Identification of BAY61-3606 derivatives with improved activity in splicing modulation that induces inclusion of cassette exons similar to the Splicing Factor 3B subunit 1 mutation

Running Title

Structure-activity relationship analysis of splicing-modulatory compounds with unknown mode of action revealed a link between their activity and the splicing change of poison exons in SR protein genes

Takanori Matsumaru[a † ‡], Toshiki Iwamatsu[b †], Kana Ishigami[b], Makoto Inai[b], Wataru Kanto[c] Ayumi Ishigaki[c], Atsushi Toyoda[d], Satoshi Shuto[a], Katsumi Maenaka[a], Shinichi Nakagawa[a] and Hiroshi Maita[a]*

matsumaru@chem.keio.ac.jp, iwamatsu_etf@mii.maruho.co.jp, ishi.kana1992.6.8@gmail.com, makinai@u-shizuoka-ken.ac.jp, bike.wck.55@gmail.com, ayumi.gakitya@gmail.com, atoyoda@nig.ac.jp, shu@pharm.hokudai.ac.jp, maenaka@pharm.hokudai.ac.jp, nakagawas@pharm.hokudai.ac.jp, maimai@pharm.hokudai.ac.jp

[a] Faculty of Pharmaceutical Sciences, Hokkaido University, Kita-12, Nishi-6, Kita-ku, Sapporo 060-0812, Japan

[b] School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

[c] Graduate School of Life Science, Hokkaido University, Kita-12, Nishi-6, Kita-ku, Sapporo 060-0812, Japan

[d] Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka, 411-8540, Japan

[†] These authors equally contributed

‡ Present address: Department of Chemistry, Faculty of Sciences and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
*To whom correspondence should be addressed

Kita 12 Nishi 6, Kita-ku, Sapporo, Hokkaido 060-0812, Japan
Tel: +81-011-706-3921
Fax: +81-011-706-4988
E-mail:maimai@pharm.hokudai.ac.jp (Hiroshi Maita)
ORCID: https://orcid.org/0000-0002-9620-4651 (Hiroshi Maita)

The authors declare no competing financial interest.

Acknowledgments
This research project owes a significant part to Professor Toshiyuki Kan, who contributed greatly until his death in July 2021. We deeply appreciate his dedication to this research. We also express our gratitude to Satoko Otsuguro, Rie Nakaido and Yuka Yoshida for their assistance. Additionally, we acknowledge the support and assistance provided by the Center for Research and Education on Drug Discovery and the Global Facility Center (GFC) at Hokkaido University. This study was supported by JSPS KAKENHI Grant Number 16K08225, 20K07028 and 23K06093 granted to H.M. and 24H00546 Grant-in-Aid for Scientific Research (A) granted to S.N. This study was also supported by JSPS KAKENHI Grant Number 16H06279 (PAGS) and Japan Agency for Medical Research and Development (AMED) Grant Number JP17am0101093 and JP22ama121037 to K.M.

Abstract

Splicing modulation by a small compound offers therapeutic potential for diseases caused by splicing abnormality. However, only a few classes of compounds that can modulate splicing have been identified. We previously identified BAY61-3606, a multiple kinase inhibitor, as a compound that relaxes the splicing fidelity at the 3' splice site recognition. We have also reported the synthesis of derivatives of BAY61-3606. In this study, we tested those compounds for their splicing modulation capabilities and identified two contrasting compounds. These compounds were further investigated for their effects on the whole transcriptome, and analysis of changes in transcription and splicing revealed that the highly active derivative in the splicing reporter assay also showed significantly higher activity in modulating the splicing of endogenously expressed genes. Particularly, cassette exon inclusion was highly upregulated by this compound, and clustering analysis revealed that these effects resembled those in splicing factor 3b subunit 1 (SF3B1) K700E mutant cells but contrasted with those of the splicing inhibitor H3B-8800. Additionally, a group of serine/arginine-rich (SR) protein genes was identified as representatively affected, likely via modulation of poison exon inclusion. This finding could guide further analysis of the mode of action of these compounds on splicing, which could be valuable for developing drugs for diseases associated with splicing abnormalities.

Keywords

pre-mRNA splicing, splicing-modulatory compound, structure-activity relationship, SR proteins, poison exon

Introduction

Splicing abnormalities are widely known to cause various genetic diseases (Anna & Monika, 2018; Daguenet, Dujardin, & Valcárcel, 2015), and recent studies have revealed that aberrant splicing is also exploited by various types of cancer cells (De Kesel, Fijalkowski, Taylor, & Ntziachristos, 2022; Yoshida & Ogawa, 2014). However, only a few classes of small compounds are available that target pre-mRNA splicing, which limits the ability to treat many diseases by targeting the splicing abnormality (Effenberger, Urabe, & Jurica, 2017; Eymin, 2021). Thus, it is important to enlarge the repertoire of small compounds that modulate splicing, which may provide a potentially useful compound.

In the splicing reaction, the splice site consists of conserved short sequences at both the 5' and 3' termini of an intron are essential for the spliceosome assembly. Consequently, substituting these sequences can have deleterious effects on the splicing reaction, and many mutations are found in these sites (Buratti, Chivers, & Královičová, 2007; Vořechovský, 2006). Because of the nature of their susceptibility to cause heredity diseases, we have been trying to find compounds that allow splicing at the mutated 3' splice sites, which results in the identification of luteolin and apigenin (Chiba, Ariga, & Maita, 2016). To further identify such compounds, we previously identified a small compound, BAY61-3606, through a compound library screening. BAY61-3606 enhanced the efficiency of splicing at the mutated 3' splice site in a reporter gene (Tomita, Nakagawa, Ariga, & Maita, 2023). Here, we screened derivatives of BAY61-3606 and identified specific moieties critical for splicing recovery by these compounds. Furthermore, to observe the feature of these compounds' effect on global splicing, transcriptomic analyses were performed using the original compound and two derivatives: one significantly more efficacious than the original and the other almost inactive. Unfortunately, the splicing analysis revealed that the improved derivative enhanced the inclusion of a cassette exon but could not induce splicing at the endogenous non-canonical 3' splice site. However, to characterize our compounds in terms of the splicing modulatory activity, we performed analysis to compare the effects of our compound with the publicly available datasets derived from samples with disease-causative splicing factor mutations or samples treated with known small compounds targeting splicing. This analysis revealed that our compound has a similarity in splicing

modulatory activity to the K700E mutation in splicing factor 3b subunit 1 (SF3B1), a mutation identified in hematologic malignancies. In contrast, it differed from H3B-8800, a small molecule splicing inhibitor that targets SF3B1 (Seiler et al., 2018). Moreover, SR protein genes showed characteristic responses to our compound. These results indicate that our compounds could be applicable as a novel class of compounds that target splicing.

Methods and Materials

Cell culture and chemical compounds

The 293T and its cell lines harboring either the snRNP reporter or the splicing reporter genes had been previously established (Chiba et al., 2016). These cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum. K562 cells were cultured in RPMI with 10% fetal bovine serum. BAY61-3606 and its derivatives utilized in this study were synthesized as previously described (Matsumaru et al., 2017), and the concentration used was 10μM, except for Fig. 1D.

snRNP reporter assay

The assay method was described in the previous study (Maita et al., 2014). In brief, cells expressing the split luciferase reporter were treated with the tested compounds for 4 hours, then lysed in BN buffer (20 mM Hepes-KOH (pH7.9), 20 mM KCl, 0.5 M 6-aminohexanoic acid, 1.5 mM $MgCl_2$, 0.5% Triton X-100 and 10% Gycerol). Following centrifugation at 11,100*g* for one minute, an aliquot of the supernatant was mixed with 100 μL of luciferase substrate solution (Steady-Glo Luciferase Assay System, Promega), and the luminescence was immediately measured using Lumat LB9507 (Berthold technologies).

Splicing reporter assay

The 293T cell line expressing pFD-Luc-Py-AT was described before (Tomita et al., 2023). HeLa cell line expressing pFD-Luc-Py-AT was also obtained by identifying the stable transfectant by GFP fluorescence. To evaluate splicing efficiency via reporter activity, 96-well plates were used. The next day after seeding cells into the plates, half of the medium was replaced with the medium containing each compound at double the

intended final concentration. After 24 hours, the medium was removed, and cells were lysed using BN buffer. The cell lysates were then manually transferred to 96-well black-walled plates, and GFP fluorescence was measured using the GloMax system (Promega). For the detection of luciferase activity, Steady-Glo, a luciferase substrate solution (Promega), was added to 96-well white-walled plates. Following GFP fluorescence measurement, the lysates were transferred to these plates containing Steady-Glo. Luminescence was measured immediately after mixing the cell lysates with Steady-Glo using GloMax. Subtracting the average counts from empty wells to account for the background signal, the relative luciferase activity was determined by calculating the ratio of luciferase luminescence to GFP fluorescence. When summarizing the compounds' activities, experimental results for Ti050 and other compounds that are typically active, yet showed a two-fold decrease in GFP fluorescence counts compared to the DMSO-treated sample, were omitted to ensure accuracy.

Reverse transcription (RT)-PCR

pFD-Luc-Py-AT was transfected into 293T cells, and compounds were added to the culture medium 24 hours post-transfection. After an additional 24 hours of incubation, RNA was extracted from the cells and used for RT-PCR analysis. RNAiso (TaKaRa) and ReverTra Ace RT master mix (Toyobo) were used for RNA extraction and reverse transcription, respectively. ExTaq was used for labeled PCR. To detect spliced mRNAs, fragments were amplified using the sense primer (ex17-up: 5′-CAACAAGAAACCACTGGCAC-3′) and the antisense primer (pFD-as-IRD1: 5′-GCCCAGCTTCACAGTAGTG-3′-IRD800). The amplicons were electrophoresed through 2% agarose gel. Labeled PCR products were detected and quantified using an Odyssey imager (Aloka).

Transcriptomic analysis

Total RNA extracted from 293T or K562 cells, treated with each compound at 10 μM for 4 hours, was utilized for RNA-seq library preparation using a TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold (Illumina). Paired-end sequencing reads, obtained using the HiSeq 2500 system (Illumina) with 2 x 100 bp

read length, were mapped to the GRCh38 reference genome using HiSAT2 (ver. 2.2.1) (Kim, Paggi, Park, Bennett, & Salzberg, 2019). Raw fastq data obtained from RNA-seq were deposited in the DDBJ Sequence Read Archive (DRA) under accession number PRJDB18151. The resulting Bam files were used as input for Rsubread (Liao, Smyth, & Shi, 2019), an R package designed to acquire count data, which was subsequently analyzed using the edgeR package (M. D. Robinson, McCarthy, & Smyth, 2010). Additionally, for splicing change analysis, these BAM files also served as inputs for rMATS (ver. 4.1.1) (Y. Wang et al., 2024). The differences in inclusion levels of significantly altered splicing isoforms, as identified by rMATS, were used to draw boxplots presented in Figures 2 and 3. The summary of the rMATS analysis outcomes is provided in Supplementary Tables.

Plot construction and statistical analysis

Most plots were generated using the ggplot2 package (Wickham, Chang, & Wickham, 2016) in R. For the dose-response curves, we utilized the drc package (ver. 3.0.1) (Ritz, Baty, Streibig, & Gerhard, 2015) to fit the data to a four-parameter logistic model. The resulting calculated values were then integrated with ggplot2 for plotting. To construct a heatmap, we utilized the heatmap.2 function from the gplots package (Warnes et al., 2016), applying it to the count data before visualization with ggplot2. The edgeR package was used to calculate gene expression change probabilities, with the result being incorporated into ggplot2 to generate MA plots. The MA plot is a two-dimensional plot where the y-axis indicates the fold change as the log ratio of gene expression levels between two different conditions, and the x-axis indicates the mean log intensity. Additionally, we determined the number of overlapping genes between two groups using R's base functions. These counts were then input into the eulerr package (Larsson, 2018) to obtain parameters for Venn diagram. The Venn diagrams were subsequently produced using ggplot2. For the gene enrichment analysis using GO terms, we employed the web tool KOBAS (http://bioinfo.org/kobas/) (Bu et al., 2021), selecting KEGG as the pathway database. The obtained data were downloaded and utilized to generate a lollipop plot using ggplot2. Sashimi plots were created using the gviz package (Hahne & Ivanek, 2016). RNA-seq was conducted on three independent samples for 293T and on two for K562 cells. One of those BAM files was utilized as

input to generate a plot with gviz for both the vehicle control (DMSO) and Ti050-treated samples. To present the gene structures in the upper section, a gff3 file obtained from GENCODE Comprehensive gene annotation, Release 45 (GRCh38.p14) served as the annotation file. Pearson correlation coefficients for transcript levels and fold changes of SR proteins were calculated using R's base functions. Publicly available datasets were analyzed using the same pipeline described above. Euclidean distance calculation and heatmap construction were performed using R. The accession IDs for each dataset from the GEO database are presented in the Supplementary Tables.

Results

Identification of more and less potent derivatives of BAY61-3606

BAY61-3606 was identified as a hit compound in the assay using the small nuclear Ribonucleoprotein complex (snRNP) reporter. The snRNP assay detects snRNPs, which are ribonucleoprotein complexes that assemble onto pre-mRNA to form the spliceosome. Specific snRNP components fused to split luciferase fragments were used as a reporter pair that reconstitutes luciferase activity only when involved in the snRNP. We hypothesized that a compound modulating cellular snRNP levels might alter splicing. We used the snRNP reporter for the assay to screen the compound library and identified BAY61-3606 (Maita, Tomita, & Ariga, 2014; Tomita et al., 2023). Therefore, derivatives of BAY61-3606, described in our previous work (Matsumaru et al., 2017), were initially tested in the snRNP reporter assay. In this assay, the activity of each compound was assessed at 10 μM, as detailed in Table 1. The results of the assay revealed that certain critical structures in these compounds were required to modulate the snRNP reporter activity. Concurrently, the compounds were also evaluated in a splicing reporter assay at the same concentration with the snRNP reporter assay (10 μM). The splicing reporter assay was designed to detect the change of splicing efficiency at the non-canonical 3' splice site by the ratio of luciferase activity encoded in the last exon to the GFP fluorescence encoded in the first exon (Fig. 1A), which was used to evaluate the splicing modulating activity of BAY61-3606 (Chiba et al., 2016; Tomita et al., 2023).

Importantly, most of the results tested in the snRNP reporter showed reduced activity (Table 1), and there is a discrepancy between the results of the snRNP reporter

and the splicing reporter. Compounds such as 7c and 7p, which were highly active in the snRNP assay, did not promote splicing recovery in the splicing reporter assay. While the compound treatment time for the snRNP reporter assay was set to 4 hours, the latter assay required a treatment duration of 24 hours. This prolonged exposure could induce toxicity or interfere with the transcription of the reporter gene. Indeed, after 7c or 7p treatment, the levels of the reporter gene expression, as detected by GFP fluorescence, decreased more than twofold compared to the vehicle control (Supplementary Figure 1), potentially leading to the observed discrepancies between the two assays. This result indicates that these compounds might target other proteins in addition to the protein involved in the splicing regulation due to their structural modifications. Given these findings, we ceased comparing derivatives via the snRNP assay. Instead, we shifted our focus to evaluating all compounds based on their efficacy in splicing recovery using the splicing reporter because our goal was to identify the splicing-modulating compounds.

7h (Ti056) and 7k (Ti050) are closely related derivatives, differing only in the substitution of 3,4-dimethoxyphenyl to 4-butoxyphenyl and 4-methylphenyl, respectively, as illustrated in Fig.1C. However, they largely differed in the splicing recovery activity (Fig. 1B). While Ti056 was less active than BAY61-3606 in promoting splicing recovery, Ti050 was much more active than BAY61-3606. Notably, Ti050 and other close derivatives occasionally exhibited decreased levels in both GFP fluorescence and splicing recovery activity when tested in the 96-well plate format (Supplementary Figure 2). Therefore, we thought that the conclusion based on the results obtained at a single concentration was insufficient. Subsequently, we evaluated the dose responses of both compounds. As depicted in Fig. 1D, the dose-response curves for Ti050 clearly show higher activity than those for BAY61-3606. Furthermore, we also compared the splicing recovery activity of Ti050 with several analogs in a different cell line, and the phenomenon was confirmed by RT-PCR (Supplementary Figure 3). These results are consistent with the conclusion that Ti050 shows improved activity in the splicing recovery of our splicing reporter gene. However, our results cannot determine whether the compound has two targets: one related to splicing modulatory activity and another one related to cell toxicity. Alternatively, it is possible that the excessive effects of the compound caused cell toxicity or undesired splicing changes of the reporter gene, resulting in the reduction of GFP fluorescence.

In addition, the splicing recovery activities of these derivatives are presented as a heatmap alongside the previously reported kinase inhibitory activities (Matsumaru et al., 2017) (Fig. 1E), which indicates no correlation between the higher activity in the splicing reporter assay and the kinase inhibitory activities against Spleen tyrosine kinase (Syk) and Germinal Center Kinase (GCK).

The difference in efficacy of Ti050 and Ti056 on the whole transcriptome is consistent with that observed in the splicing reporter

Next, we evaluated the effects of these compounds on the whole transcriptome, as we have shown that the original compound BAY61-3606 induced significant alterations in splicing patterns after 4 hours of treatment in 293T cells (Tomita et al., 2023). Then, we compared the effects of Ti056 and Ti050 in the transcriptome analysis at the same condition as the previous report (4 hours of treatment at 10μM). Firstly, count data from 293T cells treated with Ti056 and Ti050 were analyzed to assess gene expression changes. The top 500 genes exhibiting significant changes were depicted in a heatmap following clustering (Fig. 2A). The heatmap revealed that gene expression changes induced by Ti056 closely resembled those observed in the vehicle control sample treated with dimethyl sulfoxide (DMSO). Conversely, the changes in gene expression following Ti050 treatment were markedly distinguished from the vehicle control sample and Ti056. Additionally, the MA plot, which illustrates both expression levels and fold changes, showed a trend consistent with the heatmap (Fig. 2B).

To analyze changes in splicing, we employed rMATS, a package designed for alternative splicing analysis(Shen et al., 2014). rMATS categorized complex alternative splicing events into five distinct classes: cassette exon inclusion (CEI), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), mutually exclusive exon (MXE) and retained intron (RI), as illustrated in Fig. 2C. Within each class, splicing patterns may shift in both directions, shown as in a cartoon; however, active compounds tended to increase the total number of events and also biased the directionality of these changes. To visualize the number and direction of splicing events affected by each compound, we presented the significantly changed events following Ti056 and Ti050 treatment in scatter plots, shown in Fig. 2D and 2E, respectively. Although we expected that our compounds lead to an increase in A3SS, both compounds actually increased the inclusion of cassette

exons and reduced retained introns. However, the number of events significantly altered by the treatment with these compounds differed, as detailed in Supplementary Tables. The finding that Ti050 altered the expression levels and splicing patterns in a greater number of genes compared to Ti056 is consistent with the results obtained using the splicing reporter.

Higher activity of Ti050 compared to BAY61-3606 regarding their effects on the whole transcriptome

To compare Ti050 with the original compound BAY61-3606 and to evaluate the effects of these compounds in a different cell type, we conducted RNA-seq analysis using K562 cells under the same condition as for 293T cells. Since K562 cells are frequently used to evaluate the effects of splicing factor mutations found in hematologic malignancies on splicing, we expected that testing our compounds on this cell line would allow us to compare the compound's effects with those of disease-associated samples. The MA plots revealed significant changes in gene expression levels induced by both compounds (Fig. 3A). Notably, Ti050 treatment resulted in a larger number of affected genes, correlating with the finding that Ti050 is more efficacious than BAY61-3606 in the splicing reporter assay. However, when comparing gene expression levels between Ti050-treated and BAY61-3606-treated cells, most genes showed no changes in expression with only a small subset exhibiting differential expression (Fig. 3A, right panel). This suggests that the directionality of gene expression alterations is consistent across both compounds, aligning with the observation that Ti050 is a derivative of BAY61-3606 with enhanced activity. The impact on splicing by both compounds was also investigated using rMATS, and the results are shown in scatter plots (Fig. 3B). The analysis revealed a slightly higher number of genes with altered splicing patterns in Ti050-treated cells (2041 in Ti050 vs. 1589 in BAY61-3606; refer to Supplementary Tables). Notably, we repeatedly found the common genes with both compounds, showing similar splicing pattern changes. Consequently, we assessed the overlap of gene names affected by each compound. This comparison showed a significant overlap in the gene groups influenced by each treatment (Fig. 3C). However, the Ti050-treated cells exhibited a notably higher count of unique gene names compared to BAY61-3606-treated

cells, mirroring the results from the splicing reporter assay. Furthermore, the categories of splicing pattern changes that were significantly altered in K562 cells were similar to those in 293T cells, indicating a consistent mode of action across these cell lines. This suggests that the target protein of the splicing-modulating activity of these compounds could be a ubiquitous factor.

Effects of BAY61-3606 and Ti050 on cassette exon inclusion resembled the SF3B1 K700E mutant but contrasted with H3B-8800.

Next, we tried to compare splicing changes triggered by our compounds with those caused by disease-associated splicing factor mutants or known splicing inhibitors. We re-analyzed eight publicly available raw RNA-seq datasets. These included two independent datasets for each of the splicing factor mutations (Yoshida et al., 2011), such as SF3B1 K700E, U2 small nuclear RNA auxiliary factor 1 (U2AF1) S34F and Serine/arginine-rich splicing factor 2 (SRSF2) P95H (GSE187356 (Lieu et al., 2022) and GSE94528 (Seiler et al., 2018) for SF3B1; GSE224576 (Wedge et al., 2023) and GSE242092 for U2AF1 S34F; GSE235600 (Liu et al., 2024) and GSE200446 (Huang et al., 2022) for SRSF2 P95H) and one dataset for H3B-8800 (Seiler et al., 2018) and E7107 (Kotake et al., 2007) (GSE94528 (Seiler et al., 2018) and GSE114558 (E. Wang et al., 2019), respectively), all using K562 cells. Obtained rMATS results are presented in Supplementary Tables, and plots drawn in the same style as Fig. 3B are shown in Supplementary Figure 4. When the results were compared, irrespective of the direction of splicing changes, a significant increase of the total event numbers in cassette exon inclusion in both SF3B1 K700E-expressing cells and H3B-8800-treated cells was similar to the results obtained with cells treated with our compounds. We then focused on cassette exon inclusion events and performed clustering analysis using these ten datasets (eight public datasets and our two datasets). To compare the similarity of the splicing changes observed in each dataset, Euclidean distances were used, which were calculated based on the differences in Inclusion Level Difference values obtained from rMATS for common genes among each group. As shown in the heatmap (Fig. 3D), the effects of our compounds closely resembled those of the SF3B1 mutant groups. Interestingly, the H3B-8800 group clustered oppositely, suggesting that cassette exons included by our compound overlapped with those in the SF3B1 mutant, while similar genes were skipped by H3B-8800. To visualize the number of genes oppositely regulated by our compounds

and H3B-8800, a Venn diagram is shown in Fig. 3E. Although the high number of events identified in the H3B-8800-treated sample increased the likelihood of overlap, 247 genes were identified as oppositely regulated between H3B-8800 and both our compound and the SF3B1 mutant. A representative image of this regulation was visualized using Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011) (Fig. 3F).

Analysis of changes in splicing patterns of serine/arginine-rich (SR) protein genes

To further investigate the molecular mechanism underlying the effects of our compounds, we focused on their common impact on the transcriptome. We identified genes that showed alteration in both 293T and K562 cells in terms of expression levels and splicing patterns (Fig. 4A and Supplementary Figure 5). Specifically, the numbers of common differentially expressed genes (DEGs) and genes with altered cassette exon inclusion patterns (ASGs) were found to be 821 and 608, respectively. Although decreases in retained introns were another observed feature of our compounds, this category was not used in subsequent analysis because it is unclear if these decreases are due to transcriptional inhibition, which reduces nascent transcripts and could be misinterpreted as decreases in retained introns. These gene lists were then subjected to enrichment analysis using Gene Ontology (GO) terms to identify the candidate biological processes potentially influenced by these compounds (Supplementary Figure 6). The enriched processes identified within the DEGs revealed no GO terms suggesting relationships with specific pathways. However, the ASGs notably included genes associated with the spliceosome. In addition, three splicing regulatory proteins – SRSF2, SRSF6, and Muscleblind-like protein 2 (MBNL2)– remained among the common small gene set that comprised only 13 genes shared between the DEGs and ASGs (Fig. 4, listed in the box).

Co-occurrence of reduced transcripts and poison exons inclusion in certain SR proteins

Considering the detailed expression changes focusing on SRSF2 and SRSF6, it is notable that there was an upregulation of the inclusion of cassette exons both in the presence of Ti050 and BAY61-3606 (Fig. 5A and B), despite the decreased expression levels of both genes (Fig. 5C). Previous research has shown that a splicing-dependent mechanism regulates the expression of SR proteins (L. F. Lareau, Inada, Green, Wengrod,

& Brenner, 2007; Leclair et al., 2020; Ni et al., 2007; Pervouchine et al., 2019). In this mechanism, as the levels of SR protein increase, the SR proteins bind to their own mRNA, which promotes the inclusion of a 'poison exon' that contains a stop codon (Änkö et al., 2012; Sureau, Gattoni, Dooghe, Stévenin, & Soret, 2001; Wagner & Frye, 2021). This mechanism is understood as an autoregulatory feedback loop designed to stabilize the expression levels of SR proteins through non-sense mediated decay (NMD) pathway (L. Lareau & Brenner, 2015; Müller-McNicoll, Rossbach, Hui, & Medenbach, 2019). Sashimi plots, which compare splicing patterns of SRSF2 and SRSF6, demonstrated that such splicing events occurred in the presence of Ti050 (Fig. 5A and B). However, we have not tested if the expression levels of these genes increased upon inhibition of NMD. To determine whether other SR proteins were also influenced by these compounds, we searched gene lists of ASGs under each condition. This analysis identified three additional SR proteins –SRSF3, SRSF4 and SRSF11– as ASGs in both 293T and K562 cells. Additionally, SRSF1 and SRSF7 were identified as ASGs uniquely in 293T and K562 cells, respectively. Sashimi plots for these genes were presented in Supplementary Figures 7 and 8. A scatter plot illustrating the fold changes in gene expression levels upon compound treatment revealed that the data points representing each SR protein gene were closely aligned between 293T and K562 cells (Fig. 5C). Utilizing the fold changes from both cell lines for a scatter plot resulted in points of each SR protein was aligning nearly on a straight line, with a Pearson correlation coefficient of 0.94 (Fig. 5D). Further scatter plot based on expression levels also revealed the correlation between the two cell lines; however, the orders of transcript abundance and the fold changes in gene expression levels presented Fig. 5D and E were not the same. These findings indicate that the observed changes in SR protein gene expression levels by these compounds were not due to compound-induced random variations of gene expression, but rather were likely regulated by compound-induced splicing shifts. This suggests that these compounds may modulate poison exon-mediated autoregulatory feedback mechanisms of SR proteins. To confirm this, it is necessary to compare the protein levels of these genes, although we have not tested that yet.

Discussion

We found that two BAY61-3606 derivatives, differing only in a substituent group at the phenyl ring, exhibited opposite effects on the splicing recovery activity of a reporter gene with a mutated 3' splice site. Compound 7h (Ti056) features a 4-butoxy group, while 7k (Ti050) contains a 4-methyl group at this position. This suggests that small substituents at this position may be crucial for fitting into the binding pocket of the protein involved in splicing recovery. However, compounds 7p and 7c, which either lack a substituent group or have a 4-methoxy substituent at this position, showed high activity in the snRNP assay but were significantly less effective in the splicing reporter assay. Therefore, small substitutions such as the 4-methyl group at this position may be particularly effective in binding the target protein involved in splicing recovery. Alternatively, varying substitutions might influence their ability to bind the multiple target proteins, thereby affecting the balance between the compound's toxicity and its splicing recovery efficacy. However, as seen in the heatmap in Fig. 1E, the inhibitory activities of 7p and 7c against Syk and GCK are comparable with BAY61-3606. This suggests that the target should be other kinases or proteins. If the occasional reduction of splicing recovery activity only for 7a, 7c, 7g and 7k in the 96-well plate assay is due to such secondary target proteins, further structural modifications may allow the separation of splicing modulatory activity from cell toxicity.

It should also be noted that 7q, which lacks a substituent at the phenyl ring similar to 7p but includes a methyl at the amide moiety, exhibited high splicing recovery activity. Conversely, 7r, 7s, 7t, 7u 7v, and 7w, featuring various substitutions similar to the methyl group in 7p, from the amide moiety to larger alkyl groups, showed a gradual reduction in splicing recovery activity. The observation that 7p, which has no substitution at both the phenyl group and the amide moiety, exhibited low splicing recovery activity, likely due to toxicity or transcriptional inhibition, suggests that a minor modification to the amide moiety might mitigate adverse target binding. A previous study has demonstrated that BAY61-3606 inhibits ATP binding in various kinases (Lau et al., 2012), suggesting that it can bind to the enzymatic pocket of multiple target kinases. Although the structures of the catalytic core of kinases are similar, minor differences lead to variations in the spectrum of target kinases for a small compound inhibitor that binds to these cores. Consistent with this, 7p exhibited inhibitory activity against GCK kinase comparable to that of BAY61-3606, while it decreased fourfold in 7q (Matsumaru et al., 2017). This

inverse correlation between GCK inhibitory activity and the splicing recovery activity suggests that a methyl group at the amide moiety in 7q may play a role in distinguishing the target protein involved in the splicing regulation from GCK. Thus, it is possible that small changes in compounds could affect the splicing reporter activity by altering the targeted kinases. However, it is unclear whether the kinase inhibition directly causes the splicing changes or indirectly affects the reporter activity through increased toxicity or by influencing other processes than splicing.

Our transcriptomic analysis has revealed that both BAY61-3606 and Ti050 trigger changes in the splicing pattern of several SR protein genes. Notably, these alterations occur in the poison exon within SR protein genes, which is an integral part of an autofeedback loop that stabilizes SR protein expression levels (L. Lareau & Brenner, 2015; L. F. Lareau et al., 2007; Müller-McNicoll et al., 2019). This feedback mechanism plays a crucial role under stress conditions such as hypoxia (de Oliveira Freitas Machado et al., 2023). However, the reported splicing changes in hypoxia primarily show an increase in cassette exon inclusion events, which is opposite to changes observed in Ti050-treated cells. Nonetheless, Ti050 promotes the inclusion of poison exons in a distinct set of SR proteins, indicating that exploring the mechanisms behind these splicing changes in SR proteins could provide insights into the specific targets of Ti050 and BAY61-3606 in splicing modulation. Interestingly, a contrasting change was observed in SRSF7, which might be key to understanding the mechanism.

Furthermore, despite our attempt, we could not detect any obvious increase in aberrant splicing at the non-AG 3' splice sites, such as the mutated splice site used in our splicing reporter (data not shown). Although non-AG 3' splice sites are involved in the category of weak 3' splice sites, their sequences are near-cognate to the consensus and these near-cognate 3' splice sites are rejected by the proofreading mechanism of the spliceosome (Mayas, Maita, & Staley, 2006). Therefore, the finding that our compounds only increased the usage of weak but cognate 3' splice sites suggests that these compounds did not suppress the proofreading mechanism; rather, they just facilitated the recognition of the weak 3' splice site.

On the other hand, our compounds enhanced the inclusion of a normally skipped exon, resembling the effects of the SF3B1 K700E mutant but contrasting with H3B-8800. Malignant cells with splicing factor mutations rely on the wild-type allele for survival

(Lee et al., 2016, 2018; Zhou et al., 2015). The observation that our compounds induced splicing changes in normal cells similar to SF3B1 mutant indicates that it may cause lethality, more specifically in mutant cells. Furthermore, our compounds may enhance the production of neoantigens that are targeted by cancer immunotherapy. Recent studies suggest that aberrant splicing is a major source of neoantigens (Jayasinghe et al., 2018; Kahles et al., 2018; Li et al., 2024; Smart et al., 2018), and splicing factor mutations were proposed to contribute to their production (Bigot et al., 2021). Our compounds strongly induce cassette exon inclusion similar to the SF3B1 mutant, possibly leading to neoantigen production as proposed for other splicing modulatory compounds (Lu et al., 2021; Matsushima et al., 2022). Further analysis of our compounds' mode of action could provide valuable insights for developing drugs targeting hematological malignancies and other cancers associated with splicing abnormalities.

Reference

Änkö, M.-L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., … Neugebauer, K. M. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biology*, *13*, R17. doi:10.1186/gb-2012-13-3-r17

Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics*, *59*, 253–268. doi:10.1007/s13353-018-0444-7

Bigot, J., Lalanne, A. I., Lucibello, F., Gueguen, P., Houy, A., Dayot, S., … Lantz, O. (2021). Splicing patterns in SF3B1-mutated uveal melanoma generate shared immunogenic tumor-specific neoepitopes. *Cancer Discovery*, *11*, 1938–1951. doi:10.1158/2159-8290.CD-20-0555

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., … Kong, L. (2021). KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for

gene enrichment analysis. *Nucleic Acids Research*, *49*, W317–W325. doi:10.1093/nar/gkab447

Buratti, E., Chivers, M., & Královičová, J. (2007). Aberrant 5′ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids*. Retrieved from https://academic.oup.com/nar/article-abstract/35/13/4250/1203203

Chiba, M., Ariga, H., & Maita, H. (2016). A splicing reporter tuned to non-AG acceptor sites reveals that luteolin enhances the recognition of non-canonical acceptor sites. *Chemical Biology & Drug Design*, *87*, 275–282. doi:10.1111/cbdd.12656

Daguenet, E., Dujardin, G., & Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. *EMBO Reports*, *16*, 1640–1655. doi:10.15252/embr.201541116

De Kesel, J., Fijalkowski, I., Taylor, J., & Ntziachristos, P. (2022). Splicing dysregulation in human hematologic malignancies: beyond splicing mutations. *Trends in Immunology*, *43*, 674–686. doi:10.1016/j.it.2022.06.006

de Oliveira Freitas Machado, C., Schafranek, M., Brüggemann, M., Hernández Cañás, M. C., Keller, M., Di Liddo, A., … Müller-McNicoll, M. (2023). Poison cassette exon splicing of SRSF6 regulates nuclear speckle dispersal and the response to hypoxia. *Nucleic Acids Research*, *51*, 870–890. doi:10.1093/nar/gkac1225

Effenberger, K. A., Urabe, V. K., & Jurica, M. S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome. *Wiley Interdisciplinary Reviews. RNA*, *8*. doi:10.1002/wrna.1381

Eymin, B. (2021). Targeting the spliceosome machinery: A new therapeutic axis in cancer? *Biochemical Pharmacology*, *189*, 114039. doi:10.1016/j.bcp.2020.114039

Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. *Methods in Molecular Biology* , *1418*, 335–351. doi:10.1007/978-1-4939-3578-9_16

Huang, Y.-J., Chen, J.-Y., Yan, M., Davis, A. G., Miyauchi, S., Chen, L., … Zhang, D.-E. (2022). RUNX1 deficiency cooperates with SRSF2 mutation to induce multilineage hematopoietic defects characteristic of MDS. *Blood Advances*, *6*, 6078–6092. doi:10.1182/bloodadvances.2022007804

Jayasinghe, R. G., Cao, S., Gao, Q., Wendl, M. C., Vo, N. S., Reynolds, S. M., … Ding, L. (2018). Systematic analysis of splice-site-creating mutations in cancer. *Cell Reports*, *23*, 270-281.e3. doi:10.1016/j.celrep.2018.03.052

Kahles, A., Lehmann, K.-V., Toussaint, N. C., Hüser, M., Stark, S. G., Sachsenberg, T., … Rätsch, G. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell*, *34*, 211-224.e6. doi:10.1016/j.ccell.2018.07.001

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*, 907–915. doi:10.1038/s41587-019-0201-4

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., … Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nature Chemical Biology*, *3*, 570–575. doi:10.1038/nchembio.2007.16

Lareau, L., & Brenner, S. (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Molecular Biology and Evolution*, *32*, 1072–1079. doi:10.1093/molbev/msv002

Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., & Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, *446*, 926–929. doi:10.1038/nature05676

Larsson, J. (2018). *eulerr: Area-proportional Euler diagrams with ellipses*. Retrieved from https://lup.lub.lu.se/student-papers/record/8934042/file/8935805.pdf

Lau, K. S., Zhang, T., Kendall, K. R., Lauffenburger, D., Gray, N. S., & Haigis, K. M. (2012). BAY61-3606 affects the viability of colon cancer cells in a genotype-directed manner. *PloS One*, *7*, e41343. doi:10.1371/journal.pone.0041343

Leclair, N. K., Brugiolo, M., Urbanski, L., Lawson, S. C., Thakar, K., Yurieva, M., … Anczuków, O. (2020). Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. *Molecular Cell*, *80*, 648-665.e9. doi:10.1016/j.molcel.2020.10.019

Lee, S. C.-W., Dvinge, H., Kim, E., Cho, H., Micol, J.-B., Chung, Y. R., … Abdel-Wahab, O. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. *Nature Medicine*, *22*, 672–678. doi:10.1038/nm.4097

Lee, S. C.-W., North, K., Kim, E., Jang, E., Obeng, E., Lu, S. X., … Abdel-Wahab, O. (2018). Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. *Cancer Cell*, *34*, 225-241.e8. doi:10.1016/j.ccell.2018.07.003

Li, G., Mahajan, S., Ma, S., Jeffery, E. D., Zhang, X., Bhattacharjee, A., … Salomonis, N. (2024). Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Science Translational Medicine*, *16*. doi:10.1126/scitranslmed.ade2886

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*, e47. doi:10.1093/nar/gkz114

Lieu, Y. K., Liu, Z., Ali, A. M., Wei, X., Penson, A., Zhang, J., … Mukherjee, S. (2022). SF3B1 mutant-induced missplicing of MAP3K7 causes anemia in myelodysplastic syndromes. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2111703119. doi:10.1073/pnas.2111703119

Liu, X., Devadiga, S. A., Stanley, R. F., Morrow, R. M., Janssen, K. A., Quesnel-Vallières, M., … Klein, P. S. (2024). A mitochondrial surveillance mechanism activated by SRSF2 mutations in hematologic malignancies. *The Journal of Clinical Investigation*, *134*. doi:10.1172/JCI175619

Lu, S. X., De Neef, E., Thomas, J. D., Sabio, E., Rousseau, B., Gigoux, M., … Bradley, R. K. (2021). Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell*, *184*, 4032-4047.e31. doi:10.1016/j.cell.2021.05.038

Maita, H., Tomita, K., & Ariga, H. (2014). A split luciferase-based reporter for detection of a cellular macromolecular complex. *Analytical Biochemistry*, *452*, 1–9. doi:10.1016/j.ab.2014.01.015

Matsumaru, T., Inai, M., Ishigami, K., Iwamatsu, T., Maita, H., Otsuguro, S., … Kan, T. (2017). Divergent synthesis of kinase inhibitor derivatives, leading to

discovery of selective Gck inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *27*, 2144–2147. doi:10.1016/j.bmcl.2017.03.055

Matsushima, S., Ajiro, M., Iida, K., Chamoto, K., Honjo, T., & Hagiwara, M. (2022). Chemical induction of splice-neoantigens attenuates tumor growth in a preclinical model of colorectal cancer. *Science Translational Medicine*, *14*, eabn6056. doi:10.1126/scitranslmed.abn6056

Mayas, R. M., Maita, H., & Staley, J. P. (2006). Exon ligation is proofread by the DExD/H-box ATPase Prp22p. *Nature Structural & Molecular Biology*, *13*, 482–490. doi:10.1038/nsmb1093

Müller-McNicoll, M., Rossbach, O., Hui, J., & Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *Journal of Molecular Cell Biology*, *11*, 930–939. doi:10.1093/jmcb/mjz043

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., … Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes & Development*, *21*, 708–718. doi:10.1101/gad.1525507

Pervouchine, D., Popov, Y., Berry, A., Borsari, B., Frankish, A., & Guigó, R. (2019). Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic Acids Research*, *47*, 5293–5306. doi:10.1093/nar/gkz193

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-Response Analysis Using R. *PloS One*, *10*, e0146021. doi:10.1371/journal.pone.0146021

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26. doi:10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*, 139–140. doi:10.1093/bioinformatics/btp616

Seiler, M., Yoshimi, A., Darman, R. B., Chan, B., Keaney, G., Thomas, M., … Buonamici, S. (2018). H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature Medicine*, *24*, 497–504. doi:10.1038/nm.4493

Shen, S., Park, J. W., Lu, Z.-X., Lin, L., Henry, M. D., Wu, Y. N., … Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E5593-601. doi:10.1073/pnas.1419161111

Smart, A. C., Margolis, C. A., Pimentel, H., He, M. X., Miao, D., Adeegbe, D., … Van Allen, E. M. (2018). Intron retention is a source of neoepitopes in cancer. *Nature Biotechnology*, *36*, 1056–1058. doi:10.1038/nbt.4239

Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J., & Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *The EMBO Journal*, *20*, 1785–1796. doi:10.1093/emboj/20.7.1785

Tomita, K., Nakagawa, S., Ariga, H., & Maita, H. (2023). BAY61-3606 Alters snRNP Composition and Enhances Usage of Suboptimal Splice Acceptor Site. *Biological & Pharmaceutical Bulletin*, *46*, 147–157. doi:10.1248/bpb.b22-00471

Vořechovský, I. (2006). Aberrant 3′ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Research*, *34*, 4630–4641. doi:10.1093/nar/gkl535

Wagner, R. E., & Frye, M. (2021). Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *43*, e2000242. doi:10.1002/bies.202000242

Wang, E., Lu, S. X., Pastore, A., Chen, X., Imig, J., Chun-Wei Lee, S., … Aifantis, I. (2019). Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell*, *35*, 369-384.e7. doi:10.1016/j.ccell.2019.01.010

Wang, Y., Xie, Z., Kutschera, E., Adams, J. I., Kadash-Edmondson, K. E., & Xing, Y. (2024). rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nature Protocols*. doi:10.1038/s41596-023-00944-2

Warnes, M. G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., & Others. (2016). Package 'gplots.' *Various R Programming Tools for Plotting Data*, 112–119. Retrieved from https://www.researchgate.net/profile/Arni-Magnusson-2/publication/303186599_gplots_Various_R_programming_tools_for_plotting_data/data/5a8fd3b80f7e9ba4296a11a1/gplots.pdf

Wedge, E., Ahmadov, U., Hansen, T. B., Gao, Z., Tulstrup, M., Côme, C., … Kristensen, L. S. (2023). Impact of U2AF1 mutations on circular RNA

expression in myelodysplastic neoplasms. *Leukemia*, *37*, 1113–1125. doi:10.1038/s41375-023-01866-4

Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2.' *Create Elegant Data Visualisations Using the Grammar of Graphics. Version*, *2*, 1–189. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=af53fd2f5b9e81b6edec0c13e1b3babd34bda399

Yoshida, K., & Ogawa, S. (2014). Splicing factor mutations and cancer. *Wiley Interdisciplinary Reviews. RNA*, *5*, 445–459. doi:10.1002/wrna.1222

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., … Ogawa, S. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, *478*, 64–69. doi:10.1038/nature10496

Zhou, Q., Derti, A., Ruddy, D., Rakiec, D., Kao, I., Lira, M., … Stegmeier, F. (2015). A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Research*, *75*, 1949–1958. doi:10.1158/0008-5472.CAN-14-2930

Table 1

Effects of BAY 61-3606 derivatives on the snRNP reporter assay and the splicing reporter assay

Compound	R	R′	R″	snRNP assay (%)	Splicing assay (%)
1 (BAY 61-3606)	3,4-Di-OMe	H	H	100	100
7a	2-OMe	H	H	102	150
7b	3-OMe	H	H	11.6	43.7
7c	4-OMe	H	H	93.5	15.9
7d	3,4-Dioxole	H	H	37.3	51.0
7e	2,4-Di-OMe	H	H	NA	193
7f	3,5-Di-OMe	H	H	22.3	41.4
7g	4-OPr	H	H	NA	138
7h (Ti056)	4-OBu	H	H	NA	61.9
7k (Ti050)	4-Me	H	H	NA	210
7l	3-OMe	H	Me	38.3	30.9
7m	4-OMe	H	Me	36.6	30.6
7n	3,4-Di-OMe	H	Me	NA	61.9
7o	3,4-Dioxole	H	Me	35.4	30.5
7p	H	H	H	120	6.58
7q	H	H	Me	79.4	171
7r	H	H	Et	36.6	55.4
7s	H	H	Pr	42.1	43.4
7t	H	H	*i*-Pr	16.2	46.5
7u	H	H	Bu	16.2	46.2
7v	H	H	*s*-Bu	73.2	20.4
7w	H	H	Bn	69.2	19.4
7x	H	Me	Me	57.8	35.0

NA: Not Applicable

Figure Legends

Figure 1 (A) Schematic illustration of the splicing reporter and the measurement of splicing recovery utilizing this reporter. (B) Comparison of compounds' recovery abilities to enhance splicing at the mutated 3' splice site. The recovery activities were determined by normalizing the luciferase counts to the GFP counts with the activity of BAY61-3606 set as 100%. The activity is presented as the mean of two independent experiments. Each experiment consisted of two technical replicates. The data points from these technical replicates are overlaid as dots with different colors indicating independent experiments. The lower dotted line represents the average activity of the vehicle control (DMSO), serving as the standard, and the upper dotted line signifies the average activity of BAY61-3606. (C) Structural features of BAY61-3606 and its derivatives, Ti050 and Ti056. R, R' and R'' denote the positions where the substituents, as listed in Table 1, are incorporated. (D) The dose-response curve of Ti050 in comparison with BAY61-3606. The recovery activities were determined similar with (A). The activity is presented as fold change relative to the reporter activity in untreated samples. The experiment was repeated three times, each with three biological replicates, and the representative results were shown. (E) A heatmap of splicing recovery activities of BAY61-3606 derivatives, determined by the splicing reporter assay, alongside the kinase inhibitory activities referenced from our previous report.

Figure 2 Comparison of Ti056 and Ti050 through transcriptome analysis in 293T cells. (A) Heatmap illustrating the top 500 genes, showing variation in expression levels among 293T cells treated with the vehicle control (DMSO), Ti056, and Ti050. (B) MA plots display transcript levels and fold changes relative to the control sample for Ti056 (left panel) and Ti050 (right panel). Genes exhibiting significant expression changes (FDR<0.01) are highlighted in red. (C) A schematic representation of splicing patterns detected by rMATS. (D) Box plots illustrate splicing events significantly altered by each compound (left: Ti056; right: Ti050), with data points representing the inclusion differences overlaid on the box plots.

Figure 3 Comparison of Bay61-3606 and Ti050 through transcriptome analysis in K562 cells. (A) MA plots display transcript levels and fold changes relative to the control sample for BAY61-3606 (left panel) and Ti050 (center panel). Right panel shows a comparative analysis between Ti050 and BAY61-3606, visualizing the similarity in the effects of these compounds. Genes exhibiting significant expression changes (FDR<0.01) are highlighted in red. (B) Box plots illustrate splicing events significantly altered by each compound (left: BAY61-3606; right: Ti050), with data points representing the inclusion differences overlaid on the box plots. (C) A Venn diagram displays overlapping genes with significant alteration in cassette exon inclusion events detected in rMATS analysis between BAY61-3606-treated and Ti050-treated cells. (D) Heatmap for clustering analysis using Euclidean distance based on each gene's Inclusion Level Difference values in the event type CEI (Cassette exon inclusion) obtained by rMATS. Closely related groups are indicated in white or lighter colors (Ti050, BAY61-3606 and SF3B1 mutants), while darker colors with H3B-8800 indicate significant differences. (E) Venn diagram showing the number of overlapping genes among the three sample groups. Ti050 and SF3B1groups contain genes with increased cassette exon inclusion, while the H3B-8800 group contains genes with increased exon skipping. (F) Expression data of a representative common gene (RBM39) in the three groups visualized using mapped reads in IGV.

Figure 4 The analysis of transcriptome changes by Ti050 identified a small subset of genes. Venn diagrams illustrate overlapping genes that exhibit significant changes in both transcript levels (left, DEGs) and splicing patterns (right, ACGs), with a focus on cassette exon inclusion (CEI) events for ACGs. The size of each circle corresponds to the number of genes in that group, and the circle color denotes the cell type (magenta for 293T and green for K562). The overlap in gene numbers between 293T and K562 for transcript levels is 821, and for splicing patterns, it is 608. A further overlap between these two groups is depicted in a subsequent Venn diagram (below). The remaining 13 genes identified across all categories are listed in the box.

Figure 5 Visualization of alterations in SR protein genes at both the transcript and splicing levels. (A) Sashimi plots illustrating SRSF6 in 293T cells and (B) SRSF2 in

K562 cells. The upper sections show the annotated exon-intron structures with various isoforms. The upper sashimi plot represents the splicing patterns observed in the control sample, while the lower plot corresponds to the Ti050-treated sample. An arrow highlights the ‘poison exon’ inclusion significantly induced by Ti050 treatment. (C) A scatter plot compares the changes in transcript levels of SR protein genes, with point colors indicating cell types (blue for 293T and magenta for K562). Points representing the same gene are positioned closely together. The y-axis (logCPM) indicates the expression levels of each gene, and the x-axis(-logFC) indicates the fold change in gene expression levels induced by a compound treatment compared to the vehicle control sample. Except for SRSF9 and SRSF11, the expression levels of all SRSF genes were decreased. (D) This scatter plot compares the log fold changes in transcript levels of SR protein genes between 293T and K562 cells. The Pearson correlation coefficient is displayed in the upper left corner of the plot. Genes showing significant changes in cassette exon inclusion (CEI) splicing events in both cell types are enclosed in a solid box, while those in a single cell type are in a dashed box. A regression line demonstrates that most data points align closely with the line. (E) Transcript levels of SR protein genes in K562 (y-axis) vs. 293T (x-axis) cells are shown in a scatter plot with a regression line, including the Pearson correlation coefficient in the upper left corner. The gene order for SR proteins in both plots is not correlated.

Figure 1

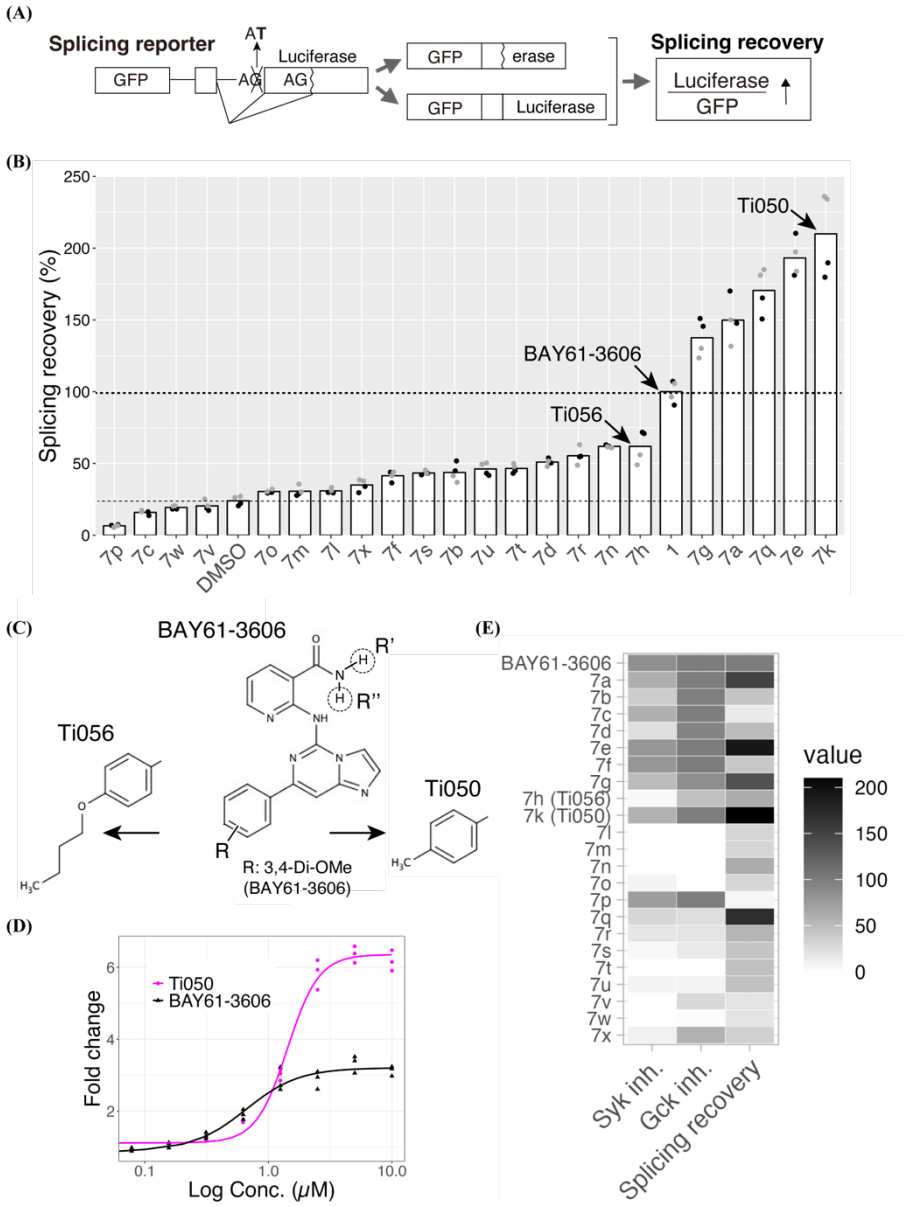

(A)
Splicing reporter
AT
Luciferase
GFP
AG
AG
GFP
erase
GFP
Luciferase
Splicing recovery
Luciferase
GFP
(B)
Splicing recovery (%)
250
200
150
100
50
0
Ti050
BAY61-3606
Ti056
7p 7c 7w 7v DMSO 7o 7m 7l 7x 7f 7s 7b 7u 7t 7d 7r 7n 7h 1 7g 7a 7q 7e 7k
(C)
BAY61-3606
R'
R''
Ti056
Ti050
R
R: 3,4-Di-OMe
(BAY61-3606)
(D)
Ti050
BAY61-3606
Fold change
6
4
2
0.1
1.0
10.0
Log Conc. (μM)
(E)
BAY61-3606
7a
7b
7c
7d
7e
7f
7g
7h (Ti056)
7k (Ti050)
7l
7m
7n
7o
7p
7q
7r
7s
7t
7u
7v
7w
7x
value
200
150
100
50
0
Syk inh.
Gck inh.
Splicing recovery

Figure 2

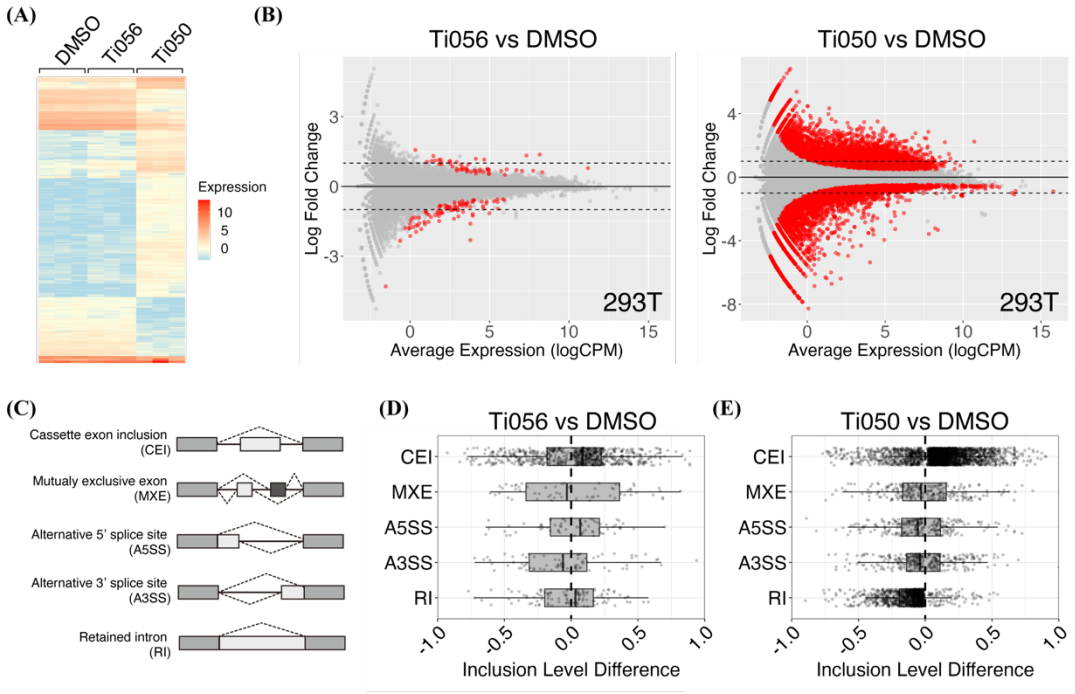

(A)
DMSO
Ti056
Ti050
Expression
10
5
0
(B)
Ti056 vs DMSO
Log Fold Change
3
0
-3
293T
0
5
10
15
Average Expression (logCPM)
Ti050 vs DMSO
Log Fold Change
4
0
-4
-8
293T
0
5
10
15
Average Expression (logCPM)
(C)
Cassette exon inclusion (CEI)
Mutualy exclusive exon (MXE)
Alternative 5' splice site (A5SS)
Alternative 3' splice site (A3SS)
Retained intron (RI)
(D)
Ti056 vs DMSO
CEI
MXE
A5SS
A3SS
RI
-1.0
-0.5
0.0
0.5
1.0
Inclusion Level Difference
(E)
Ti050 vs DMSO
CEI
MXE
A5SS
A3SS
RI
-1.0
-0.5
0.0
0.5
1.0
Inclusion Level Difference

Figure 3

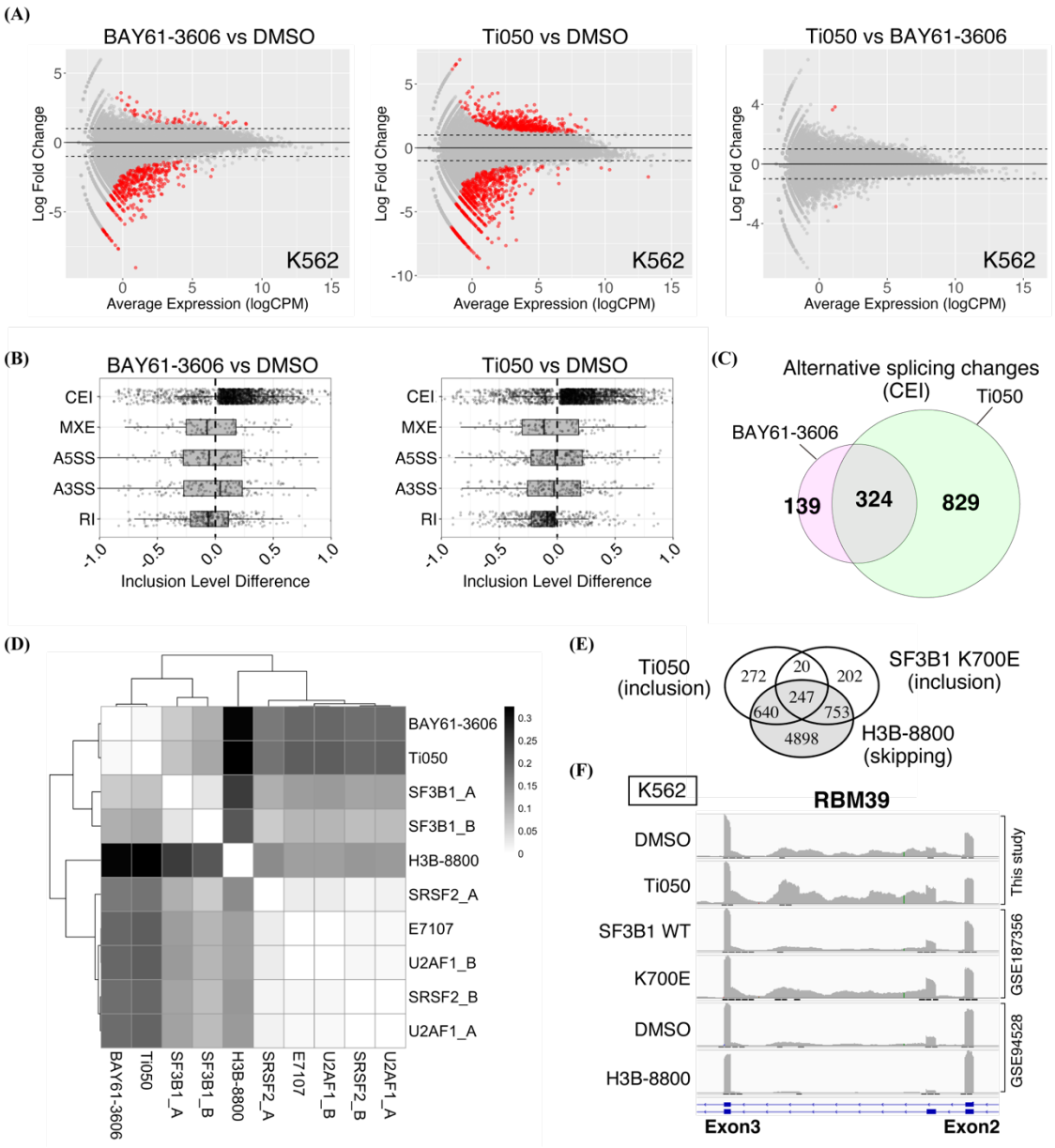

(A)
BAY61-3606 vs DMSO
Ti050 vs DMSO
Ti050 vs BAY61-3606
Log Fold Change
Average Expression (logCPM)
K562
(B)
BAY61-3606 vs DMSO
Ti050 vs DMSO
CEI
MXE
A5SS
A3SS
RI
Inclusion Level Difference
(C)
Alternative splicing changes (CEI)
BAY61-3606
Ti050
139
324
829
(D)
BAY61-3606
Ti050
SF3B1_A
SF3B1_B
H3B-8800
SRSF2_A
E7107
U2AF1_B
SRSF2_B
U2AF1_A
(E)
Ti050 (inclusion)
SF3B1 K700E (inclusion)
H3B-8800 (skipping)
272
20
202
247
640
753
4898
(F)
K562
RBM39
DMSO
Ti050
SF3B1 WT
K700E
DMSO
H3B-8800
This study
GSE187356
GSE94528
Exon3
Exon2

Figure 4

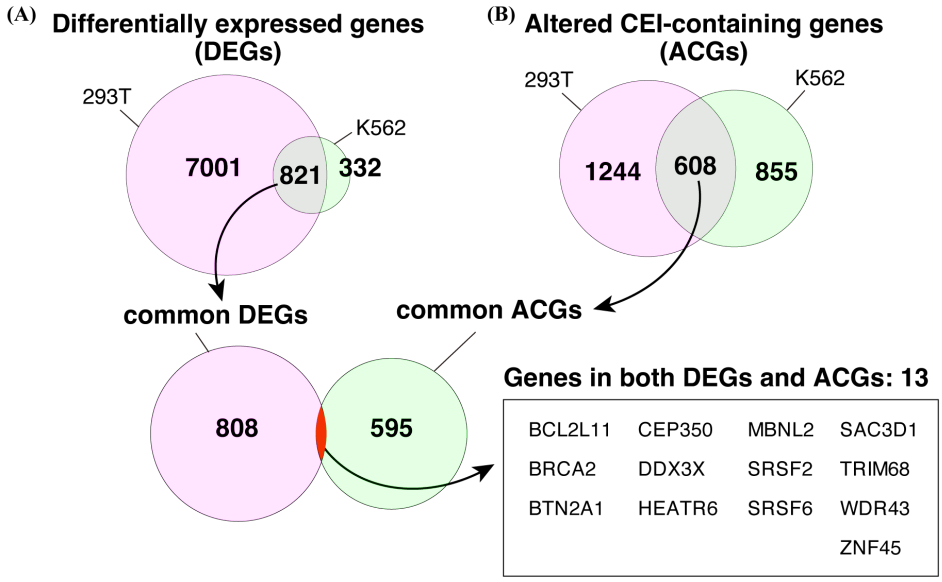
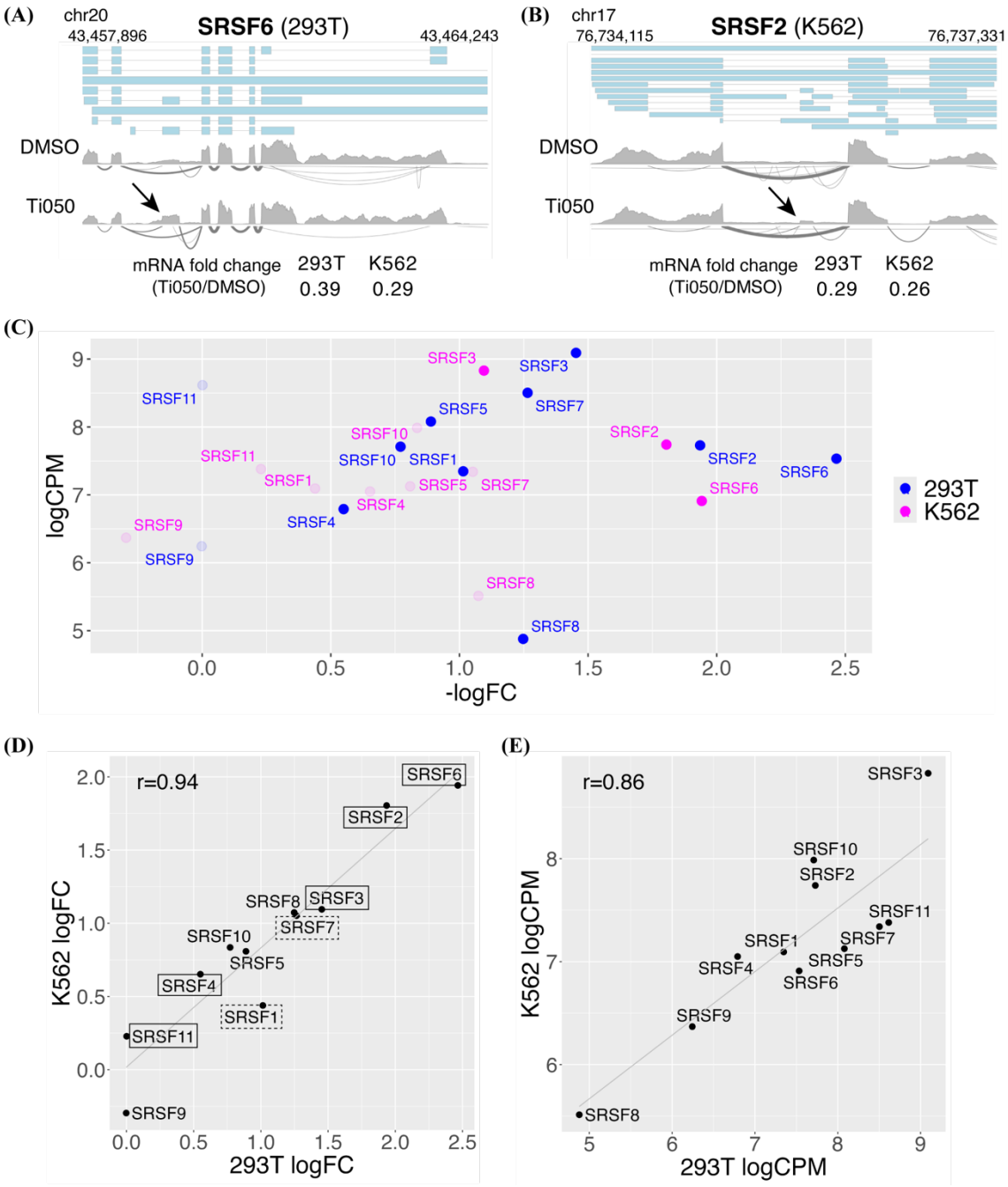

(A)
Differentially expressed genes (DEGs)
293T
K562
7001
821
332
common DEGs
808
595
(B)
Altered CEI-containing genes (ACGs)
293T
K562
1244
608
855
common ACGs
Genes in both DEGs and ACGs: 13
BCL2L11
CEP350
MBNL2
SAC3D1
BRCA2
DDX3X
SRSF2
TRIM68
BTN2A1
HEATR6
SRSF6
WDR43
ZNF45

Figure 5


(A)
chr20
43,457,896
SRSF6 (293T)
43,464,243
DMSO
Ti050
mRNA fold change (Ti050/DMSO)
293T 0.39
K562 0.29
(B)
chr17
76,734,115
SRSF2 (K562)
76,737,331
DMSO
Ti050
mRNA fold change (Ti050/DMSO)
293T 0.29
K562 0.26
(C)
logCPM
-logFC
293T
K562
SRSF3
SRSF11
SRSF5
SRSF7
SRSF10
SRSF2
SRSF1
SRSF6
SRSF4
SRSF9
SRSF8
(D)
r=0.94
K562 logFC
293T logFC
SRSF6
SRSF2
SRSF8
SRSF3
SRSF7
SRSF10
SRSF5
SRSF4
SRSF1
SRSF11
SRSF9
(E)
r=0.86
K562 logCPM
293T logCPM
SRSF3
SRSF10
SRSF2
SRSF11
SRSF1
SRSF7
SRSF4
SRSF5
SRSF6
SRSF9
SRSF8

Supplementary Tables

Ti056 (293T)

rMATS event type	Total event number	Altered events	DMSO	Ti056
CEI	63824	697	257	440
MXE	9991	60	31	29
A5SS	3984	68	30	38
A3SS	5868	78	42	36
RI	3846	113	53	60

Ti050 (293T)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	63354	2645	545	2100
MXE	9572	242	126	116
A5SS	3983	252	155	97
A3SS	5864	320	181	139
RI	3878	1111	1026	85

Ti050 (K562)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	41091	2041	447	1594
MXE	5145	126	73	53
A5SS	3282	227	119	108
A3SS	4843	206	106	100
RI	3546	533	420	113

BAY61-3606 (K562)

rMATS event type	Total event number	Altered events	DMSO	BAY61-3606
CEI	40813	1589	328	1261
MXE	3275	130	70	60
A5SS	4886	153	70	83
A3SS	5149	97	55	42
RI	3534	291	183	108

H3B-8800 (K562, GSE94528)

rMATS event type	Total event number	Altered events	DMSO	Ti056
CEI	103028	20340	18672	1668
MXE	21605	3880	2014	1866
A5SS	7041	1324	567	757
A3SS	9930	2081	693	1388
RI	6880	2701	395	2306

E7107 (K562, GSE114558)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	105793	4186	2577	1609
MXE	21266	513	264	249
A5SS	7701	356	139	217
A3SS	11075	463	179	284
RI	7317	554	143	411

SF3B1 K700E Data Set A (K562, GSE187356)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	66236	2419	793	1626
MXE	11080	281	135	146
A5SS	5930	310	133	177
A3SS	8493	603	234	369
RI	6565	843	369	474

SF3B1 K700E Data Set B (K562, GSE94528)

rMATS event type	Total event number	Altered events	DMSO	BAY61-3606
CEI	86975	5229	1817	3412
MXE	16098	805	414	391
A5SS	6900	455	197	258
A3SS	9983	871	380	491
RI	6952	845	395	450

U2AF1 S34F Data Set A (K562, GSE224576)

rMATS event type	Total event number	Altered events	DMSO	Ti056
CEI	76554	7379	4074	3305
MXE	14287	1094	571	523
A5SS	6167	497	276	221
A3SS	8935	1053	664	389
RI	6513	890	342	548

U2AF1 S34F Data Set B (K562, GSE242092)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	123745	5751	3719	2032
MXE	27519	1074	506	568
A5SS	8239	304	145	159
A3SS	11861	858	480	378
RI	7503	678	140	538

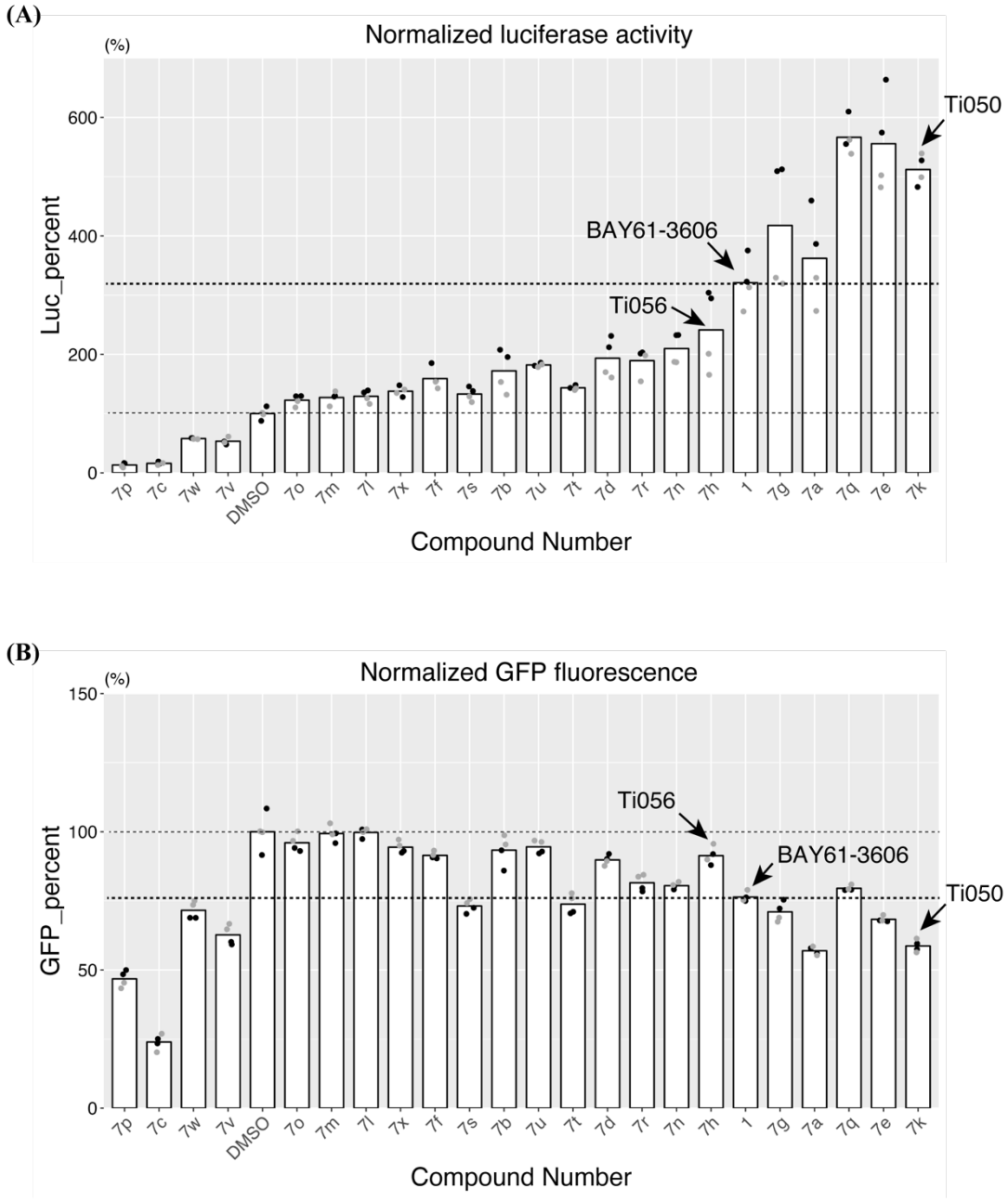
SRSF2 P95H Data Set A (K562, GSE235600)

rMATS event type	Total event number	Altered events	DMSO	Ti050
CEI	75189	904	460	444
MXE	13333	55	30	25
A5SS	6314	106	55	51
A3SS	9161	105	51	54
RI	6551	94	69	25

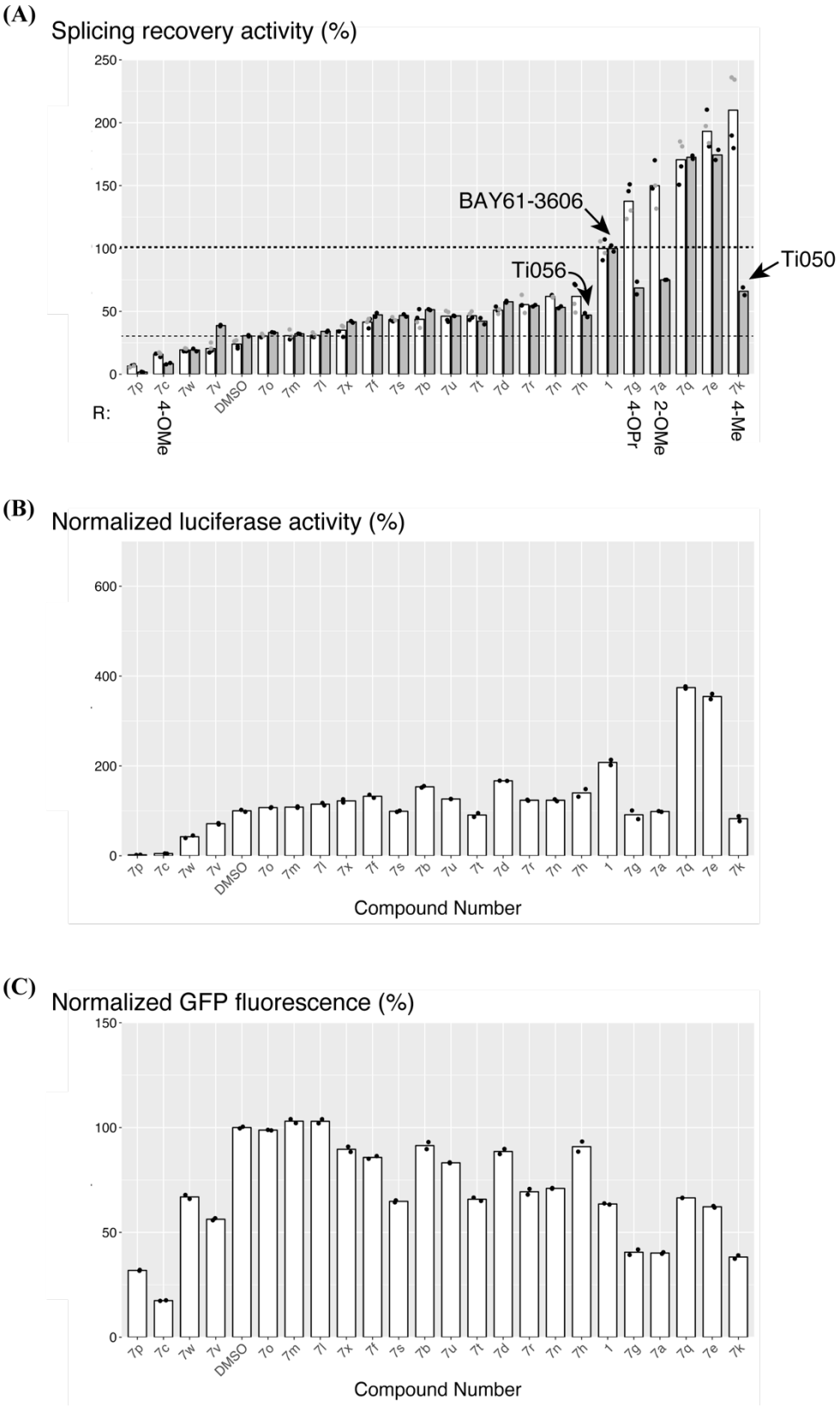
SRSF2 P95H Data Set B (K562, GSE200446)

rMATS event type	Total event number	Altered events	DMSO	BAY61-3606
CEI	145544	7846	4635	3211
MXE	34909	1613	782	831
A5SS	9044	693	356	337
A3SS	13086	954	426	528
RI	7897	875	308	567

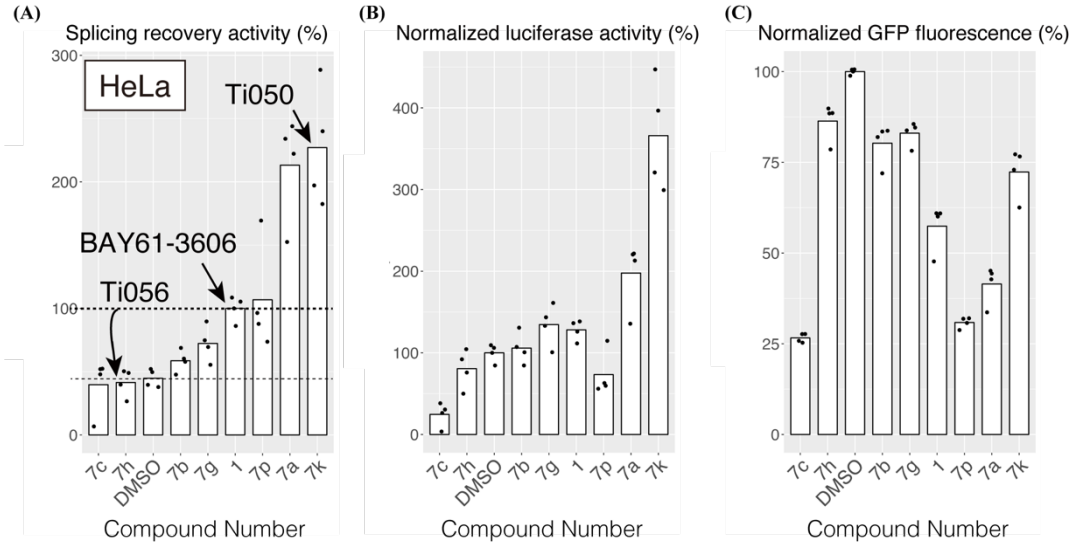
Supplementary Figure 1

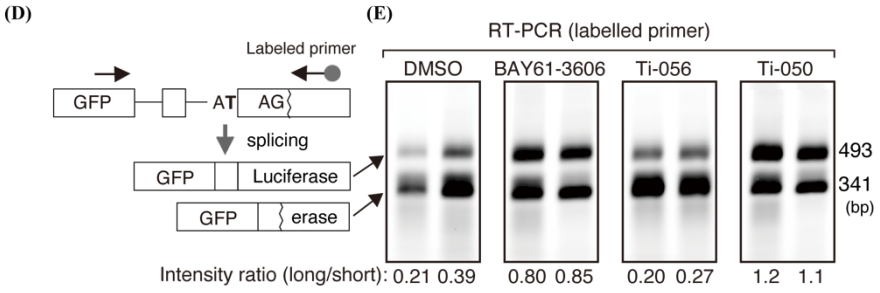


Supplementary Figure 2

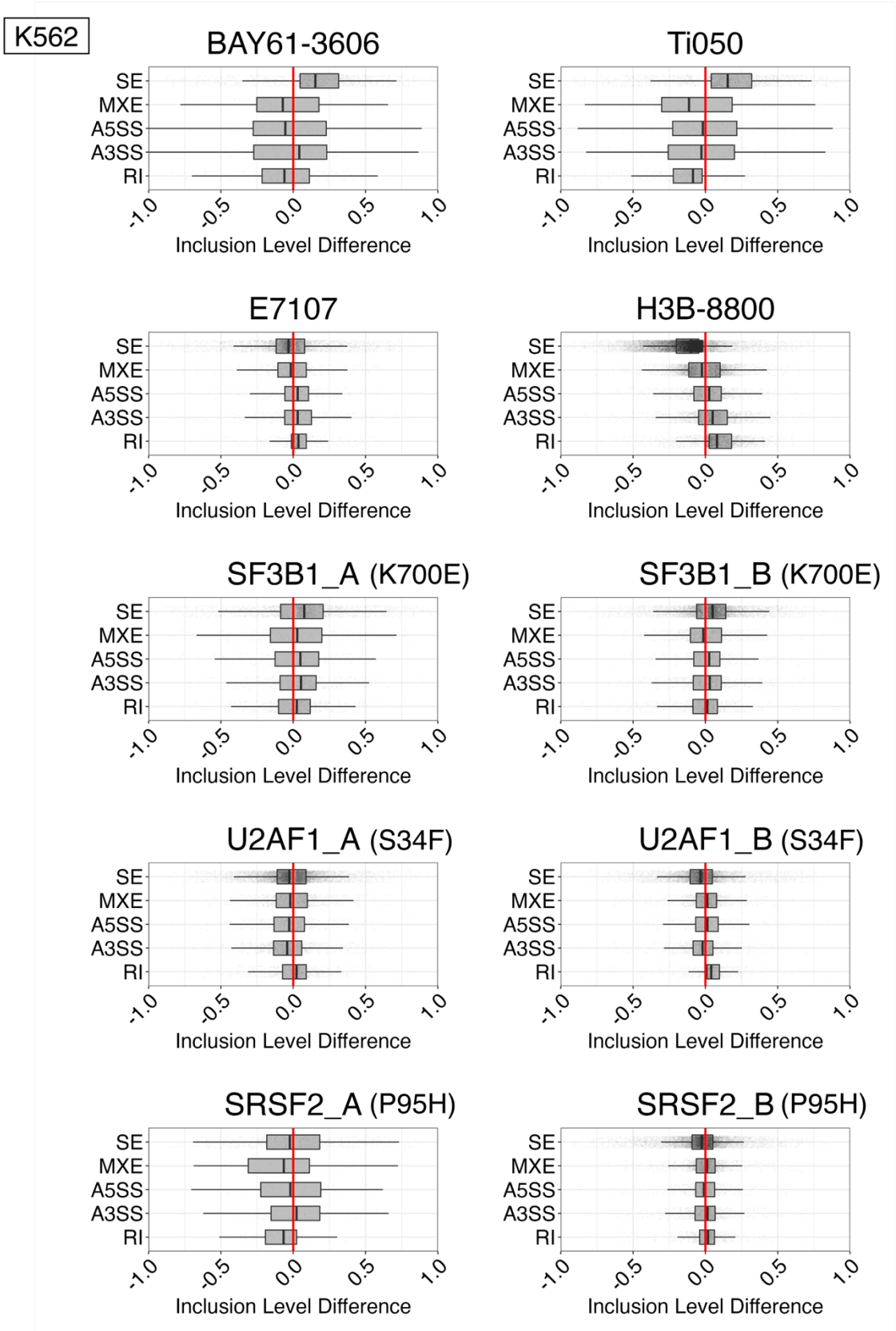


Supplementary Figure 3

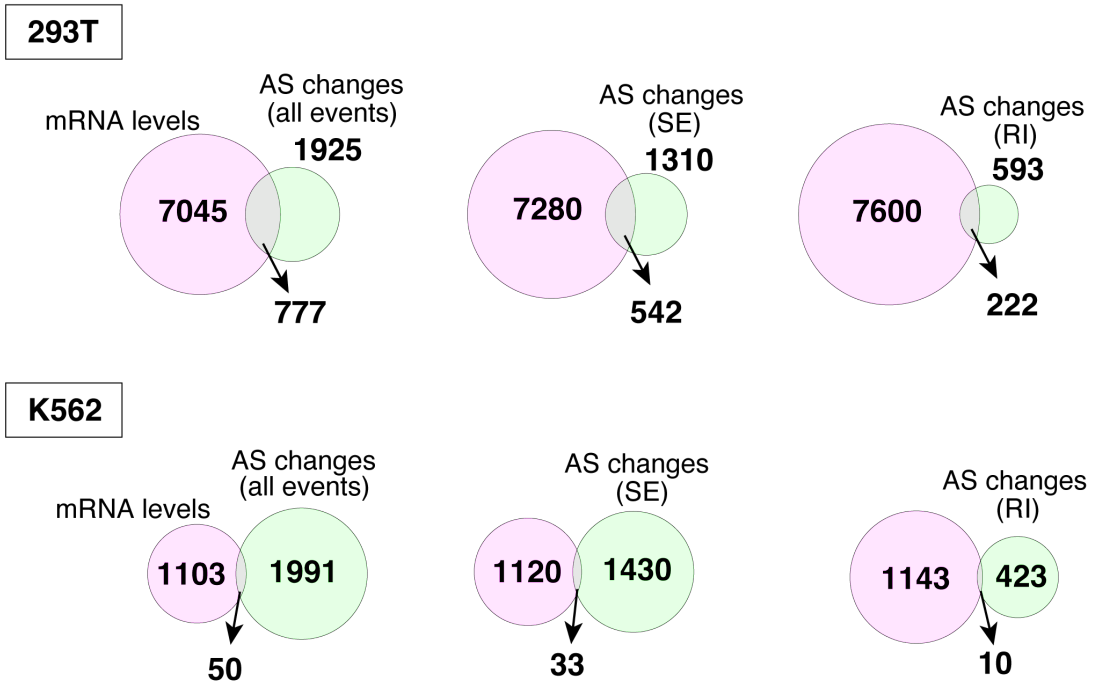




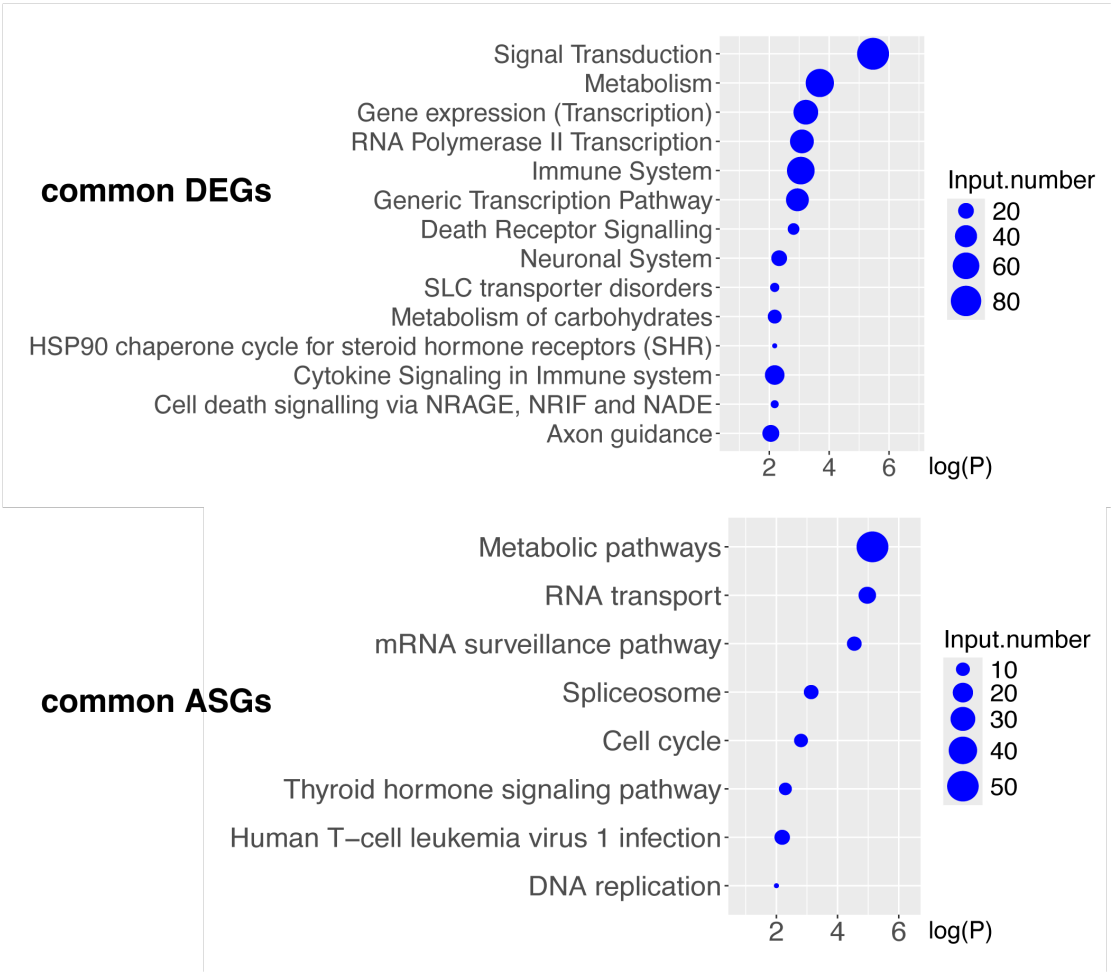
Supplementary Figure 4



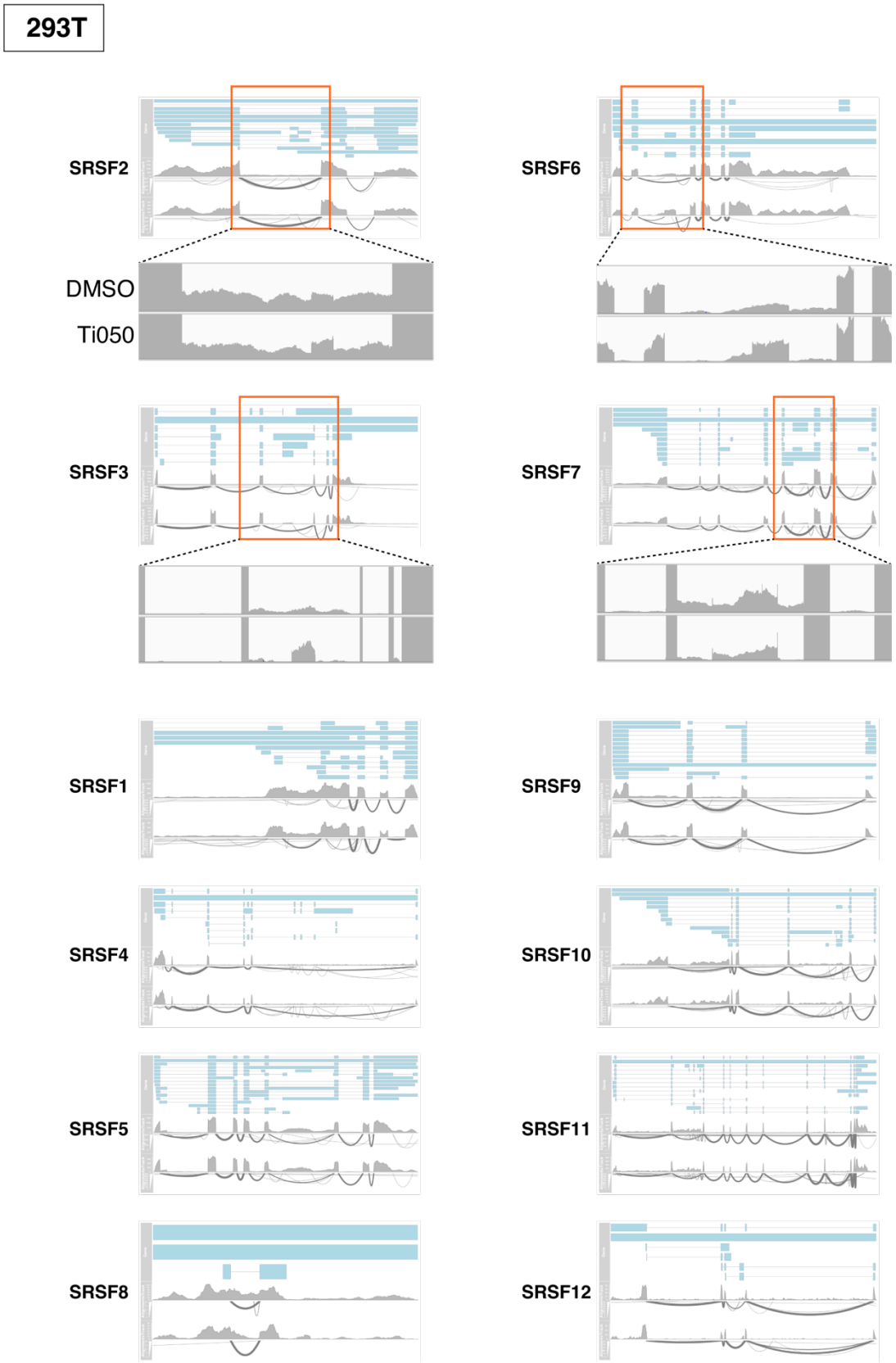
Supplementary Figure 5



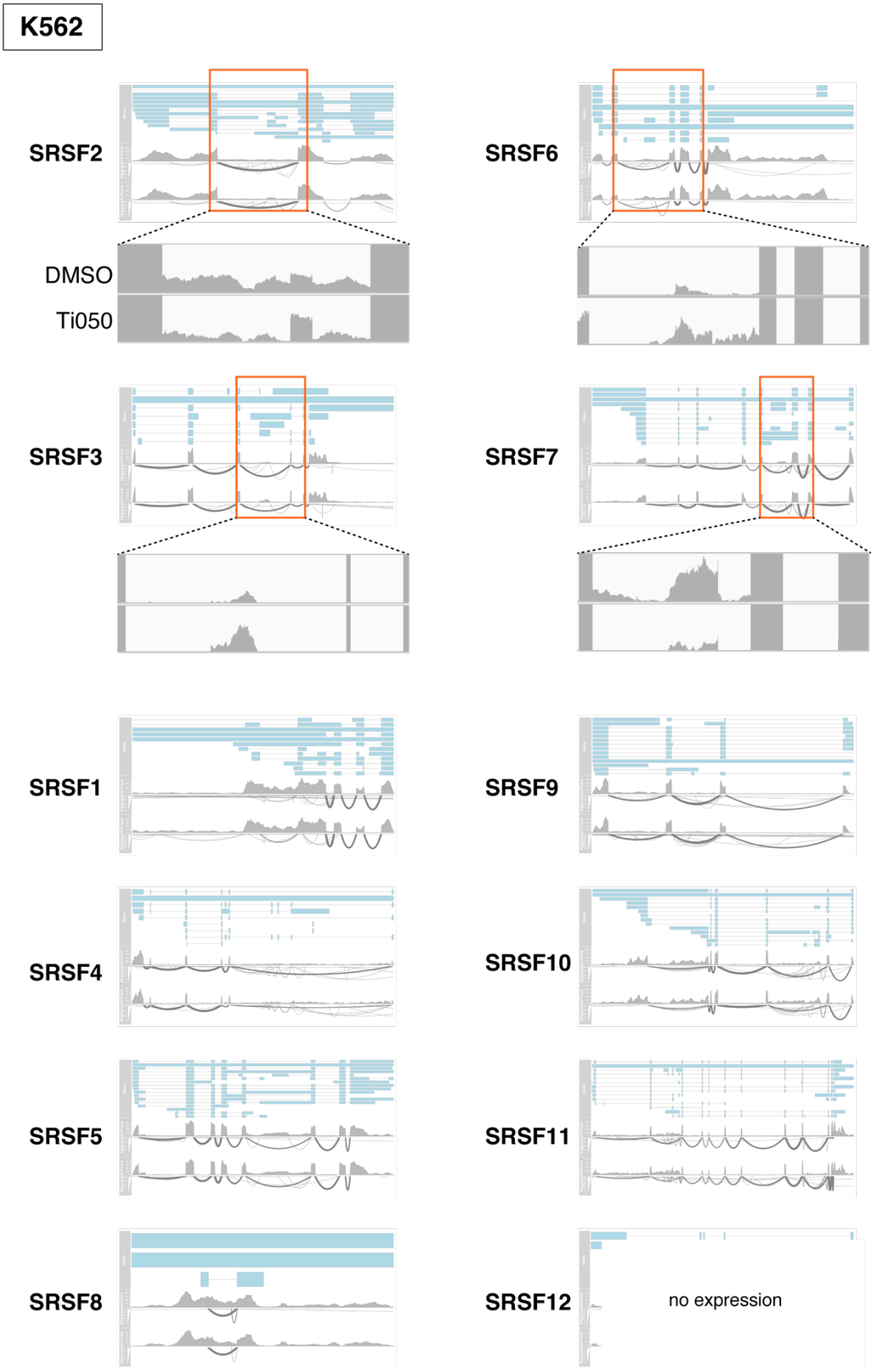
Supplementary Figure 6



Supplementary Figure 7



Supplementary Figure 8



Supplementary Table and Figure Legends

Supplementary Tables

Summary of rMATS analysis of RNA-seq data. rMATS detected differential splicing events based on the provided annotation files. These events are classified into five categories: CEI (Cassette exon inclusion; although the original name in the rMATS output is SE (Skipped exon), we renamed this category to make the direction of changes easier to understand), MXE (Mutually exclusive exon), A5SS (Altered 5' splice site), A3SS (Altered 3' splice site), and RI (Retained intron). The total number of detected events for each dataset is shown as "total events number", while the number of statistically significant events is indicated as "altered events". Altered events are further classified based on the direction of the changes. For example, if there are two columns named DMSO and Ti050 with CEI values of 545 and 2100, these numbers represent the number of genes with increased cassette exon inclusion in DMSO and Ti050, respectively. For MXE, A5SS and A3SS, the numbers indicate genes with increased usage of the upstream splice site. For RI, the numbers represent genes with increased intron retention.

Supplementary Figure 1

The measured luciferase activities (A) and GFP fluorescence (B) derived from the splicing reporter gene, which were used to calculate the splicing recovery activity, are presented as bar plots, with the activity of DMSO set to 100%. Each bar indicates the mean activity of two independent experiments. Each experiment consisted of two technical replicates. The data points from these technical replicates are overlaid as dots with different colors indicating each experiment. A thin dotted line represents the average activity of the vehicle control (DMSO), serving as the standard, while a thick dotted line represents the average activity of BAY61-3606.

Supplementary Figure 2

Occasional reduction of splicing recovery activity in specific compounds. (A) Splicing recovery activities, determined by the splicing reporter, are grouped based on GFP values for comparison. Activities were calculated by normalizing luciferase counts to GFP values. Data sets were divided into two groups: GFP values higher than 50% and lower

than 50% for Ti050. The mean activities are presented as bar plots, with the activity of BAY61-3606 set to 100%. White bars represent the group with higher GFP values, and gray bars represent the group with lower GFP values. The white bars show the mean of four data points obtained from two independent experiments, each consisting of two technical replicates, corresponding to the data shown in Fig. 1B. The thin dotted line represents the average activity of the vehicle control (DMSO) as the standard, while the thick dotted line shows the average activity of BAY61-3606. Four compounds with small substituents at the phenyl ring occasionally showed reduced splicing recovery activities. In experiments where splicing recovery activities were reduced, luciferase activities (B) and GFP values (C) are shown as bar plots, with DMSO activity set to 100%. Data points from technical replicates are overlaid as dots. Both luciferase activities and GFP fluorescence were reduced, leading to an overall decrease in splicing recovery activity.

Supplementary Figure 3

(A) Several compounds were further tested using HeLa cells expressing the splicing reporter gene, treated for 24 hors at 5 μM. The average of technical replicates (n=4) are presented as bars, with the activity of BAY61-3606 set to 100%, and individual data points are overlaid. (B) and (C) are plots showing relative luciferase activities and GFP values normalized against DMSO. (D) Splicing pattern changes were confirmed by RT-PCR analysis. Schematic indicates the position of primers. Isoforms detected by RT-PCR were quantified based on fluorescence from IRD800 dye-labeled at the 3' end of the antisense primer. (E) The longer band represents the isoform spliced at the original mutated site, while the shorter band represents the isoform spliced at the downstream cryptic site. The ratio of each band is shown below each lane to indicate splicing recovery efficiency. Ti050 showed higher efficiency in leading splicing at the mutated 3' splice site.

Supplementary Figure 4

A schematic representation of splicing patterns detected by rMATS for each dataset. The plots for BAY61-3606 and Ti050 are the same as those shown in Fig. 3B, with some adjustments.

Supplementary Figure 5

Venn diagrams reveal that only a small subset of genes overlap between those exhibiting altered transcript levels and those showing altered splicing patterns following Ti050 treatment in both 293T and K562 cells.

Supplementary Figure 6

The summary of the enrichment analysis utilizing GO terms is depicted as a lollipop plot. The circle's size of correlates with the number of genes assigned to each category by GO terms. The name of each biological process is displayed to the left of the plot. The x-axis represents the LogP, which is the logarithm of the enrichment probability value.

Supplementary Figure 7 and 8

Sashimi plots of all SR protein genes, comparing Ti050-treated samples with the vehicle control, visualize the changes in splicing pattern of their transcripts. Red boxes highlight the specific gene regions exhibiting marked changes, with magnified images of these regions presented below. To prepare the magnified images of sashimi plots, IGV was used to generate initial mapped images from BAM files with an auto-scaled y-axis. These images were further magnified by setting a smaller y-axis, while maintaining consistent y-axis ratios for both conditions. The exported SVG images were then clipped and pasted. Plotted regions are presented below, respectively:

SRSF1 chr17:58000919-58007246, SRSF2 chr17: 76734115-76737331, SRSF3 chr6: 36594362-36605600, SRSF4 chr1: 29147743-29181900, SRSF5 chr14: 69767142-69772005, SRSF6 chr20: 43457896-43464243, SRSF7 chr2: 38743599-38751361, SRSF8 chr11: 95066919-95071225, SRSF9 chr12: 120461672-120469748, SRSF10 chr1: 23964347-23980327, SRSF11 chr1: 70221380-70253052, SRSF12 chr6: 89095959-89118071