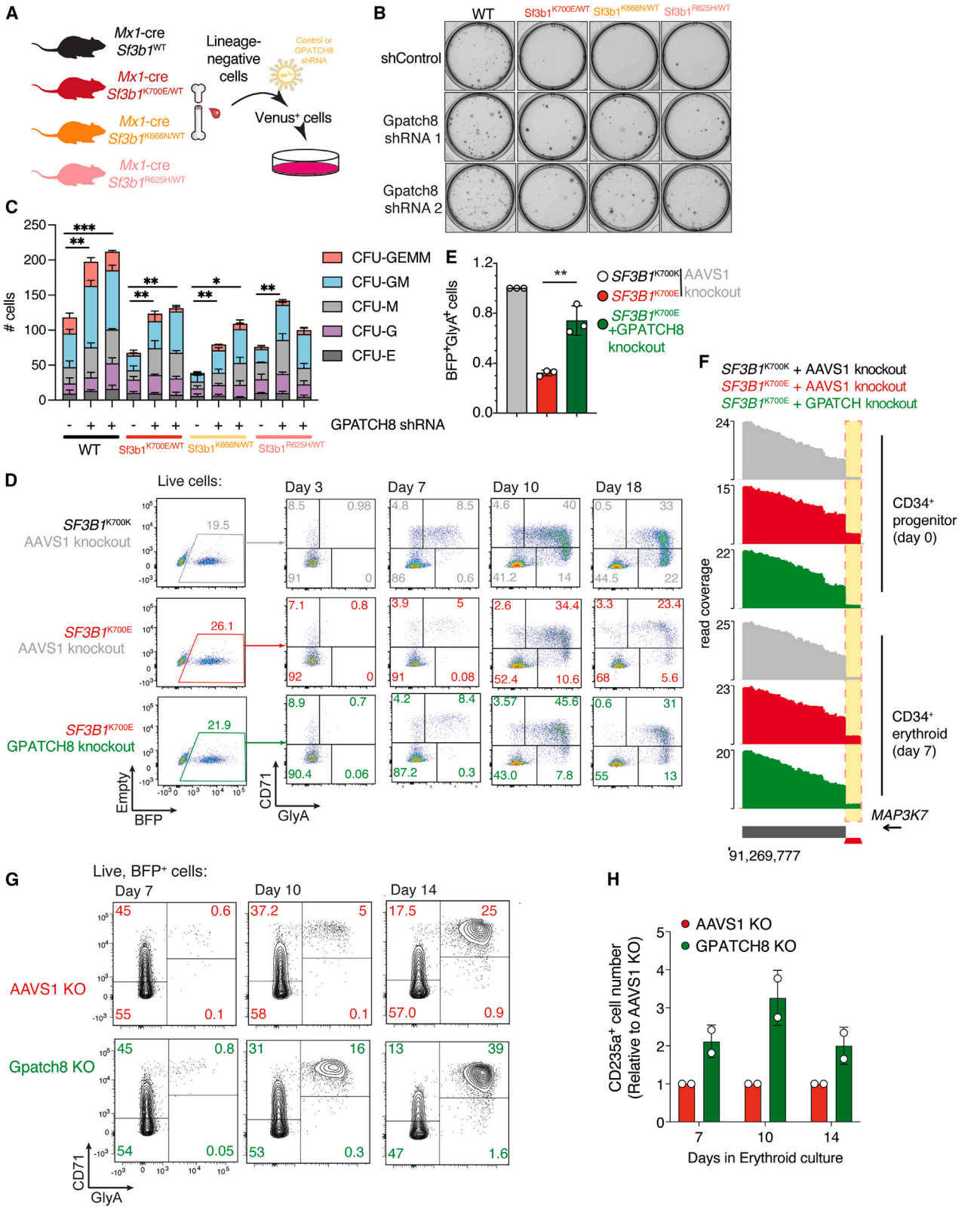
We expressed (1) DHX15 with an AP tag and (2) EX- and HA-tagged SUGP1 in 293T cells and then assessed for biotinylation in the presence versus absence of GPATCH8. This experiment revealed greatly enhanced biotin signal upon GPATCH8 deletion compared with control sgRNA (Figures 6F, 6G, and S6C). This result thereby indicates that GPATCH8 loss promotes SUGP1 and DHX15 interaction as readout by reconstitution of the split APEX signal.

Suppression of GPATCH8 rescues hematopoietic defects of SF3B1 mutant hematopoietic cells

There have been no prior attempts to correct mis-splicing of a larger number of *SF3B1* mutation-induced mis-splicing events in parallel, as therapeutic means to achieve this do not exist. Of note, GPATCH8 mRNA expression does not differ in MDS patients based on SF3B1 mutant genotype (Figure S6D). Given our discovery of GPATCH8 as required for a large fraction of *SF3B1*-mutant mis-splicing events, we next sought to evaluate the impact of GPATCH8 KO on *SF3B1*-mutant hematopoiesis. To achieve this, we generated new knockin mice with conditional expression of the *Sf3b1*^{K666N} and *Sf3b1*^{R625H} alleles to compliment previously generated *Sf3b1*^{K700E} conditional knockin mice[39] (Figures S7A–S7F). Upon Cre-recombination, these animals express each *Sf3b1* mutation at the expected ~50% allele frequency from the endogenous *Sf3b1* locus. Lineage-negative bone marrow (BM) hematopoietic cells from each model exhibited greatly impaired myeloid differentiation and clonogenic capacity in *in vitro* methylcellulose colony assays as well as the same Map3k7 mis-splicing as human cells (Figures 7A, 7B, and S7G). Importantly, however, silencing of Gpatch8 in these same cells prior to *in vitro* plating significantly rescued colony formation in each mutant-*SF3B1* genotype as well as Map3k7 mis-splicing (Figures 7B, 7C, and S7G). Of note, Gpatch8 silencing was well tolerated in WT mouse BM cells despite ~80% reduction in Gpatch8 expression.

We complimented the above murine studies with evaluation of the SF3B1 mutation and GPATCH8 deletion in human hematopoietic stem and progenitor cells (HSPCs). This was achieved by CRISPR editing with AAV6 homology-directed repair in human cord blood CD34$^+$ cells to introduce the SF3B1^{K700E} (or the control SF3B1^{K700K}) heterozygous mutation and a BFP fluorescent reporter at the endogenous *SF3B1* allele. This was combined with the deletion of either GPATCH8 or a non-targeting AAVS1 control in the same cells. A mean of 16% SF3B1 precise editing efficiency marked by BFP$^+$ and 58% GPATCH8 KO efficiency was achieved across CD34$^+$ cells from three independent cord blood donors using this approach. Edited HSPCs were then differentiated into erythroid cells

(F) Western blot of biotin and HA (left) along with Ponceau S loading control (right) in the split APEX experiment from (E). HA-EX-tagged FRB was used as negative control and "/" represents an additional negative control empty cDNA.
(G) Quantification of biotin signal normalized to respective conditions without HA-EX-tagged protein expression. Mean ± SD shown. Unpaired two-sided student t test.
See also Figure S6.



(legend on next page)

following an 18-day liquid culture (Figures S7H and S7I). A defect in erythroid differentiation was evident in the SF3B1 mutant cells, consistent with the marked anemia in patients with SF3B1 mutant MDS (Figures 7D and 7E). GPATCH8 KO partially corrected the erythroid defect, with the most pronounced rescue at the $CD71^{+}GlyA^{+}$ erythroblast stage in early erythropoiesis (Figures 7D and 7E). RNA-seq analysis of primary human $CD34^{+}$ cells immediately after editing (day 0) as well differentiated erythroid cells at day 7 of culture of these cells revealed that the key mis-splicing event induced by mutant SF3B1 in MAP3K7 was corrected by GPATCH8 deletion in this system (Figure 7F).

Finally, we had the opportunity to test the impact of GPATCH8 deletion on erythropoiesis of two distinct AML patient samples with somatic SF3B1 mutations (an $SF3B1^{K666N/WT}$ patient sample with SF3B1 mutant allele frequency of 42%, and an $SF3B1^{K700E/WT}$ patient sample with mutant allele frequency of 40%). We achieved a GPATCH8 KO efficiency of 35%–37% (as identified in purified $CD235a^{+}$ cells following 15 days in erythroid culture). Importantly, this level of GPATCH8 deletion greatly enhanced mature $CD235a^{+}$ erythroid cells in each patient, which we confirmed were derived from SF3B1 mutant leukemic cells (Figures 7G, 7H, and S7J). Altogether, these data indicate that correcting SF3B1 mutant mis-splicing via GPATCH8 deletion ameliorates aberrant hematopoiesis associated with SF3B1-mutant myeloid neoplasms.

DISCUSSION

Here, we utilized synthetic introns uniquely responsive to cancer-associated mutations in *SF3B1* to perform unbiased, positive enrichment, whole-genome CRISPR screens. This revealed GPATCH8 as required for approximately one-third of the several thousand aberrant RNA splicing events induced by mutant SF3B1. The discovery that GPATCH8 is required for the molecular and phenotypic sequelae of *SF3B1* mutations was particularly interesting given recent identification that the related protein SUGP1 is also important in branchpoint recognition.[15,18,21,22] One possible explanation for the requirement of GPATCH8 for *SF3B1* mutation-dependent mis-splicing could be that mutations in *SF3B1* promote physical interaction with GPATCH8. However, we did not identify preferential interaction of U2 snRNP or U2AF proteins with GPATCH8 in *SF3B1*-mutant cells.

SUGP1 has been shown across several studies to physically link U2 snRNP to the RNA helicase DHX15 via SUGP1's G-patch domain.[18,21,22] We identified that GPATCH8 also physically interacts with DHX15. Recruitment of DHX15 to U2 snRNP is critical for branchpoint recognition, as the ATPase helicase function of DHX15 displaces SF1 and may allow U2 snRNP to identify correct BPs. Consistent with this, Zhang et al. identified that expression of a mutant form of SUGP1, which impairs DHX15 interaction phenocopies certain aberrant splicing events induced by mutant SF3B1.[18] At the same time, reconstitution of GPATCH8-null cells with the G-patch-domain-containing N terminus of GPATCH8 alone is sufficient to recapitulate the aberrant splicing seen with mutant SF3B1. These data suggest that GPATCH8 and SUGP1 might compete for interaction with DHX15, which could explain the functional antagonism between SUGP1 and GPATCH8 in regulating *SF3B1* mutation-dependent mis-splicing. This hypothesis was confirmed with the finding that loss of GPATCH8 promotes the interaction of SUGP1 and DHX15.

Currently, the basis for allele-specific associations of individual *SF3B1* mutations and specific cancers and cancer outcomes is unknown. Nonetheless, GPATCH8 deletion was required for mis-splicing induced by each of these most commonly encountered mutations in *SF3B1*.

A number of prior studies have tested the impact of correcting individual mis-splicing events on phenotypic consequences of the SF3B1 mutation. Here, deleting GPATCH8 allowed us to correct a multitude of SF3B1 mutant mis-splicing events in parallel. In doing so, GPATCH8 deletion partially rescued the impaired hematopoietic growth and erythroid differentiation characteristic of SF3B1 mutations in mouse and human hematopoietic precursor cells. Ineffective erythropoiesis is the hallmark of SF3B1 mutant MDS,[2,40] and it was recapitulated by SF3B1-edited primary human HSPCs. GPATCH8 deletion corrected the erythroid differentiation defect, and this effect was most pronounced in early erythropoiesis, consistent with the role of MAP3K7 mis-splicing in regulating erythroid master transcription factor GATA1.[27] It is important to note that GPATCH8 is not a pan-essential gene across cancer cell lines,[41] and deletion of GPATCH8 did not abrogate growth of human or mouse hematopoietic precursors *in vitro*. Overall, these findings suggest that pharmacologic approaches to disable GPATCH8, or its interaction with DHX15, may serve as important therapeutic approaches to correct molecular and biological defects of SF3B1-mutant cancers.

Figure 7. Suppression of GPATCH8 rescues hematopoietic defects of SF3B1 mutant hematopoietic cells

(A) Schema of anti-Gpatch8 shRNA experiments in hematopoietic precursor cells from Sf3b1 wild-type (WT) and mutant mice.

(B) Photograph of methylcellulose colonies generated from hematopoietic precursor cells of Sf3b1 WT and mutant mice with or without suppression of Gpatch8.

(C) Enumeration of numbers and types of hematopoietic colonies from (B). Measurements are from three independent experiments (mean ± SD, unpaired two-sided student t test, * $p < 0.0332$, ** $p < 0.0021$, and *** $p < 0.0002$).

(D) FACS plot of live, BFP^{+} cells followed by CD71/glycophorin A (GlyA) FACS staining of one representative $CD34^{+}$ cord blood donor treated with sgRNAs to create the $SF3B1^{K700E/WT}$ mutation or $SF3B1^{K700K}$ control along with knockout of GPATCH8 (or AAVS1 negative control).

(E) Absolute numbers of BFP^{+} $GlyA^{+}$ cells at day 7 following erythroid culture from three separate donors. Mean ± SD. Two-way ANOVA.

(F) RNA-seq coverage plots of $CD34^{+}$ cells from (D) at days 0 and 7 of erythroid culture revealing MAP3K7 aberrant 3′ splice site event.

(G) FACS staining of live, BFP^{+} hematopoietic cells from an $SF3B1^{K666N/WT}$ acute myeloid leukemia patient treated with knockout of GPATCH8 (or AAVS1 negative control) and grown in erythroid culture conditions over the indicated time period.

(H) Relative number of mature $CD235a^{+}$ erythroid cells in GPATCH8 knockout of primary SF3B1 mutant myeloid leukemia patient samples relative to AAVS1 KO across two distinct patients.

See also Figure S7.

Limitations of the study

GPATCH8 regulates only approximately one-third of SF3B1 mutant aberrant splicing events. These data suggest that there may be other factors critical for *trans* regulation of mutant SF3B1 splicing activity yet to be characterized. Additionally, it is possible that GPATCH8 may act at other steps in RNA processing beyond 3′ss selection. For example, several RNA splicing factors that act in later steps of splicing catalysis enriched as GPATCH8-interacting proteins, and these interactions will be important to study in future work. Finally, the fact that GPATCH8 appears to be dispensable in SF3B1 WT cells despite exerting an impact on SF3B1 mutant cells begs the question of the normal physiologic role of GPATCH8 in cell differentiation and proliferation. Future biological studies aimed at characterizing the requirement of GPATCH8 in cell differentiation *in vivo* will be critical to further elucidate the function of GPATCH8.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
  - Mice
- METHOD DETAILS
  - Isolation and gene editing of human cord blood and AML bone marrow HSPCs
  - Human HSPC erythroid differentiation
  - Vector cloning, lentiviral production, and cell transduction
  - 3X-FLAG-tag knockin into endogenous *GPATCH8*
  - Whole-genome CRISPR-Cas9 Screening
  - RT-PCR and quantitative RT-PCR
  - Cell viability and proliferation measurements
  - Flow cytometry
  - Split APEX proximity labeling
  - Western blotting
  - RNA-sequencing library preparation and sequencing
  - RNA-seq data analysis
  - Splice site sequence analyses
  - Colony-forming assays
  - GPATCH8 and SUGP1 TRIBE-seq
  - Analysis of TRIBE-seq data
  - CRISPR tiling screen
  - Alphafold2 and ColabFold modeling of protein-protein interactions
  - Immunoprecipitation for mass spectrometry
  - Protein digestion for proteomic analyses
  - Mass spectrometry analyses
  - DIA data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.04.006.

ACKNOWLEDGMENTS

S.B. was supported by the Lauri Strauss Leukemia Foundation and the Aplastic Anemia and MDS International Foundation. O.A.-W. and R.K.B. were supported by the Edward P. Evans Foundation, NIH/National Cancer Institute (NCI) (R01 CA251138 and R01 CA283364), and NIH/National Heart Lung and Blood Institute (NHLBI) (R01 HL128239). O.A.-W. was supported by the NIH/NCI (R01 CA242020 and P50 CA254838-01) and the Leukemia & Lymphoma Society. R.K.B. was supported by the NIH/NHLBI (R01 HL151651) and the Blood Cancer Discoveries Grant program through the Leukemia & Lymphoma Society, the Mark Foundation for Cancer Research, and the Paul G. Allen Frontiers Group (8,023–20). R.K.B. is a Scholar of the Leukemia & Lymphoma Society (1,344–18) and holds the McIlwain Family Endowed Chair in Data Science. S.D. is supported by the NIH/NHLBI (R01 HL151651 and R01 HL169156), the Kuni Foundation, and Edward P. Evans Foundation. S.D. is a Scholar of the Leukemia and Lymphoma Society (1,391–24). Computational studies were supported in part by FHCRC's Scientific Computing Infrastructure (ORIP S10 OD028685). We acknowledge the use of the Integrated Genomics Operation Core, supported by the NCI Cancer Center Support Grant (CCSG; P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

AUTHOR CONTRIBUTIONS

Conceptualization, S.B., B.L., R.K.B., and O.A.-W.; CRISPR screens and validation, S.B., B.L., E.W., Q.Z., S.J.H., K.L., A.D., Y.A., S.M., R.G., and R.F.S.; RNA-seq analyses, J.M.B.P., J.B., and R.K.B.; TRIBE-seq analyses, J.B.; mass spectrometry, J.O.-P., Z.L., and M.M.; primary human cell experiments, L.B.G., M.S., and S.D.; mouse model experiments, S.B., A.M.L., and N.F.; bioinformatics, data analyses, and visualization, J.M.B.P., J.B., J.R., and R.K.B.; resources, S.D., R.K.B., and O.A.-W.; writing – review & editing, all authors; supervision, S.D., R.K.B., and O.A.-W.; project administration, S.B., B.L., R.K.B., and O.A.-W.; funding acquisition, S.D., R.K.B., and O.A.-W.

DECLARATION OF INTERESTS

R.K.B. and O.A.-W. are founders and scientific advisors of Codify Therapeutics, hold equity in this company, and receive research funding from this company. S.B., J.M.B.P., R.K.B., and O.A.-W. are named inventors on a patent related to this study. R.K.B. is a founder and scientific advisor of Synthesize Bio and holds equity in this company. O.A.-W. has served as a consultant for Foundation Medicine Inc., Merck, Prelude Therapeutics, Amphista Therapeutics, MagnetBio, and Janssen and is on the Scientific Advisory Board of Envisagenics Inc., Harmonic Discovery Inc., and Pfizer Boulder. O.A.-W. has received prior research funding from H3B Biomedicine, Nurix Therapeutics, Minovia Therapeutics, and LOXO Oncology unrelated to the current manuscript.

Received: September 26, 2023
Revised: March 4, 2024
Accepted: April 9, 2024
Published: April 29, 2024

REFERENCES

1. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69. nature10496 (pii). https://doi.org/10.1038/nature10496.
2. Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med. *365*, 1384–1395. https://doi.org/10.1056/NEJMoa1103283.
3. Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D.S., Zhang, L., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N. Engl. J. Med. *365*, 2497–2506. https://doi.org/10.1056/NEJMoa1109016.
4. Quesada, V., Conde, L., Villamor, N., Ordóñez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Beà, S., Pinyol, M., Martínez-Trillos, A., et al. (2011). Exome sequencing identifies recurrent mutations of the

splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat. Genet. *44*, 47–52. https://doi.org/10.1038/ng.1032.

5. Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 10615. https://doi.org/10.1038/ncomms10615.

6. Furney, S.J., Pedersen, M., Gentien, D., Dumont, A.G., Rapinat, A., Desjardins, L., Turajlic, S., Piperno-Neumann, S., de la Grange, P., Roman-Roman, S., et al. (2013). SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov. *3*, 1122–1129. https://doi.org/10.1158/2159-8290.CD-13-0330.

7. Liu, B., Liu, Z., Chen, S., Ki, M., Erickson, C., Reis-Filho, J.S., Durham, B.H., Chang, Q., de Stanchina, E., Sun, Y., et al. (2021). Mutant SF3B1 promotes AKT- and NF-kappaB-driven mammary tumorigenesis. J. Clin. Invest. *131*. https://doi.org/10.1172/JCI138315.

8. Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–1120. https://doi.org/10.1016/j.cell.2012.08.029.

9. Rahman, M.A., Krainer, A.R., and Abdel-Wahab, O. (2020). SnapShot: splicing alterations in cancer. Cell *180*, 208–208.e1. https://doi.org/10.1016/j.cell.2019.12.011.

10. Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al. (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3? Splice Site Selection through Use of a Different Branch Point. Cell Rep. *13*, 1033–1045. https://doi.org/10.1016/j.celrep.2015.09.053.

11. Inoue, D., Chew, G.L., Liu, B., Michel, B.C., Pangallo, J., D'Avino, A.R., Hitchman, T., North, K., Lee, S.C., Bitner, L., et al. (2019). Spliceosomal disruption of the non-canonical BAF complex in cancer. Nature *574*, 432–436. https://doi.org/10.1038/s41586-019-1646-9.

12. Wahl, M.C., and Lührmann, R. (2015). SnapShot: Spliceosome Dynamics II. Cell *162*, 456–456.e1. https://doi.org/10.1016/j.cell.2015.06.061.

13. Wahl, M.C., and Lührmann, R. (2015). SnapShot: Spliceosome Dynamics I. Cell *161*, 1474–14e1. https://doi.org/10.1016/j.cell.2015.05.050.

14. Yang, F., Bian, T., Zhan, X., Chen, Z., Xing, Z., Larsen, N.A., Zhang, X., and Shi, Y. (2023). Mechanisms of the RNA helicases DDX42 and DDX46 in human U2 snRNP assembly. Nat. Commun. *14*, 897. https://doi.org/10.1038/s41467-023-36489-x.

15. Zhang, J., Ali, A.M., Lieu, Y.K., Liu, Z., Gao, J., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2019). Disease-Causing Mutations in SF3B1 Alter Splicing by Disrupting Interaction with SUGP1. Mol. Cell *76*, 82–95.e7. https://doi.org/10.1016/j.molcel.2019.07.017.

16. Tang, Q., Rodriguez-Santiago, S., Wang, J., Pu, J., Yuste, A., Gupta, V., Moldón, A., Xu, Y.Z., and Query, C.C. (2016). SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev. *30*, 2710–2723. https://doi.org/10.1101/gad.291872.116.

17. Zhao, B., Li, Z., Qian, R., Liu, G., Fan, M., Liang, Z., Hu, X., and Wan, Y. (2022). Cancer-associated mutations in SF3B1 disrupt the interaction between SF3B1 and DDX42. J. Biochem. *172*, 117–126. https://doi.org/10.1093/jb/mvac049.

18. Zhang, J., Huang, J., Xu, K., Xing, P., Huang, Y., Liu, Z., Tong, L., and Manley, J.L. (2022). DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer. Proc. Natl. Acad. Sci. USA *119*, e2216712119. https://doi.org/10.1073/pnas.2216712119.

19. Liu, Z., Zhang, J., Sun, Y., Perea-Chamblee, T.E., Manley, J.L., and Rabadan, R. (2020). Pan-cancer analysis identifies mutations in SUGP1 that recapitulate mutant SF3B1 splicing dysregulation. Proc. Natl. Acad. Sci. USA *117*, 10305–10312. https://doi.org/10.1073/pnas.1922622117.

20. Alsafadi, S., Dayot, S., Tarin, M., Houy, A., Bellanger, D., Cornella, M., Wassef, M., Waterfall, J.J., Lehnert, E., Roman-Roman, S., et al. (2021). Genetic alterations of SUGP1 mimic mutant-SF3B1 splice pattern in lung adenocarcinoma and other cancers. Oncogene *40*, 85–96. https://doi.org/10.1038/s41388-020-01507-5.

21. Beusch, I., Rao, B., Studer, M.K., Luhovska, T., Šukytė, V., Lei, S., Oses-Prieto, J., SeGraves, E., Burlingame, A., Jonas, S., and Madhani, H.D. (2023). Targeted high-throughput mutagenesis of the human spliceosome reveals its in vivo operating principles. Mol. Cell *83*, 2578–2594.e9. https://doi.org/10.1016/j.molcel.2023.06.003.

22. Feng, Q., Krick, K., Chu, J., and Burge, C.B. (2022). Splicing quality control mediated by DHX15 and its G-patch activator, SUGP1. Cell Rep. *42*, 113223. https://doi.org/10.1016/j.celrep.2023.113223.

23. Studer, M.K., Ivanović, L., Weber, M.E., Marti, S., and Jonas, S. (2020). Structural basis for DEAH-helicase activation by G-patch proteins. Proc. Natl. Acad. Sci. USA *117*, 7159–7170. https://doi.org/10.1073/pnas.1913880117.

24. Bohnsack, K.E., Ficner, R., Bohnsack, M.T., and Jonas, S. (2021). Regulation of DEAH-box RNA helicases by G-patch proteins. Biol. Chem. *402*, 561–579. https://doi.org/10.1515/hsz-2020-0338.

25. North, K., Benbarche, S., Liu, B., Pangallo, J., Chen, S., Stahl, M., Bewersdorf, J.P., Stanley, R.F., Erickson, C., Cho, H., et al. (2022). Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. Nat. Biotechnol. *40*, 1103–1113. https://doi.org/10.1038/s41587-022-01224-2.

26. Lee, S.C., North, K., Kim, E., Jang, E., Obeng, E., Lu, S.X., Liu, B., Inoue, D., Yoshimi, A., Ki, M., et al. (2018). Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell *34*, 225–241.e8. https://doi.org/10.1016/j.ccell.2018.07.003.

27. Lieu, Y.K., Liu, Z., Ali, A.M., Wei, X., Penson, A., Zhang, J., An, X., Rabadan, R., Raza, A., Manley, J.L., and Mukherjee, S. (2022). SF3B1 mutant-induced missplicing of MAP3K7 causes anemia in myelodysplastic syndromes. Proc. Natl. Acad. Sci. USA *119*, e2111703119. https://doi.org/10.1073/pnas.2111703119.

28. Dolatshad, H., Pellagatti, A., Fernandez-Mercado, M., Yip, B.H., Malcovati, L., Attwood, M., Przychodzen, B., Sahgal, N., Kanapin, A.A., Lockstone, H., et al. (2015). Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia *29*, 1092–1103. https://doi.org/10.1038/leu.2014.331.

29. Pineda, J.M.B., and Bradley, R.K. (2018). Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev. *32*, 577–591. https://doi.org/10.1101/gad.312058.118.

30. Thanaraj, T.A., and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res. *29*, 2581–2593. https://doi.org/10.1093/nar/29.12.2581.

31. Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res. *36*, 2257–2267. https://doi.org/10.1093/nar/gkn073.

32. Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394. https://doi.org/10.1089/1066527041410418.

33. Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell *27*, 617–630. https://doi.org/10.1016/j.ccell.2015.04.006.

34. Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. *25*, 14–26. https://doi.org/10.1101/gr.181016.114.

35. Xu, W., Rahman, R., and Rosbash, M. (2018). Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. RNA *24*, 173–182. https://doi.org/10.1261/rna.064691.117.

36. Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat. Biotechnol. *33*, 661–667. https://doi.org/10.1038/nbt.3235.

37. Feng, Q., Krick, K., Chu, J., and Burge, C.B. (2023). Splicing quality control mediated by DHX15 and its G-patch activator SUGP1. Cell Rep. *42*, 113223. https://doi.org/10.1016/j.celrep.2023.113223.

38. Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K., and Ting, A.Y. (2015). Directed evolution of APEX2 for electron microscopy and proximity labeling. Nat. Methods *12*, 51–54. https://doi.org/10.1038/nmeth.3179.

39. Obeng, E., McConkey, M.E., Campagna, D., Schneider, R.K., Chen, M.C., Schmidt, P.J., Mullally, A., Fleming, M.D., and Ebert, B.L. (2014). Mutant Splicing Factor 3b Subunit 1 (SF3B1) Causes Dysregulated Erythropoiesis and a Stem Cell Disadvantage Blood. (ASH Abstract 828; 56th Annual American Society of Hematology meeting)*124*.

40. Malcovati, L., Karimi, M., Papaemmanuil, E., Ambaglio, I., Jädersten, M., Jansson, M., Elena, C., Gallì, A., Walldin, G., Della Porta, M.G., et al. (2015). SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. Blood *126*, 233–241. https://doi.org/10.1182/blood-2015-03-633537.

41. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607. https://doi.org/10.1038/nature11003.

42. Griewank, K.G., Yu, X., Khalili, J., Sozen, M.M., Stempke-Hale, K., Bernatchez, C., Wardell, S., Bastian, B.C., and Woodman, S.E. (2012). Genetic and molecular characterization of uveal melanoma cell lines. Pigment Cell Melanoma Res. *25*, 182–187. https://doi.org/10.1111/j.1755-148X.2012.00971.x.

43. Gee, P., Lung, M.S.Y., Okuzaki, Y., Sasakawa, N., Iguchi, T., Makita, Y., Hozumi, H., Miura, Y., Yang, L.F., Iwasaki, M., et al. (2020). Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. Nat. Commun. *11*, 1334. https://doi.org/10.1038/s41467-020-14957-y.

44. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784. https://doi.org/10.1038/nmeth.3047.

45. Chan, W., Gottschalk, R.A., Yao, Y., Pomerantz, J.L., and Germain, R.N. (2021). Efficient Immune Cell Genome Engineering with Enhanced CRISPR Editing Tools. Immunohorizons *5*, 117–132. https://doi.org/10.4049/immunohorizons.2000082.

46. Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A., Kwon, J.S., Guan, Y., et al. (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell Rep. *5*, 1704–1713. https://doi.org/10.1016/j.celrep.2013.11.020.

47. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191. https://doi.org/10.1038/nbt.3437.

48. Biswas, J., Rahman, R., Gupta, V., Rosbash, M., and Singer, R.H. (2020). MS2-TRIBE Evaluates Both Protein-RNA Interactions and Nuclear Organization of Transcription by RNA Editing. iScience *23*, 101318. https://doi.org/10.1016/j.isci.2020.101318.

49. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

50. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

51. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 3. https://doi.org/10.14806/ej.17.1.200.

52. Biswas, J., Rosbash, M., Singer, R.H., and Rahman, R. (2021). Protocol for using TRIBE to study RNA-protein interactions and nuclear organization in mammalian cells. Star Protoc. *2*, 100634. https://doi.org/10.1016/j.xpro.2021.100634.

53. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

54. Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. *39*, D19–D21. https://doi.org/10.1093/nar/gkq1019.

55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

56. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

57. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25. https://doi.org/10.1186/gb-2010-11-3-r25.

58. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120.

59. Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods *7*, 1009–1015. https://doi.org/10.1038/nmeth.1528.

60. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software *4*, 1686. https://doi.org/10.21105/joss.01686.

61. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.

62. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLOS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

63. Bembom, O. (2007). seqLogo: Sequence logos for DNA sequence alignments. R package version 1.30.0.

64. Obeng, E.A., Chappell, R.J., Seiler, M., Chen, M.C., Campagna, D.R., Schmidt, P.J., Schneider, R.K., Lord, A.M., Wang, L., Gambe, R.G., et al. (2016). Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell *30*, 404–417. https://doi.org/10.1016/j.ccell.2016.08.006.

65. Sarchi, M., Clough, C.A., Gallì, A., Todisco, G., Creamer, J.P., Malcovati, L., and Doulatov, S. (2022). Clonal Trajectories and Therapeutic Targeting of High-Risk SF3B1-Mutant Myelodysplastic Syndromes. Blood *140* (*Supplement 1*), 211–212. https://doi.org/10.1182/blood-2022-164833.

66. Clough, C.A., Pangallo, J., Sarchi, M., Ilagan, J.O., North, K., Bergantinos, R., Stolla, M.C., Naru, J., Nugent, P., Kim, E., et al. (2022). Coordinated missplicing of TMEM14C and ABCB7 causes ring sideroblast formation in SF3B1-mutant myelodysplastic syndrome. Blood *139*, 2038–2049. https://doi.org/10.1182/blood.2021012652.

67. Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. *42*, D764–D770. https://doi.org/10.1093/nar/gkt1168.

68. Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., et al. (2013). Ensembl 2013. Nucleic Acids Res. *41*, D48–D55. https://doi.org/10.1093/nar/gks1236.

69. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

70. Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121. https://doi.org/10.1038/nmeth.3252.

71. Liu, Q., and Gregory, R.I. (2019). RNAmod: an integrated system for the annotation of mRNA modifications. Nucleic Acids Res. *47*, W548–W555. https://doi.org/10.1093/nar/gkz479.

72. Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601. https://doi.org/10.1073/pnas.1419161111.

73. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MAP3K7	Cell Signaling Technology	Cat# 5206; RRID: AB_10694079
β-actin	Cell Signaling Technology	Cat# 4970; RRID: AB_2223172
IRDye 800CW Goat anti-Rabbit IgG	LI-COR	Cat# 926-32211; RRID: AB_2925189
Monoclonal ANTI-FLAG M2 antibody	Sigma-Aldrich	Cat# F3165; RRID: AB_259529
Mouse IgG2a Isotype Control	Sigma-Aldrich	Cat# M5409; RRID: AB_1163691
DHX15 Polyclonal antibody (for IP-MS)	Bethyl Laboratories	Cat# A300-390A; RRID: AB_2091992
DHX15 Polyclonal antibody (for WB)	Bethyl Laboratories	Cat# A300-389A; RRID: AB_386100
Rabbit IgG Isotype Control	Thermo Fisher Scientific	Cat# 02-6102; RRID: AB_2532938
IRDye 800CW Streptavidin	LI-COR	Cat# 926-32230; RRID: N/A
HA-Tag (C29F4) Rabbit monoclonal antibody	Cell Signaling Technology	Cat# 3724; RRID: AB_1549585
IRDye 680RD Goat anti-Rabbit IgG Secondary Antibody	LI-COR	Cat# 926-68071; RRID: AB_10956166
Anti-rabbit IgG HRP-linked Antibody	Cell Signaling Technology	Cat# 7074; RRID: AB_2099233
Biological samples		
Cord blood $CD34^+$ cells	BloodWorks Northwest	N/A
Bone marrow cells from MDS and AML patients	The Hematologic Oncology Tissue Bank of Memorial Sloan Kettering Cancer Center	N/A
Chemicals, peptides, and recombinant proteins		
Polyethylenimine (PEI)	Polysciences	24765
Polybrene	Sigma Aldrich	TR-1003-G
Recombinant Murine SCF (mCSF)	PeproTech	250-03
Recombinant Murine Flt3-Ligand (mFLT3L)	PeproTech	250-31L
Recombinant Murine TPO (mTPO)	PeproTech	315-14
4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI)	Invitrogen	D1306
Doxycycline	MilliporeSigma	D9891
Biotin Phenol	MilliporeSigma	SML2135
H2O2	MilliporeSigma	H1009
Sodium azide	MilliporeSigma	26628-22-8
Sodium L-ascorbate	MilliporeSigma	A7631-25G
Trolox(R), 97%	Thermo Fisher Scientific	218940010
Lipofectamine 3000	Invitrogen	L3000008
Critical commercial assays		
RNeasy Plus Mini Kit	Qiagen	74136
Verso cDNA synthesis kit	ThermoFisher Scientific	AB1453B
MethoCult GF M3434	StemCell Technologies	03434
Murine CD117 MicroBeads	Miltenyi Biotech	130-091-224
Qubit Protein Assay	Invitrogen	Q33212
CellTiter-Glo® Luminescent Cell Viability Assay	Promega	G7572
Deposited data		
FLAG-GPATCH8 immunoprecipitation mass spectrometry data	ProteomeXchange	PXD044973
DHX15 immunoprecipitation mass spectrometry data	ProteomeXchange	PXD044973
RNA-seq and TRIBE-seq data	GEO repository	GSE242094

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Human: 293T	ATCC	CRL-3216
Human: MCF10A	ATCC	CRL-10317
Human: K562	ATCC	CCL-243
Human: Panc05.04	ATCC	CRL-2557
Human: MEL202	Obtained from B. Bastian[42]	N/A
Human: MEL270	Obtained from B. Bastian[42]	N/A
Experimental models: Organisms/strains		
Mx-1 cre transgenic mice	The Jackson Laboratory	Strain #:003556
Sf3b1 K700E conditional knockin mice	The Jackson Laboratory	Strain #:029766
Sf3b1 K666N conditional knockin mice	This paper	N/A
Sf3b1 R625H conditional knockin mice	This paper	N/A
Oligonucleotides		
mCardinal-P2A cloning forward primer: CTTTCGACC TGCAGCCCAAGGCCGCCATGGTGAGCAAGGGCGAG	This paper	N/A
mCardinal-P2A cloning reverse primer: CCAGCCTGCTTCAGCAGAGAGAAGTTGGTGG CTCCGCTTCCCTTGTACAGCTCGTCCATG	This paper	N/A
P2A-mEmerald-synMAP3K7i250 cloning forward primer: CTGCTGAAGCAGGCTGGTGACGTGGAGGAGA ATCCCGGCCCTATGGTGAGCAAGGGCGAG	This paper	N/A
P2A-mEmerald-synMAP3K7i250 cloning reverse primer: CAAATTTTGTAATCCAGAGGTTGATTACTTGTA CAGCTCGTCCATG	This paper	N/A
GPATCH8 sgRNA #1: ATGCTGAAGATGCTACCGAA	This paper	N/A
GPATCH8 sgRNA #2: CATGTTCAAACCAACCACAG	This paper	N/A
GPATCH8 sgRNA #3: GGTCTGAGACAGAAGACACA	This paper	N/A
SUGP1 sgRNA #1: CTATGAGAAGGGGAAGCCTG	This paper	N/A
SUGP1 sgRNA #2: AACTTTTCCAGCTCGTCTGG	This paper	N/A
Renilla sgRNA: GGAACACGGCCGTATTAGGG	This paper	N/A
Gpatch8 shRNA #1: TGCTGTTGACAGTGAGCGATA CAAGGATTATGTTGACAAATAGTGAAGCCACAGAT GTATTTGTCAACATAATCCTTGTACTGCCTACTGCCTCGGA	This paper	N/A
Gpatch8 shRNA #2: TGCTGTTGACAGTGAGCG ATCCTACAGTTGTAATCCTTTATAGTGAAGCCAC AGATGTATAAAGGATTACAACTGTAGGAGT GCCTACTGCCTCGGA	This paper	N/A
NTC shRNA: TGCTGTTGACAGTGAGCGCTTAAATAACT ACTGACGTCCGTAGTGAAGCCACAGATGTACGGACGT CAGTAGTTATTTAATTGCCTACTGCCTCGGA	This paper	N/A
GPATCH8 (human) forward primer: tggaagctgggccagggatt	This paper	N/A
GPATCH8 (human) reverse primer: gcgccgttcggtagcatctt	This paper	N/A
Gpatch8 (murine) forward primer: TGGAAGCTGGGCCAGGGATT	This paper	N/A
Gpatch8 (murine) reverse primer: GACGCCGTTCGGTAGCATCC	This paper	N/A
SUGP1 forward primer: GTGCAGAGGGACGTGGATGC	This paper	N/A
SUGP1 reverse primer: GCCCACTAGACCCACAGGCT	This paper	N/A
GAPDH (human) forward primer: CTTTTGCGTCGCCAGCCGAG	This paper	N/A
GAPDH (human) reverse primer: CCAGGCGCCCAATACGACCA	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gapdh (murine) forward primer: AGGTCGGTGTGAACGGATTTG	This paper	N/A
Gapdh (murine) reverse primer: TGTAGACCATGTAGTTGAGGTCA	This paper	N/A
mEmerald forward primer: CACATGAAGCAGCACGACTTC	This paper	N/A
mEmerald reverse primer: GTCCTCCTTGAAGTCGATG	This paper	N/A
MAP3K7 (human) forward primer: TGTCTTGTGATGGAATATGCTG	This paper	N/A
MAP3K7 (human) reverse primer: TCCCTGTGAATTAGCGCTTT	This paper	N/A
Map3k7 (murine) forward primer: GTTGTCACGTGTGAACCATC	This paper	N/A
Map3k7 (murine) reverse primer: CATGAGCAGCAGTGTAGTAAG	This paper	N/A
SMURF2 forward primer: AGCCCTGGCAGACCTCTTAGCT and rev reverse primer: ACCTGGCCTTGTTGCGTTGTCC	This paper	N/A
SMURF2 reverse primer: ACCTGGCCTTGTTGCGTTGTCC	This paper	N/A
PPM1M forward primer: TGTCCAACGAGCAGGTGGCATG	This paper	N/A
PPM1M reverse primer: CTTGGCCCTGACTGTGCAAGGG;	This paper	N/A
ZDHHC16 forward primer: CAGACCCCACCACCCACCTTCT	This paper	N/A
ZDHHC16 reverse primer: GCCTGTAGCCGACGTCTCTCCT	This paper	N/A
sgRNA for human SF3B1 K700E mutation: UGGAUGAGCAGCAGAAAGUU	This paper	N/A
sgRNA for editing AAVS1: GGGCCACUAGGGACAGGAU	This paper	N/A
sgRNA for GPATCH8 knockout in primary human cells: ATGCTGAAGATGCTACCGAA	This paper	N/A
mEmerald sgRNA #1: GGCGAGGGCGATGCCACCTA	This paper	N/A
mEmerald sgRNA #2: GGGCACGGGCAGCTTGCCGG	This paper	N/A
mEmerald sgRNA #3: GAGCTGGACGGCGACGTAAA	This paper	N/A
mCardinal sgRNA #1: GAGGCTGGTGTCCTGGGTAA	This paper	N/A
mCardinal sgRNA #2: GGGGAGGGCAAGCCCTACGA	This paper	N/A
mCardinal sgRNA #3: GGGGTACAGGGTCTCGGTGG	This paper	N/A
Recombinant DNA		
Plasmid: pVSVG	Gift from Akitsu Hotta[43]	Addgene: 138479
Plasmid: psPAX2	Gift from Didier Trono	Addgene: 12260
Plasmid: hPGK-mCardinal-P2A-mEmerald-synMAP3K7i250	This paper	N/A
Plasmid: mCardinal-pBiCMV-mEmerald-synMAP3K7i4-250	North et al.[25]	N/A
Plasmid: hPGK-PuroR-P2A-HSV-TK	North et al.[25]	N/A
Plasmid: LentiCas9-Blast	Gift from Feng Zhang[44]	Addgene: 52962
Plasmid: LentiGuide-Puro	Gift from Feng Zhang[44]	Addgene: 52963
Plasmid: LentiGuide-mTagBFP	Gift from Ronald Germain[45]	Addgene: 167929
Plasmid: SGEP	Gift from Johannes Zuber[46]	Addgene: 111170
Human pooled genome-wide sgRNA library (Brunello) in lentiGuide-Puro	Gift from David Root and John Doench[47]	Addgene: 73178
Plasmid: pUBC-stdMCP-serinemod-E488QADAR-p2A-yGFP	Gift from Robert Singer[48]	Addgene: 154787
Plasmid: pUBC-GPATCH8-serinemod-E488QADAR-p2A-yGFP	This paper	N/A
Plasmid: pUBC-serinemod-E488QADAR-SUGP1-p2A-yGFP	This paper	N/A
Plasmid: pUBC-serinemod-E488QADAR-stdMCP-p2A-yGFP	This paper	N/A
Plasmid: Codon optimized GPATCH8-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: Truncated Codon optimized GPATCH8-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Mutant Truncated Codon optimized GPATCH8-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Codon optimized SUGP1-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Mutant Codon optimized SUGP1-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Swap Truncated Codon optimized GPATCH8-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Swap Codon optimized SUGP1-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: Luciferase-Linker-NLS-3xFLAG-T2A-BFP_pGenlenti	This paper	GenScript
Plasmid: pRD-FLAG-AP-DHX15.WT	Feng et al.[37]	N/A
Plasmid: pX304-HA-SUGP1-EX.FL	Feng et al.[37]	N/A
Plasmid: pX304-HA-FRB-EX-NLS	Feng et al.[37]	N/A
Plasmid: pRRL_SRSF2_WT_mCherry	Kim et al.[33]	Addgene: 84020
Plasmid: pRRL_SRSF2_P95H_mCherry	Kim et al.[33]	Addgene: 84023
Plasmid: pRRL_U2AF1_WT_mCherry	Ilagan et al.[34]	Addgene: 84017
Plasmid: pRRL_U2AF1_S34F_mCherry	Ilagan et al.[34]	Addgene: 84019
Software and algorithms		
FlowJo V8.7	TreeStar (BD Biosciences)	https://www.flowjo.com/
Prism 9.0	GraphPad	https://www.graphpad.com
BEDTools	Quinlan and Hall[49]	https://github.com/arq5x/bedtools2
Bowtie	Langmead and Salzberg[50]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Cutadapt	Martin[51]	https://cutadapt.readthedocs.io/en/stable/installation.html
Fastqc	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
HyperTRIBE software	Biswas et al.[48,52]	https://github.com/rosbashlab/HyperTRIBE
MariaDB (v10.1 or later)	MariaDB Foundation	https://mariadb.org/download/
Perl (v5.8.8, v5.12.5, v5.22.1)	The Perl Programming Language	https://www.perl.org/get.html
Perl Module DBI.pm (v1.631, v1.636) and DBD mysql (v4.042)	Metacpan	https://www.perl.org/get.html
Picard (v2.8.2)	Broad Institute	https://broadinstitute.github.io/picard/
Python (v2.7.2 or later)	Python Software	https://www.python.org/downloads/
SAMtools	Li et al.[53]	https://samtools.sourceforge.net/
SRA Toolkit	Leionen et al.[54]	https://github.com/ncbi/sra-tools
STAR (v2.5.2b or later)	Dobin et al.[55]	https://github.com/alexdobin/STAR
RSEM v1.2.4	Li and Dewey[56]	https://deweylab.github.io/RSEM/
edgeR	Robinson et al.[57]	Bioconductor
TopHat v2.0.8b	Trapnell et al.[58]	https://ccb.jhu.edu/software/tophat/index.shtml
MISO v2.0	Katz et al.[59]	http://hollywood.mit.edu/burgelab/miso/
tidyverse	Wickham et al.[60] (https://doi.org/10.21105/joss.01686)	Bioconductor

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UpSetR	Conway et al.[61]	Bioconductor
GenomicRanges	Lawrence et al.[62]	Bioconductor
seqLogo	Bembom and Ivanek[63]	Bioconductor

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Omar Abdel-Wahab (abdelwao@mskcc.org).

Materials availability

All materials generated or analyzed during this study are included in this article and its supplemental information files. Please contact the lead contact for unique material requests. Any material that can be shared will be released via a material transfer agreement for non-commercial usage.

Data and code availability

- Data supporting the findings of this study are available from the corresponding authors upon reasonable request. RNA-seq and TRIBE-seq data have been deposited to GEO: GSE242094. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD044973.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

K562 (CCL-243), MCF10A (CRL-10317), 293T (CRL-3216), Panc05.04 (CRL-2557), cells were obtained from ATCC. Isogenic K562 and MCF10A cells with or without defined *SF3B1* mutations were generated by Horizon Discovery using adenoviral-mediated homologous recombination and previously described.[7,11,26] Uveal melanoma cell lines, MEL202 and MEL270, were obtained from B. Bastian and previously described.[11,42] MEL270 cells ectopically expressing *SF3B1*^{WT} or *SF3B1*^{K700E} were generated using a doxycycline-inducible expression system as previously described[7] and doxycycline (1 μg/ml) was added every 3 days to the media for induction. K562 cells ectopically expressing SRSF2^{WT}, SRSF2^{P95H}, U2AF1^{WT} or U2AF1^{S34F} were generated using lentiviral expression systems as previously described[33,34] and mCherry positive cells were sorted before downstream experiments. K562 cells were grown in Iscove's modified Dulbecco's media (IMDM) with 10% FBS (Gibco). MCF10A cells were grown in DMEM/F12 supplemented with 5% horse serum (Gibco), 20 ng/ml EGF (MilliporeSigma), 10 μg/ml insulin (MilliporeSigma), 0.5 μg/ml hydrocortisone (MilliporeSigma) and 0.1 μg/ml cholera toxin (MilliporeSigma). MEL270 and MEL202 cells were grown in RPMI1640 with 10% FBS (Gibco). MEL202 cells were additionally supplemented with 1% GlutaMAX (Gibco). Panc05.04 were grown in RPMI1640 with 20% FBS and 20 U ml^{-1} human recombinant insulin (Sigma). 293T cells were grown in DMEM supplemented with 10% FBS, 4.5g/L glucose and 1% GlutaMAX (Gibco). All cell lines were grown at 37 °C and 5% atmospheric CO_2 in the presence of 100 μg/ml penicillin and 100 mg/ml streptomycin.

Mice

All animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committees (IACUC) at Memorial Sloan Kettering Cancer Center (protocols 13-04-003 and 18-05-008). All mice were housed at Memorial Sloan Kettering Cancer Center with 12 h light/dark cycles and controlled temperature (20–22 °C) and humidity (40%–70%). 6–8-week-old female C57BL/6 Cd45.1^{+} mice were purchased from the Jackson Laboratory (stock no. 002014).

Sf3b1 K700E conditional knockin mouse was previously described.[64] *Sf3b1* K666N and *Sf3b1* R625H conditional knockin mice were generated by Ingenious Targeting Laboratory (Ronkonkoma, NY) using a targeted homologous recombination approach. Briefly, C57BL/6 FLP embryonic stem (ES) cells were electroporated with linearized targeting constructs encoding either inverted exon 12-14 (with K666N mutation) or inverted exon 12-17 (with R625H mutation) and the neomycin selection cassette. After selection with G418 antibiotic, surviving clones were expanded for PCR analysis to identify recombinant ES clones. Correctly targeted ES cells were injected into BALB/c blastocysts. The resulting chimeric animals were bred to C57BL/6 wildtype mice to generate Germline Neo Deleted mice. Final germline-transmitted mice were confirmed for the deletion of the Neomycin cassette and the presence of Sf3b1 mutant allele by genotyping tail DNA from mice with black coat color and sequencing the PCR product.

Sf3b1 mutant mice were bred to *Mx1*-cre mice (The Jackson Laboratory, Strain #:003556) to generate *Mx1*-cre *Sf3b1*$^{mut/WT}$ mice. Mouse genotypes from tail biopsies were determined using real-time PCR with specific probes designed for each allele (Transnetyx, Cordova, TN). Expression of Cre recombinase was induced with three polyinosine-polycytosine injections (pIpC; 12 mg/kg; GE Healthcare) every other day to achieve recombination in hematopoietic tissue.

METHOD DETAILS

Isolation and gene editing of human cord blood and AML bone marrow HSPCs

Umbilical cord blood was obtained from Bloodworks Northwest following institutional guidelines approved by the University of Washington. Acute myeloid leukemia (AML) bone marrow samples were obtained from the Memorial Sloan Kettering biospecimen repository following approved institutional guidelines. Cord blood and AML samples were processed using the Miltenyi CD34 Microbead kit according to manufacturer's instructions for human CD34$^+$ cell enrichment. Human CD34$^+$ cells isolated from cord blood were edited at the SF3B1 locus via CRISPR/Cas9 and adeno-associated virus (AAV6) template delivery system described in Sarchi et al.[65] Edited cells integrated a BFP fluorescent reporter in *SF3B1* intron 14 and either p.2098 A>G K700E mutation or wild-type K700K sequence.

Gene editing of either GPATCH8 or AAVS1 non-targeting control sgRNAs was performed with the Cas9 complex at a 1:2.5 molar ratio for each sgRNA in our engineered SF3B1$^{+/K700E}$ cord blood hematopoietic/stem progenitor cells and CD34$^+$ AML patient cells. The sequences of sgRNAs used are as follows: SF3B1 5′-UGGAUGAGCAGCAGAAAGUU-3′, AAVS1 5′-GGGCCACUAGGGACAG GAU-3′, GPATCH8 5′-ATGCTGAAGATGCTACCGAA-3′.

Editing efficiency was assessed 48 hours post-editing via flow cytometry for BFP. Genomic DNA was collected after 7 days of culture for PCR amplification of edited loci, followed by Sanger sequencing and Inference of CRISPR Edits (ICE) analysis (Synthego).

Human HSPC erythroid differentiation

Erythroid differentiation was performed as described in Clough et al.2022.[66] Briefly, primary edited CD34$^+$ HSPCs were cultured for 6 days in IMDM (ThermoFisher) + 15% FBS (Sigma) + 1% BSA (ThermoFisher) with 100 U/mL penicillin/streptomycin (Fisher), 2mM L-glutamine (ThermoFisher), 500ug/mL holo-Transferrin (Sigma), 10ug/mL Insulin (Cell Sciences), 100 ng/mL SCF, 5 ng/mL IL-3, and 6 U/mL erythropoietin (Procrit). After 6 days, IL-3 was removed and SCF reduced to 50ng/mL, and the culture continued for another 14 days. Cells were initially plated at a density of 1-2 x 10^5/ml and maintained <10^6/ml.

Vector cloning, lentiviral production, and cell transduction

To generate hPGK-mCardinal-P2A-mEmerald-synMAP3K7i250 vector, the mCardinal-P2A and P2A-mEmerald-synMAP3K7i250 fragments were amplified from mCardinal-pBiCMV-mEmerald-synMAP3K7i4-250 using primers described in the Key Resources Table. Fragments with correct sizes were purified on agarose gel and cloned into hPGK-PuroR-P2A-HSV-TK to replace the PuroR-P2A-HSV-TK sequence by Gibson assembly (NEBuilder HiFi, New England Biolabs). Orientation of fragments was as illustrated in Figure 1A.

Individual sgRNAs were cloned into LentiGuide-Puro and used along with LentiCas9-Blast construct to generate stable knockout cell lines. Correspondent untransduced cell lines were used for Puromycin and Blasticidin selection control. shRNAs were cloned into a modified version of SGEP (constitutive lentiviral miR-E) where GFP was replaced by Crimson fluorescent protein.

TRIBE plasmids used either RBP-ADAR or ADAR-RBP fusions with a downstream p2A-GFP to allow for fluorescence-based sorting of transfected cells. GPATCH8 was synthesized (GenScript) upstream of the *Drosophila* ADAR catalytic domain containing both hyperediting (E488Q) and auto editing blocking mutations (S458 synonymized)[48] (Addgene #154787). SUGP1 was synthesized downstream of the *Drosophila* ADAR catalytic domain containing both hyperediting (E488Q) and auto editing blocking mutations (S458 synonymized). For control plasmids either synonymized MS2 coat protein dimer (MCP) fused to the ADAR catalytic domain (stdMCP-ADAR) was used as a GPATCH8-ADAR control[48] (Addgene #154787) or in a flipped orientation (ADAR-stdMCP) for the ADAR-SUGP1 control. All TRIBE plasmids were subcloned from Addgene #154787 and sequence verified by GenScript.

For GPATCH8 rescue experiments in K562 cells, codon optimized GPATCH8 and its variants were synthetized and cloned into pGenLenti expression vector (GenScript) with a downstream Linker-NLS-3xFLAG-T2A-BFP. All pGenLenti variant plasmids were cloned, and sequence verified by GenScript. The same plasmids were used for transient expression of GPATCH8 and its variants in 293T cells using Lipofectamine 3000 (Invitrogen) transfections. After 48h, cells were analyzed by flow cytometry to confirm BFP expression prior to protein extraction and FLAG immunoprecipitation.

All lentiviruses were produced in 293T cells transfected with 4:2:3 ratios of expression plasmid:pVSVG:psPAX2 and 3:1 ratio of Polyethylenimine (PEI):DNA. Viral-containing supernatant was collected 48 and 72 h after transfection, filtered through 0.22 μm filter, supplemented with 4 ug/mL Polybrene, and used for cell transduction.

3X-FLAG-tag knockin into endogenous *GPATCH8*

Isogenic SF3B1^{WT} and SF3B1$^{K700E/WT}$ K562 cells were further edited by Synthego (Redwood City, CA) using CRISPR-Cas9 to insert 3X-FLAG-tag into endogenous *GPATCH8* N terminus. Single positive clones with homozygous insertion were selected. The sgRNA

sequence was GAGGAGUGAAGGCGGCAAAA, the donor DNA sequence was GGCGACGCCCGTGTTTACCTGAAAGTCTCGGTCT TCGTTGAAGCGGGAGAAGCGGTCCGCTCCGGAGCTGCCCTTGTCGTCGTCGTCCTTGTAGTCGATGTCGTGGTCCTTGTAGTCAC CGTCGTGGTCCTTGTAGTCCATGTTGCCGCCTTCACTCCTCTCAGGACGACGCTCTCCGGTTCGCTCCTTCCCTCCTGC, the PCR and sequencing primers were GGGAACCGGGGAGGGGTGGGTG (forward) and AGAGGGCTGGGAGCCGGAGATGA (reverse).

Whole-genome CRISPR-Cas9 Screening

Lentiviruses carrying human CRISPR Brunello lentiviral pooled sgRNA library was produced in 293T cells. Virus titer was determined by testing the percentage of puromycin-resistant cells after transducing MCF10A cells. A titer resulting in 40% of transduced cells (MOI 0.4) was selected for the screening. MCF10A WT or SF3B1 K700E mutant cells stably expressing Cas9 and hPGK-mCardinal-P2A-mEmerald-synMAP3K7i250 were transduced with lentivirus expressing human CRISPR Brunello sgRNA library. Puromycin selection (2 μg/mL) was performed 2 days after transduction in growth media for 5 days. The surviving cells were sorted by FACS to select the top 10% and bottom 10% cells by the ratio of mEmerald and mCardinal intensities. Cell pellets were lysed, and genomic DNA was extracted (EZNA Tissue DNA kit, Omega Biotek) and quantified by Qubit (Thermo Scientific). A quantity of gDNA covering 500× representation of gRNAs was PCR-amplified using TaKaRa Ex Taq DNA Polymerase (Takara) to add Illumina adapters and multiplexing barcodes. Amplicons were quantified by Qubit and Bioanalyzer (Agilent) and sequenced on an Illumina HiSeq 2500. Sequencing reads were aligned to the Brunello sgRNA library, and counts were obtained for each sgRNA. The probe level analysis was performed using the standard edgeR workflow with the glmLRT option for the model fitting/statistical test.

RT-PCR and quantitative RT-PCR

Total RNA was isolated using RNeasy Mini kit (Qiagen) and reverse-transcribed (RT) to cDNA using Verso cDNA synthesis kit (Thermo Scientific). PCRs were performed using GoTaq Green Master Mix (Promega) and amplicons were analyzed using 3% agarose gel electrophoresis. Quantitative RT-PCR (qRT-PCR) was performed using PowerUp SYBR Green PCR Master Mix and analyzed on QuantStudio 6 Flex Cycler (Applied Biosystems). Relative gene expression levels were calculated using the comparative threshold cycle method.

Cell viability and proliferation measurements

Cell viability and proliferation was measured using CellTiter-Glo Luminescent Cell Viability Assay (Promega) that determines the number of viable cells based on quantification of the ATP present in culture. Briefly, cells with or without GPATCH8 knockout were seeded at a density of 5000 cells per well in a 96-well white walled plate and viable cells were quantified at day 0 and day 5 (for K562, MEL202, MEL270 cells) or 7 (for MCF10A and Panc05.04 cells) of culture. Proliferation rate of cells was calculated by normalization of ATP signal at final time point (day 5 or 7 measurement) to ATP signal at initial time point (day 0 measurement).

Flow cytometry

mCardinal and mEmerald visualization was assessed at least 4 days after cell transduction with hPGK-mCardinal-P2A-mEmerald-synMAP3K7i250 vectors, using 7AAD and GFP filters respectively. All cells were resuspended in DPBS containing 1% BSA and 0.4 ng/mL DAPI (Invitrogen) prior to analysis to gate on viable (DAPI negative) cells. For GPATCH8 rescue experiments, mCardinal/mEmerald analysis was gated on BFP positive cells and DAPI was not used. All flow cytometry analyses were performed on BD LSRFortessa cytometer and data were analyzed on FlowJo (BD Biosciences).

Split APEX proximity labeling

Plasmids used for split APEX experiments (pRD-FLAG-AP-DHX15.WT, pX304-HA-SUGP1.FL-EX and pX304-HA-FRB-EX) were a gift from Feng.[37] 293T cells stably expressing Cas9 and sgRenilla or sgGPATCH8 #1 were co-transfected with inducible AP-tagged DHX15 and HA-EX-tagged SUGP1 at 1:1 ratio and treated with 2.5 mg/mL doxycycline for 48h. Cells were then incubated with 500 μM biotin-phenol for 30 min at 37°C followed by addition of H2O2 at a final concentration of 1mM for 30 sec at room temperature. The H2O2-containing medium was immediately discarded, and cells were washed three times with pre-chilled quenching solution (10 mM Sodium L-ascorbate, 5 mM Trolox, and 10 mM Sodium Azide in 1X PBS) before protein extraction with IP lysis buffer (Pierce) for western blots.

Western blotting

For MAP3K7 western blot, protein samples were extracted from cultured cells with RIPA buffer (Thermo Fisher Scientific) and quantified by BCA assay. Protein fractionated on NuPAGE 4%–12% Bis-Tris gels (Life Technologies) was transferred onto nitrocellulose membranes. Primary antibody was used at 1:1000 dilution in TBST containing 5% BSA and 0.02% sodium azide. IRDye 800CW Goat anti-Rabbit IgG Secondary Antibody was used at 1:10000 dilution in TBST containing 5% skim milk. The blot was then visualized by fluorescence on Odessey Imaging System (LI-COR). Precision Plus Protein Kaleidoscope Prestained Protein Standards (Bio-Rad) was used for protein size markers.

For split APEX western blots, proteins were extracted with IP lysis buffer (Pierce) and quantified by Qubit Protein Assay (Invitrogen). 200 μg of proteins were fractioned on NuPAGE 4%–12% Bis-Tris gels (Life Technologies) then transferred on PVDF membrane (Bio-Rad). Ponceau S staining was performed before membrane was blocked using Intercept (TBS) Blocking Buffer (LI-COR) for 30 min at

room temperature then incubated with HA-Tag antibody 1:1000 diluted in Intercept T20 (TBS) Antibody Diluent (LI-COR) overnight at 4°C. After three washes in TBST, membrane was incubated with IRDye 680RD Goat anti-Rabbit IgG Secondary Antibody and IRDye 800CW Streptavidin, both 1:5000 diluted in Intercept T20 (TBS) Antibody Diluent containing 0.02% SDS, for 1 hour at room temperature. After three washes in TBST, the blot was visualized by fluorescence on Odessey Imaging System (LI-COR). Chameleon Duo Pre-Stained Protein Ladder (LI-COR) was used for protein size.

For DHX15 western blot following FLAG immunoprecipitation, proteins were extracted with IP lysis buffer and quantified by Qubit Protein Assay. 4 mg of proteins were incubated overnight at 4°C with Protein A agarose beads (Millipore) conjugated to FLAG antibody (clone M2, Sigma) resuspended in IP lysis buffer. Beads were subsequently washed three times using IP lysis buffer before protein elution by boiling in the presence of Laemmli SDS-Sample buffer 6X (Boston BioProducts). Eluates and 40 μg of proteins (1% input) were fractioned on NuPAGE 4%–12% Bis-Tris gels then transferred on PVDF membrane. Membrane was blocked then incubated with DHX15 antibody 1:1000 diluted in Intercept T20 (TBS) Antibody Diluent overnight at 4°C. After three washes in TBST, membrane was incubated with anti-rabbit IgG HRP-linked Antibody 1:10000 diluted in Intercept T20 (TBS) Antibody Diluent overnight at 4°C, washed, then visualized by chemiluminescence on Amersham ImageQuant 800 (Cytiva). PageRuler Plus Prestained Protein Ladder (ThermoFisher Scientific) was used for protein size.

RNA-sequencing library preparation and sequencing

For cell line RNA sequencing (RNA-seq), RNA was extracted from MCF10A, K562, or 293T cells using the Qiagen RNeasy extraction kit, according to the manufacturer's instructions. A minimum of 500 ng of high-quality RNA (as determined by Agilent Bioanalyzer) per replicate was used as input for library preparation. Poly(A)-selected, strand-specific (dUTP method) Illumina libraries were prepared by the Integrated Genomics Operation (IGO) at Memorial Sloan Kettering with a modified TruSeq protocol and sequenced on the Illumina HiSeq 2000 to obtain ~60-80M paired-end 100 bp reads per sample.

For primary cell RNA-seq, gene edited BFP^+CD34^+ hematopoietic stem and progenitor cells (day 0) were isolated 72 hours after editing and $BFP^+CD235a^+$ cells were isolated after 7 days of erythroid culture. Cells were flow sorted directly into Trizol (ThermoFisher). Library preparation and sequencing were performed by Omega Biosciences. Following RNA extraction and QC, 10 ng of RNA was used as input for SMART-Seq v4 Ultra Low Input RNA Kit (Takara). Illumina sequencing was performed to obtain a minimum 100M paired-end 150 bp reads per sample.

RNA-seq data analysis

A genome annotation (UCSC hg19/GRCh37) was created by combining the UCSC knownGene,[67] Ensembl 71,[68] and MISO v2.0[59] annotations. Transcriptome mapping was accomplished using RSEM v1.2.4[56] with the argument ''-v 2'' which indicates Bowtie invocation.[69] This mapping strategy also generates gene expression values, in units of transcripts per million (TPM). All RSEM-generated gene expression estimates were normalized by applying the trimmed mean of M values (TMM) method.[57] Unaligned reads were mapped to the genome with TopHat v2.0.8b,[58] and to a custom annotation which includes all gene-level pairwise combinations of annotated 5′ and 3′ splice sites. The combined RSEM/Bowtie and TopHat alignments were inputted to MISO v2.0[59] to quantify isoform expression levels. Differential isoform expression between sample groups was ascertained using the two-sided t test. Differentially spliced events were defined as containing at least 20 isoform-specific reads, a minimum absolute difference of 10% in isoform expression, and a p value < 0.05.

Splice site sequence analyses

Differentially spliced cassette exon events following GPATCH or SUGP1 deletion were identified. 10-mer and 54-mer sequences which contain the 5′ and 3′ splice site dinucleotide motifs, respectively, were generated from treatment-responsive splicing events. We previously identified branchpoint nucleotide coordinates through a large-scale analysis of intron lariat-derived reads from thousands of RNA-seq datasets.[29] These positions were used to define 9-mer branchpoint sequences. The extracted sequences were used to generate sequence logo plots, visualizations of the associated position weight matrices, via the seqLogo package.[63] seqLogo: Sequence logos for DNA sequence alignments. R package version 1.66.0.). The enrichment of purines/pyrimidines in skipped cassette exons (relative to included) was measured within and around the exonic regions using the GenomicRanges.[62] The 95% confidence interval was estimated with bootstrapping (1000 resampling iterations). The analyses were performed within the R Programming environment with tools from Bioconductor,[70] visualized using the tidyverse (https://doi.org/10.21105/joss.01686) packages.

Colony-forming assays

Four weeks after pIpC treatment of *Mx1*-cre *Sf3b1* mutant or WT mice, bone marrow cells were collected and c-Kit^+ hematopoietic precursors were magnetically separated using murine CD117 MicroBeads according to the manufacturer's instructions (Miltenyi Biotech) then cultured overnight in IMDM with 20% FBS supplemented with mSCF (20 ng ml^{-1}), mFLT3L (10 ng ml^{-1}) and mTPO (20 ng ml^{-1}). For the next 2 day, cells were subjected to spinfections at 2,300 r.p.m. for 90 min at 32 °C with lentiviral supernatant expressing shRNAs against *Gpatch8* or *Renilla* non-targeting control in the presence of 8 μg/ml polybrene (Sigma Aldrich). 48h after transduction, viable shRNA-expressing cells ($DAPI^-Crimson^+$) were FACS-sorted, seeded into cytokine-supplemented methylcellulose medium (Methocult M3434; STEMCELL Technologies) and incubated at 37 °C with 5% CO_2 and ≥ 95% humidity. Colonies were scored after 10 days of culture, manually using inverted microscope.

GPATCH8 and SUGP1 TRIBE-seq

293T were transfected with GPATCH8-ADAR, ADAR-SUGP1, or respective MCP/ADAR non-targeting control, fusion expression plasmids using a 3:1 ratio of Polyethylenimine (PEI):DNA. 24 hours after transfections cells were suspended in sorting buffer (DPBS, 1% BSA supplemented with DAPI) and passed through a 35μm strainer mesh. FACS was performed by selecting for GFP positive DAPI negative single cells. One million GFP positive cells were sorted on a BD FACSymphony S6 or BD FACSAria III Cell Sorter, and RNA was isolated using RNeasy Plus Mini Kit. A minimum of 500 ng of high-quality RNA (as determined by Agilent Bioanalyzer) per replicate was used as input for library preparation. Ribo-depleted Illumina libraries were prepared by the Integrated Genomics Operation (IGO) at Memorial Sloan Kettering with a modified TruSeq protocol and sequenced on the Illumina HiSeq 2000 to obtain ~60-80M paired-end 100 bp reads per sample.

Analysis of TRIBE-seq data

TRIBE-seq data was analyzed following the previously published pipeline.[48,52] In brief, reads from the sequencing libraries are trimmed (Cutadapt[51]) and aligned (STAR[55]) to the hg19 human reference genome. The aligned BAM file is converted to a SAM file and the nucleotide frequency at each position in the transcriptome is recorded from aligned reads and placed into a SQL table. For each nucleotide in the transcriptome, the TRIBE RNA nucleotide frequency is compared with the wild-type mRNA library (WT RNA) nucleotide frequency to identify RNA editing sites. For a legitimate edit site, the frequency of A is greater than 80% and the frequency of G is 0.5% in the WT RNA and the frequency of G is greater than 0 in the TRIBE RNA (using the reverse complement if annotated gene is in the reverse strand). Modifying the nucleotide frequency of G in WT RNA from 0 to 0.5% allows very low-level sequence heterogeneity in WT RNA and does not disrupt the identification of legitimate editing sites in deeply sequenced libraries. We used an editing threshold of at least 5% in each replicate. The TRIBE edit sites are required to be present in both replicates and any sites that overlap with MCP/ADAR control edit sites with at least 1% editing are removed. After background subtraction, BED files containing all edit sites present in both experimental replicates are analyzed using RNAmod.[71]

To compare TRIBE-seq data to RNA splicing changes, FASTQ files from 293T cell RNA-sequencing were trimmed (Cutadapt[51]) aligned to UCSC hg19 human reference genome using STAR[55] in two pass mode with the following parameters: -twopassMode Basic, -alignSJoverhangMin 8, -alignSJDBoverhangMin 1 -alignIntronMin 20 -alignIntronMax 1000000 -alignMatesGapMax 1000000) and sorted BAM files were generated. These files were fed into rMATs-turbo[72] with the following parameters (-t paired -novelSS -variable-read-length -allow-clipping – readLength 101 -cstat 0.0001 -libType fr-firststrand). Significant splicing changes were those determine to have at least 20 reads, an FDR < 0.05 and a delta PSI value of > 10%. Significant gene symbols were used for comparison to TRIBE-seq genes as well as gene ontology analysis using R.

CRISPR tiling screen

sgRNA sequences targeting the coding regions of human GPATCH8 and human SUGP1 were designed using Benchling's CRISPR design tool and cloned into LentiGuide-mTagBFP vector as two separate libraries, each containing 10% of non-targeting and safe harbor sgRNAs. Lentiviral particles containing sgRNA libraries were produced in 293T cells and lentiviral supernatant was pre-titrated in K562 cells to obtain 10%–20% infection rate (monitored by flow cytometry for BFP expression). K562 WT or SF3B1 K700E mutant cells stably expressing Cas9 and hPGK-mCardinal-P2A-mEmerald-synMAP3K7i250 were transduced with the pre-titrated sgRNA library viruses. 48h post-transduction, top 10% and bottom 10% mEmerald/mCardinal expressing cells, gated on BFP+ cells, were sorted. The number of collected cells for each population covered at least 1000× the number of constructs in each library. Genomic DNA was extracted and representation of sgRNAs was analyzed similarly to whole-genome CRISPR-Cas9 Screening described above.

Alphafold2 and ColabFold modeling of protein-protein interactions

Alphafold2 and ColabFold[73] was utilized to model the interactions between proteins of interest using default settings. To facilitate computational modeling of the proteins, ColabFold models were generated using the structured parts of the proteins and unstructured regions were excluded. For the SUGP1 DHX15 structure, we used the recently published crystal structure (PDB: 8EJM).[18] This structure included amino acids 113-795 of DHX15 and G-patch amino acids 543-614 of SUGP1. For the DHX15 and GPATCH8 models we used amino acids 68-795 of DHX15 and amino acids 1-188 of GPATCH8 (encompassing the coiled coil, G-patch and zinc finger domains of the protein).

Immunoprecipitation for mass spectrometry

Proteins were extracted from whole cells using IP lysis buffer (Pierce) and quantified using Qubit Protein Assay (Invitrogen). 2 mg of proteins were incubated overnight at 4°C with Protein A agarose beads (Millipore) conjugated to FLAG antibody (clone M2, Sigma), DHX15 antibody (A300-390A, Bethyl) or their respective isotype controls resuspended in IP lysis buffer (Pierce). Beads were subsequently washed three times using IP lysis buffer and three times using 10mM Tris-HCl/150mM NaCl pH 7.5.

Protein digestion for proteomic analyses

The beads were resuspended in 40 μL of 2M Urea, 50mM Ammonium Bicarbonate pH 8.5 and treated with DL-dithiothreitol (1mM final concentration) for 30 minutes at 37°C with shaking (1100 rpm) on a Thermomixer (Thermo Fisher). Free cysteine residues were

alkylated with 2-iodoacetamide (3.67mM final concentration) for 45 minutes at 25°C at 1100 rpm in the dark. LysC (750 ng) was added, followed by incubation for 1h at 37°C at 1150 rpm. Finally, trypsin (750 ng) was added, followed by incubation for 16 hours at 37°C at 1150 rpm.

After incubation, the digest was acidified to pH <3 with the addition of 50% of trifluoroacetic acid (TFA), and the peptides were desalted on 3-plug C18 (3M Empore™ high performance extraction disks) stage tips. Briefly, the stage tips were conditioned by sequential addition of: i) 100 μL 100% Acetonitrile (ACN), ii) 100 μL 70% ACN/0.1% TFA, iii) 100 μL 0.1% Formic Acid (FA), iv) 100 μL 0.1% FA. Following conditioning, the acidified peptide digest was loaded onto the stage tip. The stationary phase was washed once with 100 μL of 0.1% FA. Finally, samples were eluted using 50μL of 70% ACN/0.1% FA twice. Eluted peptides were dried under vacuum followed by reconstitution in 12 μL of 0.1% FA, sonication and transfer to an autosampler vial. Peptide yield was quantified by NanoDrop (Thermo Fisher).

Mass spectrometry analyses

Peptides were separated on a 50 cm column composed of C18 stationary phase (Thermo Fisher ES903) using a gradient from 0.5% to 25% buffer B over 100 minutes, to 50% in 15 minutes, to 90% in 5 minutes (buffer A: 0.1% FA in HPLC grade water; buffer B: 99.9% ACN, 0.1% FA) with a flow rate of 300nL/min using a nanoAcquity Hplc system (Waters). MS data were acquired on an Eclipse mass spectrometer (Thermo Fisher Scientific) using a data-independent acquisition (DIA) method. The method consisted of one MS1 scan, AGC target Standard, maximum injection time of 50 msec, scan range of 380-985 m/z and a resolution of 120K. Fragment ions were analyzed in 60 DIA windows at a resolution of 15K.

DIA data analysis

Raw data files were processed using Spectronaut version 17.4 (Biognosys) and searched with the PULSAR search engine with a Homo sapiens UniProt protein database downloaded on 2022/09/23 (226,953 entries). Cysteine carbamidomethylation was specified as fixed modifications, while methionine oxidation, acetylation of the protein N terminus and deamidation (NQ) were set as variable modification. A maximum of two trypsin missed cleavages were permitted. Searches used a reversed sequence decoy strategy to control peptide false discovery rate (FDR) and 1% FDR was set as threshold for identification. Unpaired t test was used to calculate p value in differential analysis, volcano plot was generated based on $\log_2$ fold change and q value (multiple testing corrected p value). A q value of $\leq$0.05 was considered the statistically significant cut-off.