Contents:

Supplemental Tables

Supplemental Table S1. Clinical and technical information on the samples used. Each sheet lists samples used for: i) Iso-Seq long-read transcriptome sequencing; ii) short-read RNA-seq; iii) iCLIP experiment; and iv) list of public RNA-seq reads downloaded from the Short Read Archive.

Supplemental Table S2. Union of significant alternative splicing events between *SF3B1*$^{mut/wt}$ and *SF3B1*$^{wt/wt}$ detected in cell lines, CLL, or MDS patients.

Supplemental Table S3. Differential expression analysis results using Iso-Seq count reads between MDS and CLL samples.

Supplemental Table S4. Significant alternative splicing events (ASEs) detected in merged dataset using two methods: Iso-Seq and IsoTools, as well as RNA-seq and rMATS. The table contains gene overrepresentation analysis of the genes with ASEs and the summary of the two methods comparison.

Supplemental Table S5. Oligonucleotide sequences used in the validation and minigene experiments.

Supplemental Table S6. SF3B1 binding regions detected with the regions' position and class.

Supplemental Notes

SF3B1 mutation effect depends on the distance and sequence context of alternative 3' AS splice sites

We co-transfected HEK293T cells with the minigene constructs and either *SF3B1*^{wt} or *SF3B1*^{K700E} for 48 h and analyzed the expression by semiquantitative RT-PCR.

As expected, with the original intronic sequence from *THOC1*, overexpression of *SF3B1*^{K700E} resulted in an increased usage of AG', both for endogenous *THOC1* and the *THOC1* minigene (lanes 1–4, Fig. 5B). Of note, the insertion of the longer fragments from the *PABCLI* and *USP1* intronic sequences abolished the usage of the alternative splice site for the minigene constructs (lanes 7–8, 11–12, Fig. 5B). Moreover, the insertion of the longer intronic fragment of the *ZNF124* led to the sole usage of AG' and increased intron retention (lanes 15–16, Fig. 5B). This could be explained by the weak canonical AG strength compared to the alternative AG' (0.72 vs. 0.95, as calculated with SpliceRover (Zuallaert et al. 2018), Fig. 5C). Importantly, no difference in splice site usage was observed when comparing *SF3B1*^{K700E} and *SF3B1*^{wt} overexpressing cells for each of the constructs tested. These results suggested that increasing the AG'–AG was sufficient to remove the 3'AS from SF3B1 regulation.

Next, we asked if the AG'–AG distance was the main factor that influenced the 3'AS and created minigene assays with shorter, approx. 20nt versions of the previously used constructs with preserved polypyrimidine (Py) tracts. Indeed, for the *PABCL1*-Py construct the usage of AG' was stronger upon overexpression of *SF3B1*^{K700E} compared to *SF3B1*^{wt} (lanes 9–10, Fig. 5B). This indicated that the AG'–AG distance rather than the specific sequence of this construct was determinant for the SF3B1 regulation. This was different when assessing the shortened versions of the other two constructs.

The *USP1*-Py construct was spliced mainly at the canonical AG and to a little extent at AG', but also led to IR (lanes 13–14, Fig. 5B). The *ZNF124*-Py construct was only weakly spliced, but did not show any preference towards AG or AG' and resulted in increased IR instead (lanes 17–18, Fig. 5B). Interestingly, in this instance again the canonical AG was weak (0.498, Fig. 5C). Neither the *USP1*-Py nor the *ZNF124*-Py constructs were differentially spliced when comparing *SF3B1*^{wt} and *SF3B1*^{K700E} overexpressing cells.

We therefore hypothesized that not only the distance, but also the sequence content between AG' and AG plays a role for the alternative splicing in *SF3B1*$^{mut/wt}$. To test this, we replaced the short AG'–AG region of the *THOC1* minigene with short AG'–AG fragments (12–21 nt) from three non-differentially spliced 3'AS events (in *GPR98B*, *UROD*, and *CELF2*). Interestingly, we noticed differential 3'AS usage between *SF3B1*^{K700E} and *SF3B1*^{wt} in two of the three constructs tested, indicating that the Py-tract of *GPR98* and *UROD* surrounded by *THOC1* sequences resulted in increased *THOC1* AG' usage in *SF3B1*^{mut}-expressing cells (lanes 19–22, Fig. 5B). Most likely *THOC1* specific sequences, such as the branch point region upstream of AG' were responsible for the differential splicing between *SF3B1*^{wt} and *SF3B1*^{K700E} expressing cells.

In contrast, the *CELF2* insert led to a slight usage of AG' and an increased IR in both, *SF3B1*^{wt} and *SF3B1*^{K700E} expressing cells (lanes 23–24, Fig. 5B). As for the weakly spliced *ZNF*-Py construct, also the

CELF2 construct had a weak canonical AG (0.301, Fig. 5C). This suggested that a strong AG is required for AG' usage, in line with previously published work (Darman et al. 2015).

Supplemental Methods

Validation of cell lines

Cell line authenticity was tested by short tandem repeat (STR) profiling at the Cologne Center for Genomics, Cologne, Germany with 13 markers: *CSF1PO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D19S433*, *D21S11*, *THO1*, *TPOX*, *vWA* and *AMELOGENIN* for sex determination. Subsequently, STR results were analyzed with the Cellosaurus STR similarity search tool CLASTR 1.4.4 and compared to the Cellosaurus data set 46.0 (algorithm: Tanabe, mode: non-empty markers, *AMELOGENIN* not included) run on 20230724. HEK293-FT and the K562 cell line pair were correctly assigned and the Nalm6 cell line pair matched 100 % to the Nalm6 derived cell lines Nalm6/H (RRID:CVCL-B7AM) and Nalm6/HDR (RRID:CVCL_B7AN) (Li et al., 2021), compared to 83 or 81 % to Nalm6 (RRID:CVCL_0092). Moreover, the Nalm6 cell line pair differed at two marker loci (Supplemental Table S6). K562 cells were cultivated in Iscove Modified Dulbecco Media (IMDM) supplemented with 2 mM L-glutamine (Sigma-Aldrich), Nalm6 cells in RPMI-1640 (Thermo Fisher Scientific, #21875034) and HEK293-FT in DMEM, high Glucose, GlutaMAX™ (Thermo Fisher Scientific, 10569010). All media were supplemented with 10% fetal bovine serum (SIGMA-Aldrich) and 100 U/ml penicillin/streptomycin (Sigma-Aldrich). Cells were cultivated at 37°C with 5% CO_2 and 95% humidity. Cell lines were tested negative for *Mycoplasma* using the Mycoalert Plus Mycoplasma detection kit (Lonza) according to the manufacturer's protocol.

Transcript coding potential

We estimated coding potential with Coding-Potential Assessment Tool (CPAT) (Wang et al., 2013). An ORF was deemed to possess a translation initiation if its 5' intron chain coincided with that of an annotated protein-coding transcript and the corresponding annotated CDS began at that exact same location. Moreover, we predicted additional translation initiation for the first ORF that fulfilled criteria we inferred from the distribution of the annotated CDS (Supplemental Fig. S25) i.e.,: i) an ORF spanned a length of ≥ 300 nt, (ensuring a CDS length of ≥ 300 nt); ii) the ORF was situated within the initial 500 nt of the transcript (thus ensuring a 5'UTR length of ≤ 500 nt); iii) and the similarity to the Kozak consensus sequence was at least moderate (log odds ratio > -2.5), ensuring a favorable initiation sequence (Kozak, 1987) context. Furthermore, the potential for nonsense-mediated mRNA decay (NMD) was assessed for all ORFs, using the 55 nt rule (Popp et al.), wherein transcripts with a termination codon located more than 55 nt upstream of the last exon-exon junction are likely subjected to NMD (Popp & Maquat, 2016).

SF3B1-mRNA model generation

The structure of SF3B1 was taken from the cryo-electron microscopy (cryo-EM) structure of the human activated spliceosome (PDB ID: 5Z56 (Zhang et al., 2018)). Besides SF3B1 and the pre-mRNA bound to SF3B1, proteins interacting with SF3B1 and the bound pre-mRNA (RNA-binding motif protein, x-linked 2 (RBMX2); splicing factor 3A subunit 2 (SF3A2); PHD finger-like domain-containing protein 5A (PHF5A); cell division cycle 5-like protein (CDC5L); and the U2-snRNA) were also extracted from the structure. Only structural information resolved in the PDB was considered for the final model except for the RNA-binding motif protein, x-linked 2. There, to also include the N-terminal part in proximity to the

pre-mRNA bound to SF3B1, structural information for the amino acids 1–7 and 113–140 was taken from the cryo-EM structure of the activated human minor spliceosome (PDB ID: 7DVQ (Bai et al., 2021)). The RBMX2 from PDB ID 7DVQ was aligned to the RBMX2 in our model using UCSF Chimer (Pettersen et al., 2004). Atoms not resolved in our model were included into our model, missing bonds within the RBMX2 were created, and a rotamer of K8 not overlapping with the N-terminus was selected using Schrödinger Maestro, v. 2022-4. Amino acids 82–89, 173–190, 205–259, and 311–335 of the SF3B1 not connected to the rest of the protein and not part of pre-mRNA binding were removed. The protein was protonated, missing side chains were created, and N-termini and C-termini of chains were capped with ACE or NME using Schrödinger Maestro, v. 2022-4 (*Schrödinger Release 2023-4: Maestro, Schrödinger, LLC, New York, NY, 2023.*, n.d.). Missing loops (amino acids 450–452, and 486–489 in SF3B1 and 36–44 in splicing factor 3A subunit 2) were modelled using a *de novo* loop generation approach, as implemented in Maestro, v. 2022-4 (*Schrödinger Release 2023-4: Maestro, Schrödinger, LLC, New York, NY, 2023.*, n.d.). All side chains of the pre-mRNA were removed. Because side chain creation using Maestro led to unsatisfactory results with incomplete bonds within prolines (Supplemental Fig. S42), all proline side chains were further minimized with CCG MOE2022.02 (*Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.*, n.d.) using default parameters, resulting in the final model used for all pre-mRNA transcripts Supplemental Fig. S43). For the K700E mutant, the lysine was mutated to glutamate using CCG MOE2022.02 and the most favored rotamer according to MOE was selected. For each RNA construct, missing side chains of the pre-mRNA were filled using *tleap*, as implemented in AmberTools22 (Case et al., 2023).

Molecular dynamics simulations

The models were solvated in a capped octahedral water box using OPC (Case et al., 2022) water with a water shell of at least 12 Å around solute atoms and neutralized using sodium ions. The AMBER22 package of molecular simulations software (Case et al., 2005, 2022) was used in combination with the *ff19SB* force field for protein atoms (Maier et al., 2015) and the RNA OL3 force field (Zgarbová et al., 2011) for RNA atoms. MD simulations were performed as described earlier (Pauly et al., 2022). In short, a combination of steepest descent and conjugate gradient minimization was performed while lowering positional restraints on solute atoms from 25 kcal mol^{-1} $Å^{-2}$ to zero, followed by a stepwise heating procedure and lowering harmonic restraints on solute atoms from 1 kcal mol^{-1} $Å^{-2}$ to zero. For each RNA-model complex (downstream BP / $SF3B1^{wt}$; alternative upstream BP' / $SF3B1^{wt}$; downstream BP / $SF3B1^{K700E}$; alternative upstream BP' / $SF3B1^{K700E}$) four replicas of 200 ns length each were performed; resulting in 16 simulations per transcript and a total simulation time of 64 μs (20 * 4 * 4 * 200 ns).

We selected 14 transcripts which were differentially spliced between *SF3B1*$^{mut/wt}$ and *SF3B1*$^{wt/wt}$ (*PRPF38A*, *RWDD4*, *SLC3A2*, *THOC1*, *TNPO3*, *SEPTIN2*, *IMMT*, *FDPS*, *INTS13*, *CDC27*, *PHKB*, *EIF4B*, *TRIP12*, and *LETMD1*) and 6 non-differentially spliced transcripts (*NAPG*, *SNX13*, *RIC8A*, *PIGB*, *PDCD4*, and *RMDN1*). The 20 transcripts were selected based on: overlap with iCLIP binding site (DoubleNarrow or DoubleWide); median Iso-Seq read coverage ≥ 10; AG'–AG distance ≤ 50 nt; both variants not predicted to undergo NMD; and q-value for differential splicing between *SF3B1*$^{mut/wt}$ and *SF3B1*$^{wt/wt}$ of < 0.01 for differentially spliced and > 0.999 for non-differentially spliced transcripts. These filters were applied before the latest version of the protein functionality prediction was implemented in the

IsoTools and four of the differentially spliced, as well as three non-differentially spliced transcripts did no longer fulfilled these criteria for the following reasons: longer isoform predicted for NMD: *PHKB, PDCD4*; shorter isoform predicted for NMD: *RWDD4*, *TNPO3*, *SEPTIN2*, *RIC8A*, *NAPG* (in addition, q-value for *NAPG* decreased to 0.998).

For each transcript, we performed four replicas of 200 ns long MD simulations of the mRNA with: i) the BP of the downstream AG (BP) bound to $SF3B1^{wt}$; ii) the BP of the upstream AG (BP') bound to $SF3B1^{wt}$; iii) the BP bound to the $SF3B1^{K700E}$ mutant; and iv) the BP' bound to the $SF3B1^{K700E}$ mutant. During the MD simulations, SF3B1 remained structurally invariant (Supplemental Fig. S28–S32).

BP binding between all transcript–protein combinations was analysed as the number of contacts of all heavy atoms of the BP adenine to the heavy atoms of aromatic amino acids surrounding this nucleobase (first binding pocket): phenylalanine (F) 1153, and tyrosine (Y) 1157 (both located in SF3B1), or Y36 (located in PHF5A).

In the SF3B1 wild-type, K700 interacts with the negatively charged oxygens of the mRNA backbone (second binding pocket). The binding at the second binding pocket was analyzed as the frequency of contacts between all heavy atoms of the SF3B1 residue 700 and all mRNA heavy atoms.

For temperature control the Berendsen thermostat with a collision frequency of 10 ps^{-1} was used. Final simulations were analyzed using CPPTRAJ(Roe & Cheatham, 2013) from AmberTools (Case et al., 2023). Contacts between residues were considered if the distance was < 4 Å, as done previously (Gopalswamy et al., 2022). A cutoff of 4 Å was used to describe salt bridges (Barlow & Thornton, 1983). For each frame, the backbone (CA, C, N) of SF3B1 was fitted to the first frame within the respective simulation before computing the RMSF.

References

Bai, R., Wan, R., Wang, L., Xu, K., Zhang, Q., Lei, J., & Shi, Y. (2021). Structure of the activated human minor spliceosome. *Science*, *371*(6535), eabg0879. https://doi.org/10.1126/science.abg0879

Barlow, D. J., & Thornton, J. M. (1983). Ion-pairs in proteins. *Journal of Molecular Biology*, *168*(4), 867–885. https://doi.org/https://doi.org/10.1016/S0022-2836(83)80079-5

Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Berryman, J. T., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cisneros, G. A., Cruzeiro, V. W. D., Darden, T. A., Forouzesh, N., Giambaşu, G., Giese, T., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, J., Izadi, S., … Kollman, P. A. (2022). Amber 2022. *University of California, San Francisco.*

Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., Forouzesh, N., Giese, T. J., Götz, A. W., Gohlke, H., Izadi, S., Kasavajhala, K., Kaymak, M. C., King, E., Kurtzman, T., Lee, T.-S., Li, P., Liu, J., Luchko, T., … Merz, K. M. Jr. (2023). AmberTools. *Journal of Chemical Information and Modeling*, *63*(20), 6183–6191. https://doi.org/10.1021/acs.jcim.3c01153

Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz Jr., K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688. https://doi.org/https://doi.org/10.1002/jcc.20290

Gopalswamy, M., Kroeger, T., Bickel, D., Frieg, B., Akter, S., Schott-Verdugo, S., Viegas, A., Pauly, T., Mayer, M., Przibilla, J., Reiners, J., Nagel-Steger, L., Smits, S. H. J., Groth, G., Etzkorn, M., & Gohlke, H. (2022). Biophysical and pharmacokinetic characterization of a small-molecule inhibitor of RUNX1/ETO tetramerization with anti-leukemic effects. *Scientific Reports*, *12*(1), 14158. https://doi.org/10.1038/s41598-022-17913-6

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, *15*(20), 8125–8148. https://doi.org/10.1093/nar/15.20.8125

Li, Y., Zuo, C., & Gu, L. (2021). Characterization of a novel glucocorticoid-resistant human B-cell acute lymphoblastic leukemia cell line, with AMPK, mTOR and fatty acid synthesis pathway inhibition. *Cancer Cell International*, *21*(1), 623. https://doi.org/10.1186/s12935-021-02335-7

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, *11*(8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

Molecular Operating Environment (MOE), 2022.02 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023. (n.d.).

Pauly, T., Bolakhrif, N., Kaiser, J., Nagel-Steger, L., Gremer, L., Gohlke, H., & Willbold, D. (2022). Met/Val129 polymorphism of the full-length human prion protein dictates distinct pathways of amyloid formation. *Journal of Biological Chemistry*, *298*(10). https://doi.org/10.1016/j.jbc.2022.102430

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/https://doi.org/10.1002/jcc.20084

Popp, M. W., & Maquat, L. E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell*, *165*(6), 1319–1322. https://doi.org/https://doi.org/10.1016/j.cell.2016.05.053

Roe, D. R., & Cheatham, T. E. I. I. I. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, *9*(7), 3084–3095. https://doi.org/10.1021/ct400341p

Schrödinger Release 2023-4: Maestro, Schrödinger, LLC, New York, NY, 2023. (n.d.).

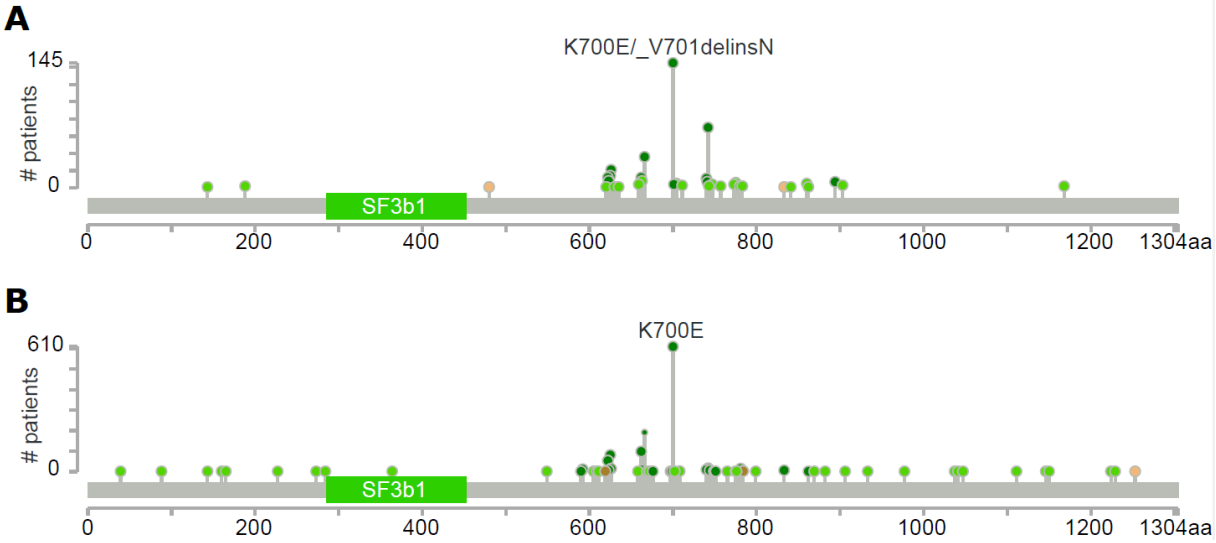
Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., & Li, W. (2013). CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucl Acids Res*, *41*. https://doi.org/10.1093/nar/gkt006

Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T. E. I. I. I., & Jurečka, P. (2011). Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation*, *7*(9), 2886–2902. https://doi.org/10.1021/ct200162x

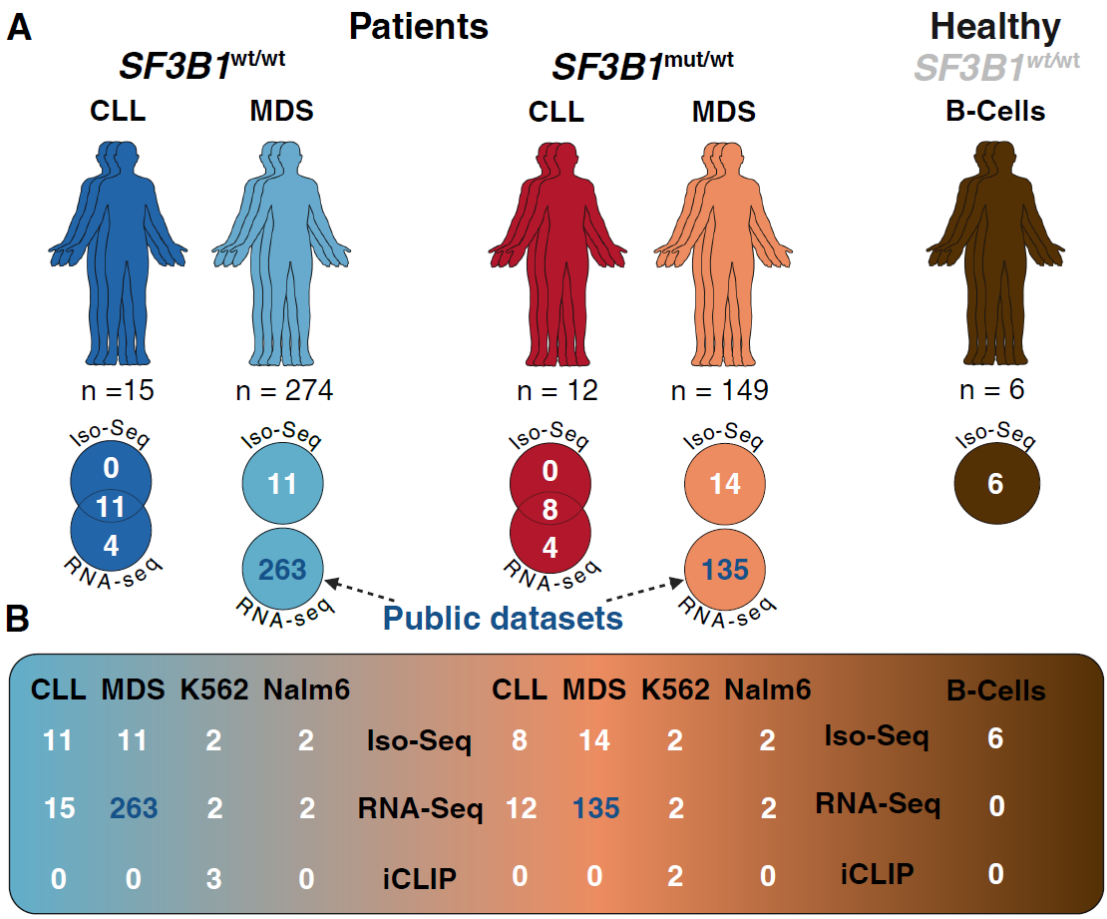
Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., & Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Research*, *28*(3), 307–322. https://doi.org/10.1038/cr.2018.14

Supplemental Figures

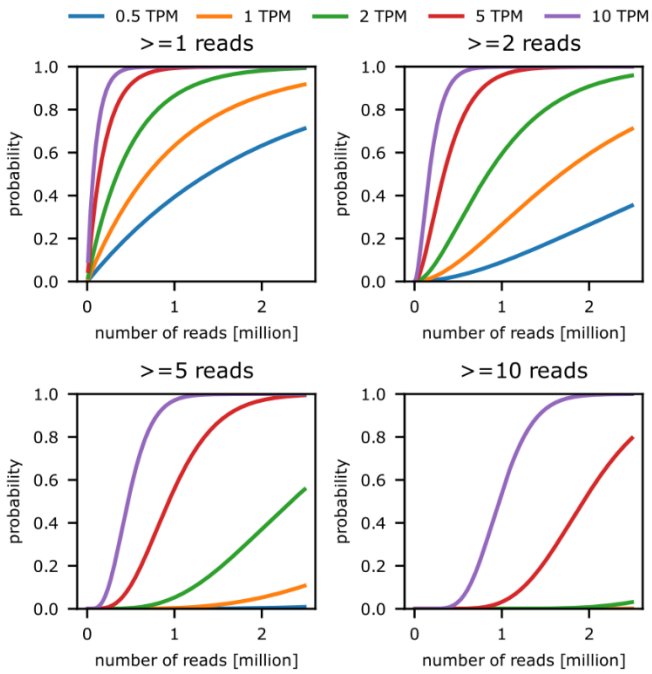
Patients'



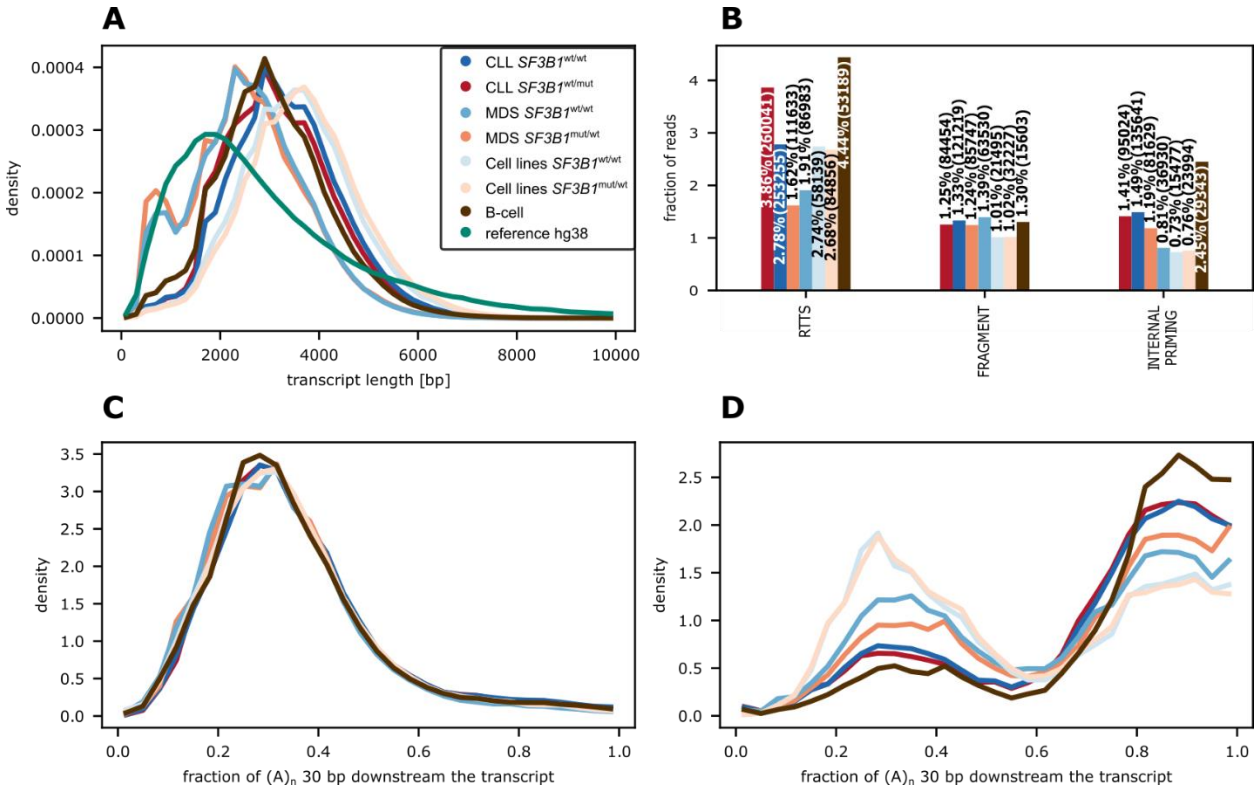
Supplemental Fig. S1. *SF3B1* mutation distribution in CLL (***A***) and MDS (***B***) patients from the cBioPortal (www.cbioportal.org accessed on 13th June 2024). The patients' cohorts are from the following studies: i) CLL patients: Broad Cell 2013, Broad Nature 2015, Broad Nature Genetics 2022, IUOPA Nature 2015, ICGC Nature Genetics 2011); ii) MDS patients: UTokyo, Nature 2011, MSK 2020, MDS IWG, IPSSM, NEJM Evidence 2022.



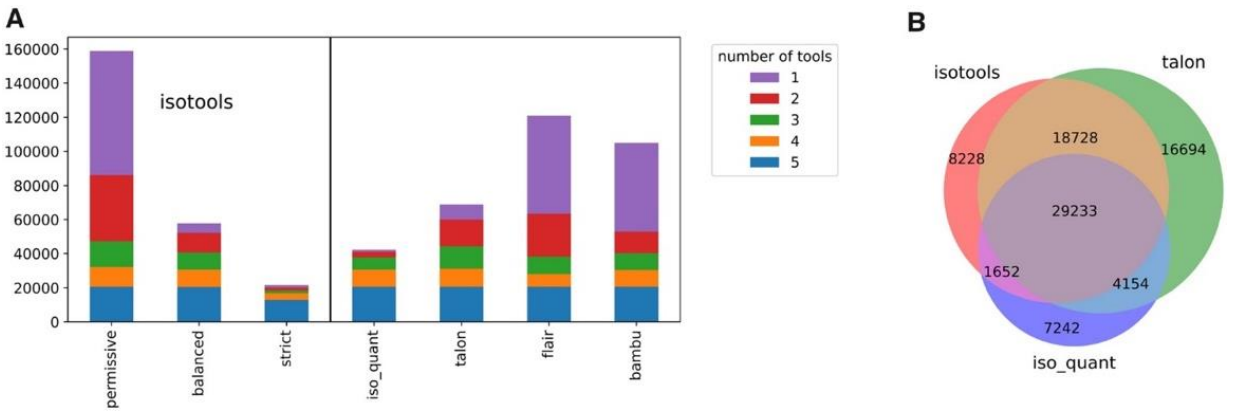
Supplemental Fig. S2. Schematic view of the samples used for this study. (*A*) The total number of patients' samples and the overlap between patients' samples used for Iso-Seq and RNA-seq are shown. Samples obtained from public repositories are marked in green. (*B*) Summary table showing the total number of samples used for each sequencing method. Samples from public repositories (MDS RNA-Seq) are marked in blue font. Figure created with BiorRender.com.



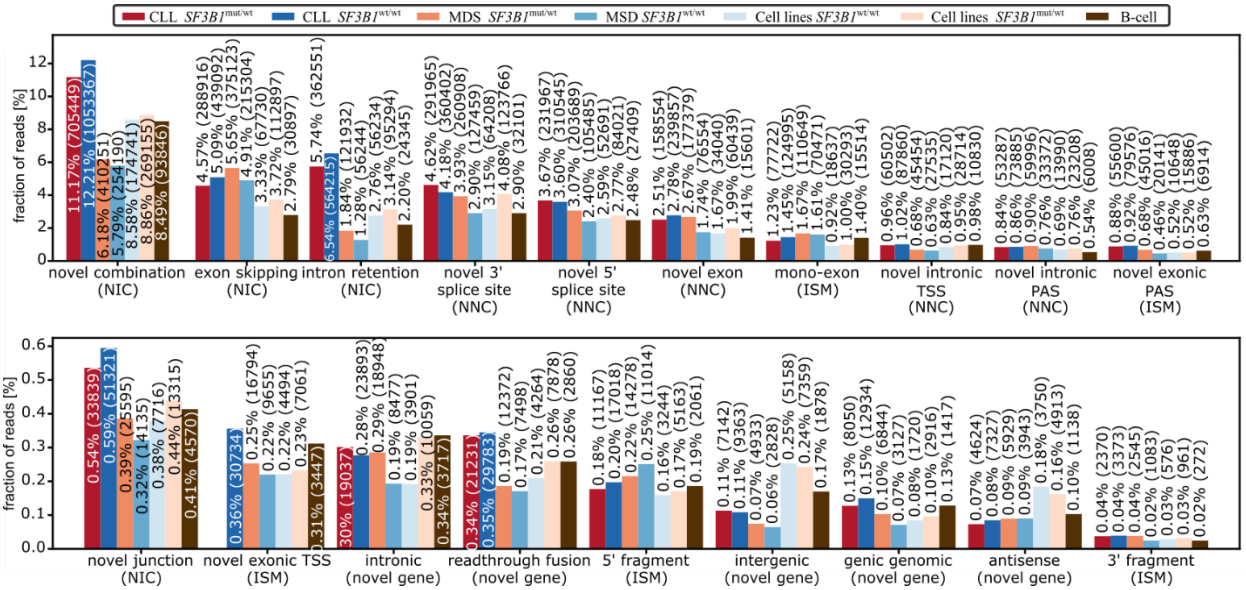
Supplemental Fig. S3. Read saturation analysis in Iso-Seq data analysed. The probability of a transcript expressed at 0.5, 1, 2, 5, or 10 transcripts per million (TPM) to be identified at sequence depth of 0 to 2.5 million reads with at least 1, 2, 5, or 10 reads.



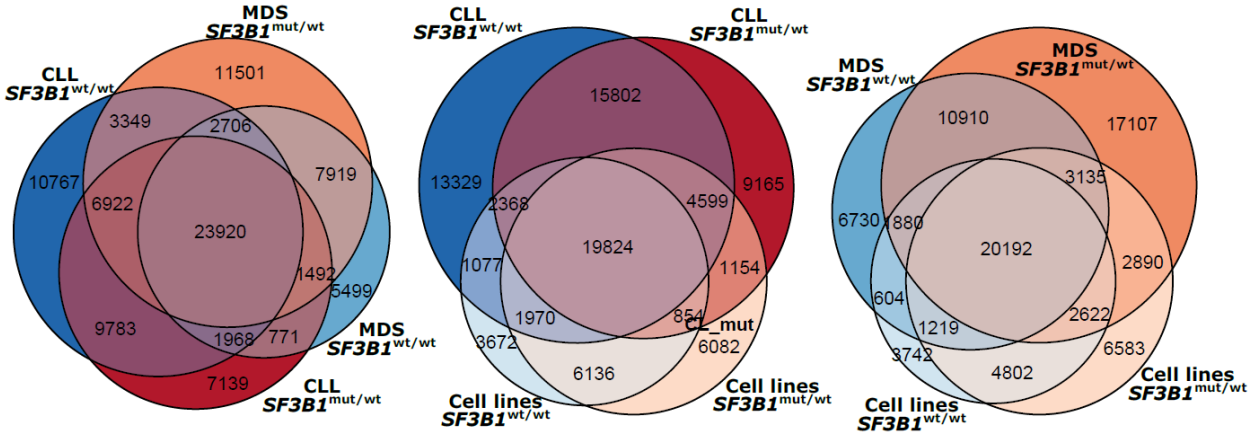
Supplemental Fig. S4. Quality control of the Iso-Seq reads in eight datasets: cell lines (CL), CLL, MDS, each separated by the *SF3B1* mutational status, as well as normal B cells from this and the ENCODE studies. (***A***) Density distribution of transcript length sequenced in each of the sample group. The x axis was cut at 10,000 nt. (***B***) Percentage of reads within each of artefact type per group of samples. (***C***) Density distribution of the fraction of adenosines (A) within 30 bp downstream the transcript within known transcript annotated in GENCODE version 36. (***D***) Same as (*C*) but for novel transcripts.



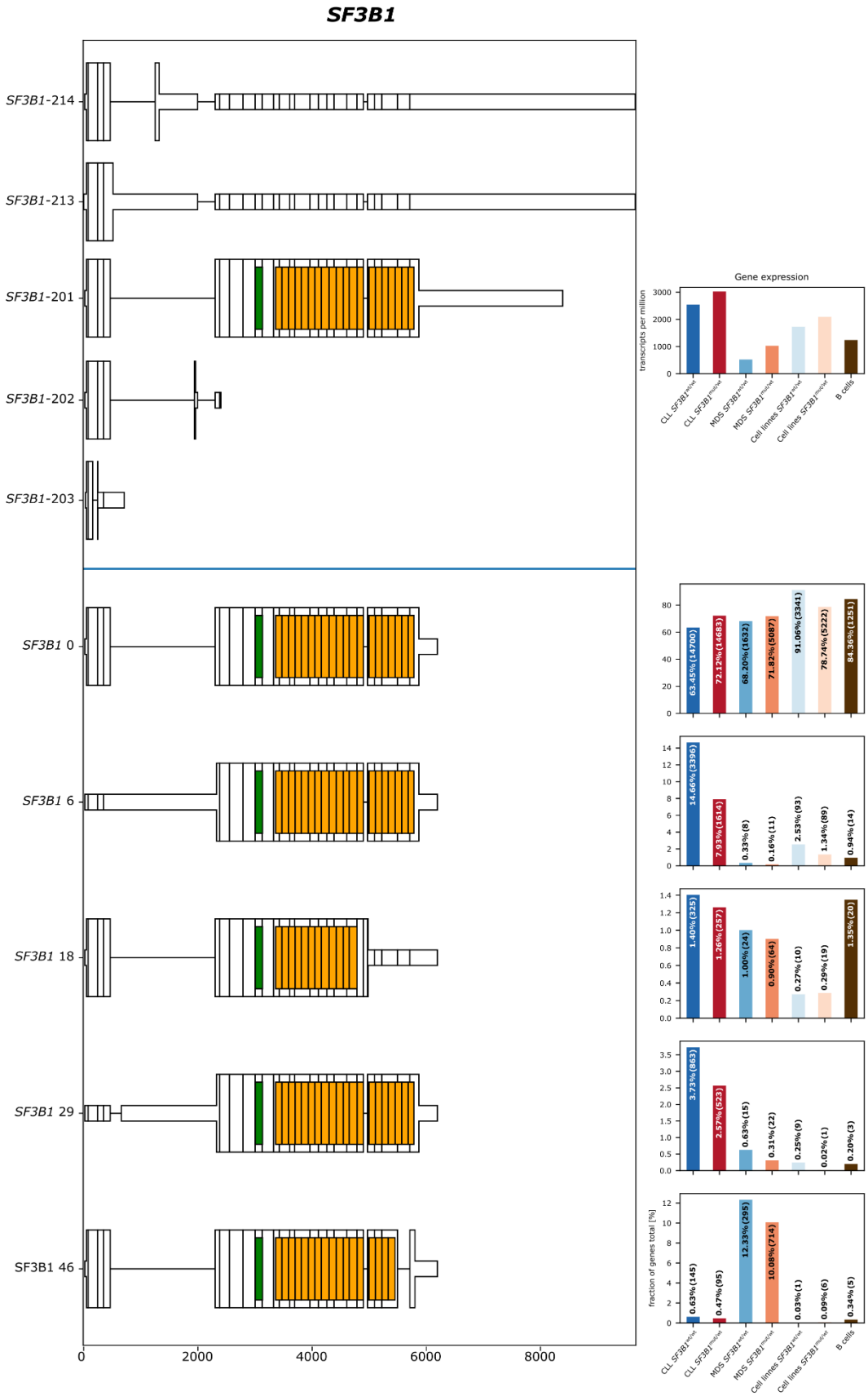
Supplemental Fig. S5. Comparison of recovered transcripts from IsoTools using different filter queries, with alternative tools for transcriptome reconstructions for long reads that participated in a challenge reported in (Pardo-Palacios et al. 2024) and dataset from (Lienhard et al. 2023). IsoQuant, TALON, FLAIR, and Bambu were used with default filtering criteria. (***A***) Barplots (top) depict the number of reported transcripts per software and the color indicates the number of tools reporting a specific transcript. (***B***) Venn diagrams (bottom) depict the overlap between IsoTools with balanced filter, IsoQuant, and TALON. The more stringent filtering leads to the detection of isoforms with stronger overlap to other tools and thus a higher reliability of the resulting transcripts and ASEs to be detected.



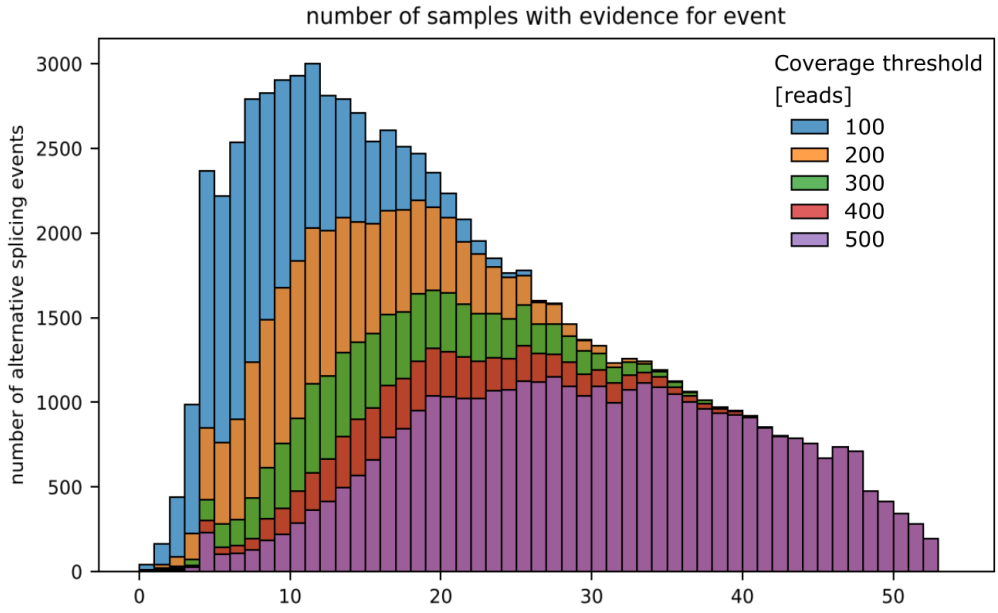
Supplemental Fig. S6. Novel isoforms identified with Iso-Seq separated by transcript type in all group investigated. Only substantial transcripts were used for calculation.



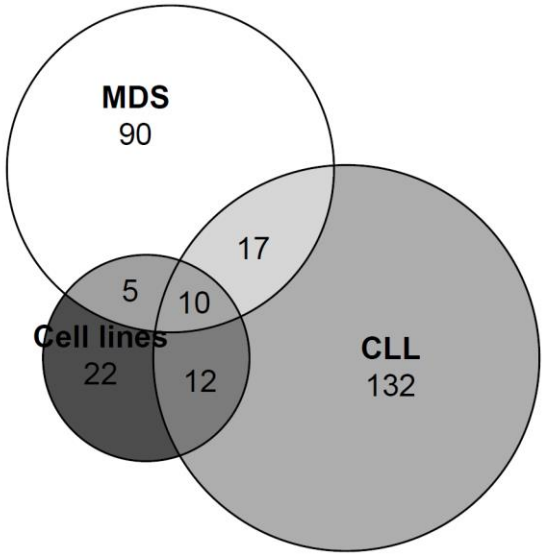
Supplemental Fig. S7. Venn diagrams showing the overlap between transcript expressed in at least one sample per group with ≥1 read coverage.


SF3B1
SF3B1-214
SF3B1-213
SF3B1-201
SF3B1-202
SF3B1-203
SF3B1 0
SF3B1 6
SF3B1 18
SF3B1 29
SF3B1 46
0
2000
4000
6000
8000
Gene expression
transcripts per million
3000
2000
1000
0
CLL SF3B1 mut/wt
CLL SF3B1 mut/mut
MDS SF3B1 mut/wt
MDS SF3B1 mut/mut
Cell lines SF3B1 mut/wt
Cell lines SF3B1 mut/mut
B cells
63.45%(14700)
72.12%(14683)
68.20%(1632)
71.82%(5087)
91.06%(3341)
78.74%(5222)
84.36%(1251)
14.66%(3396)
7.93%(1614)
0.33%(8)
0.16%(11)
2.53%(93)
1.34%(89)
0.94%(14)
1.40%(325)
1.26%(257)
1.00%(24)
0.90%(64)
0.27%(10)
0.29%(19)
1.35%(20)
3.73%(863)
2.57%(523)
0.63%(15)
0.31%(22)
0.25%(9)
0.02%(1)
0.20%(3)
fraction of genes total [%]
0.63%(145)
0.47%(95)
12.33%(295)
10.08%(714)
0.03%(1)
0.09%(6)
0.34%(5)

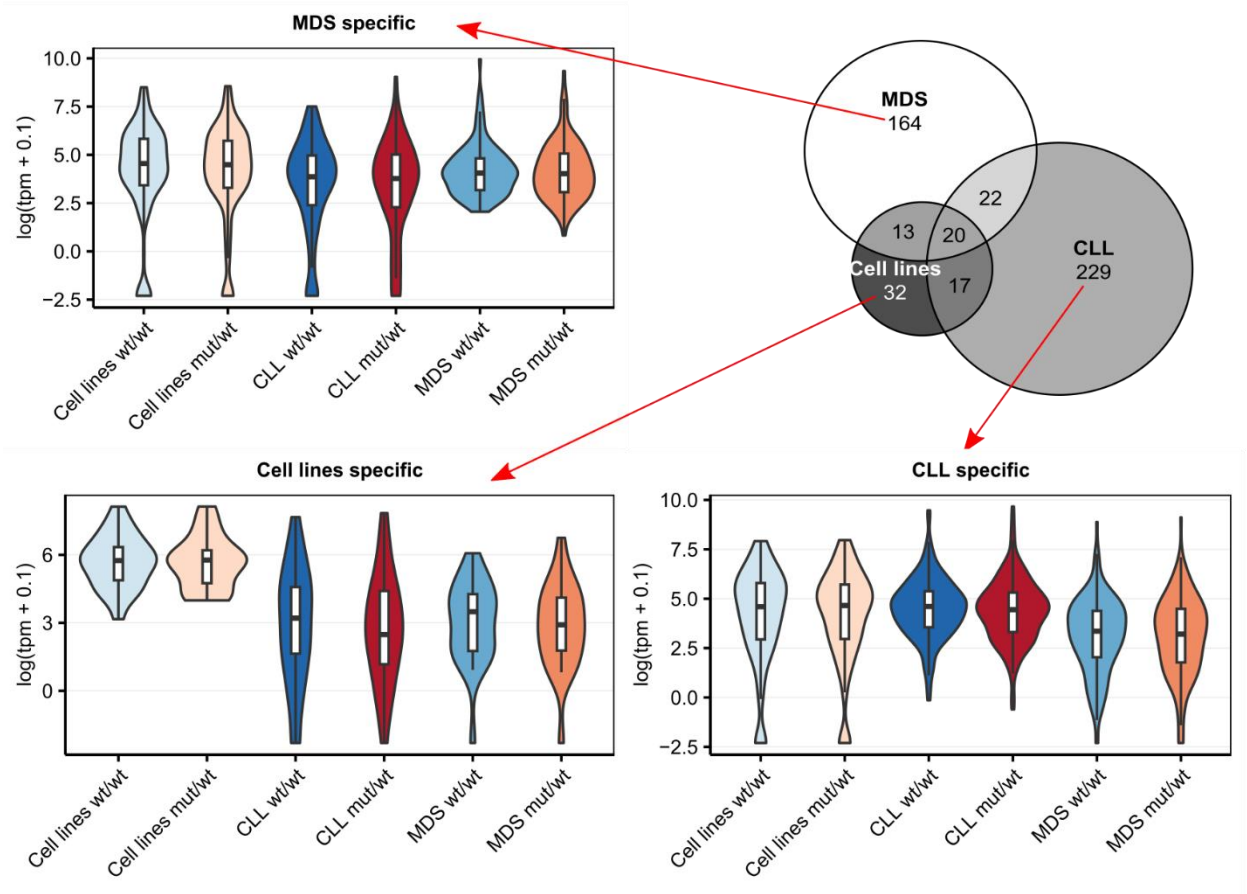
Supplemental Fig. S8. *SF3B1* substantially expressed isoforms. Blue lines separate GENCODE annotated isoforms (top) and Iso-Seq isoforms (bottom). For each Iso-Seq isoform, the fraction of reads identified for each isoform is shown on the right. Green box represents the p14 interacting region, the yellow boxes represent HEAT domains.



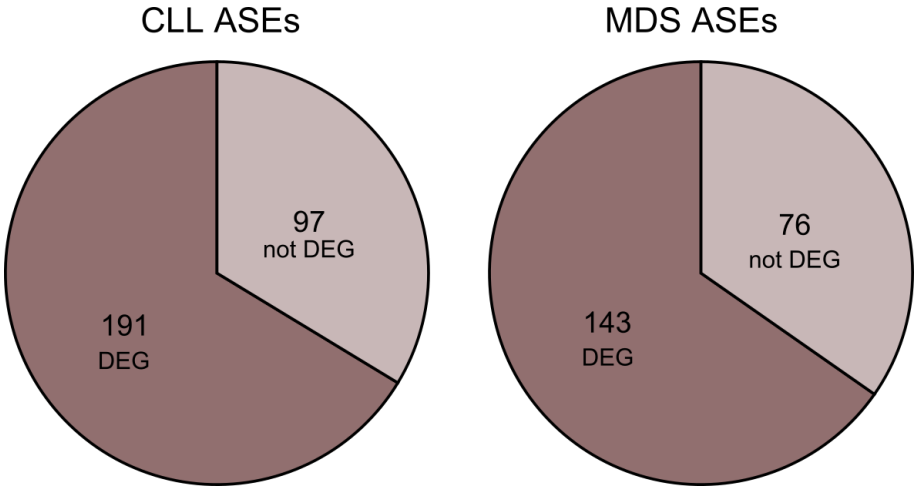
Supplemental Fig. S9. Distribution of the number of samples supporting an alternative splicing event (ASE) after applying different thresholds for minimum read number supporting the ASE.



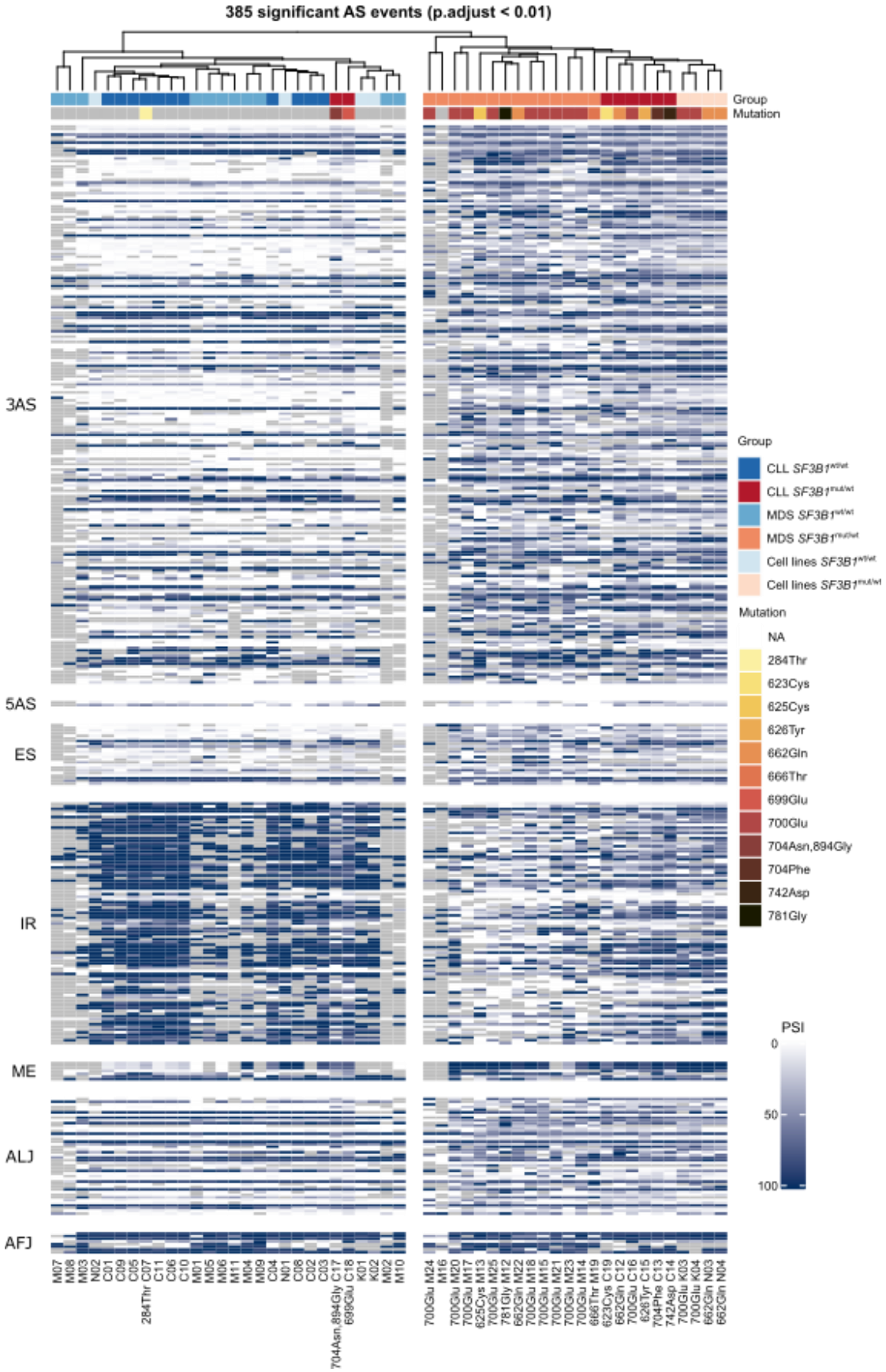
Supplemental Fig. S10. Overlap between highly significant (q-value < 0.05) alterations in alternative splicing events (ASEs) in samples with *SF3B1* mutation identified in the three datasets used (cell lines, CLL patients, or MDS patients).



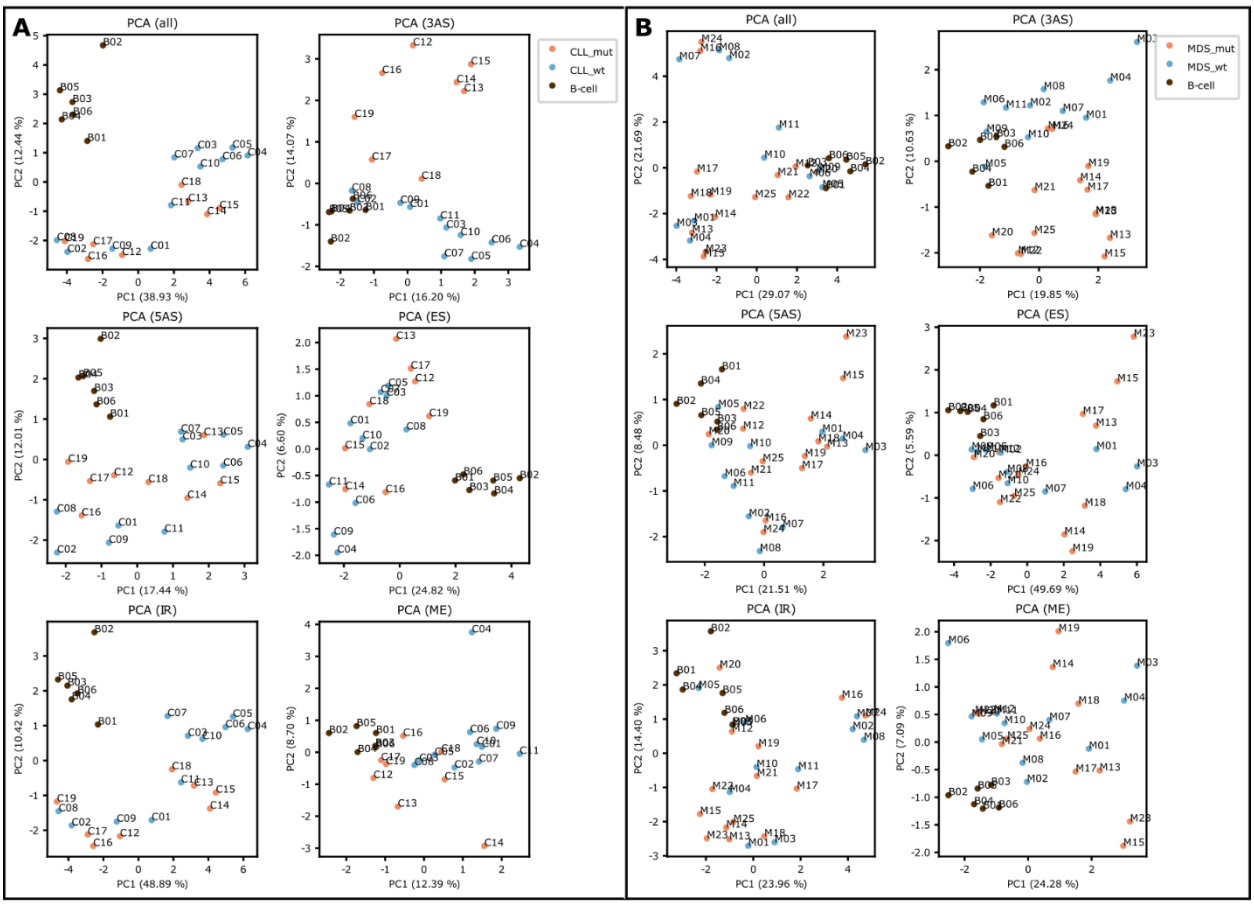
Supplemental Fig. S11. Overlap of significant alternative splicing events (ASEs) detected using the three sets (cell lines, CLL and MDS patients separately. Although most of the events seem set-specific, the dPSI has high correlation among the datasets (see Fig. 2). For each subset-specific ASE, the overall expression at gene level per group is shown to highlight the fact, the specificity seems to be related to cell-specific transcriptomic profile, rather than specific SF3B1^{mut} mode of action in each dataset.



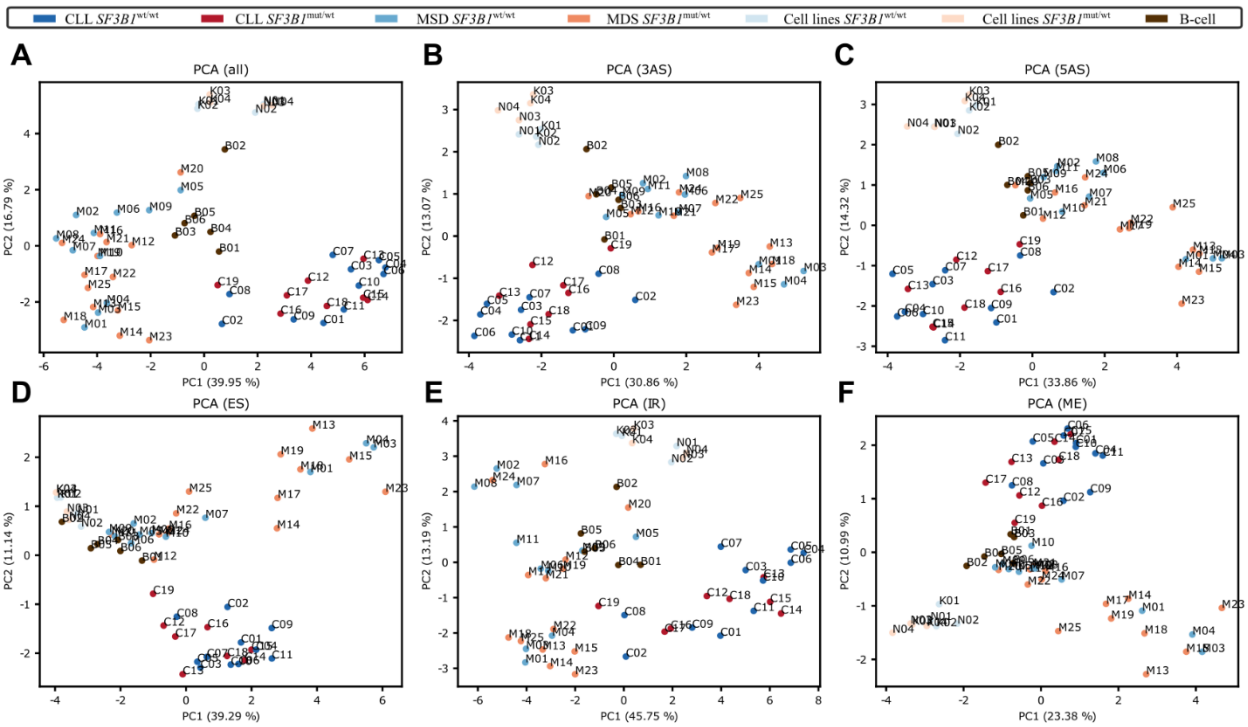
Supplemental Fig. S12. Proportion of significant alternative splicing events (ASEs) detected using CLL and MDS patients separately with differentially expressed genes identified with DESeq2 between MDS and CLL (FDR < 0.05).


385 significant AS events (p.adjust < 0.01)
Group
Mutation
3AS
5AS
ES
IR
ME
ALJ
AFJ
Group
CLL SF3B1wt/wt
CLL SF3B1mut/wt
MDS SF3B1wt/wt
MDS SF3B1mut/wt
Cell lines SF3B1wt/wt
Cell lines SF3B1mut/wt
Mutation
NA
284Thr
623Cys
625Cys
626Tyr
662Gln
666Thr
699Glu
700Glu
704Asn,894Gly
704Phe
742Asp
781Gly
PSI
0
50
100
M07 M08 M03 N02 C01 C09 C05 284Thr C07 C11 C06 C10 M01 M05 M06 M11 M04 M09 C04 N01 C08 C02 C03 704Asn,894Gly C17 699Glu C18 K01 K02 M02 M10
700Glu M24 M16 700Glu M20 700Glu M17 625Cys M13 700Glu M25 781Gly M12 662Gln M22 700Glu M18 700Glu M15 700Glu M21 700Glu M23 700Glu M14 666Thr M19 623Cys C19 662Gln C12 700Glu C16 626Tyr C15 704Phe C13 742Asp C14 700Glu K03 700Glu K04 662Gln N03 662Gln N04

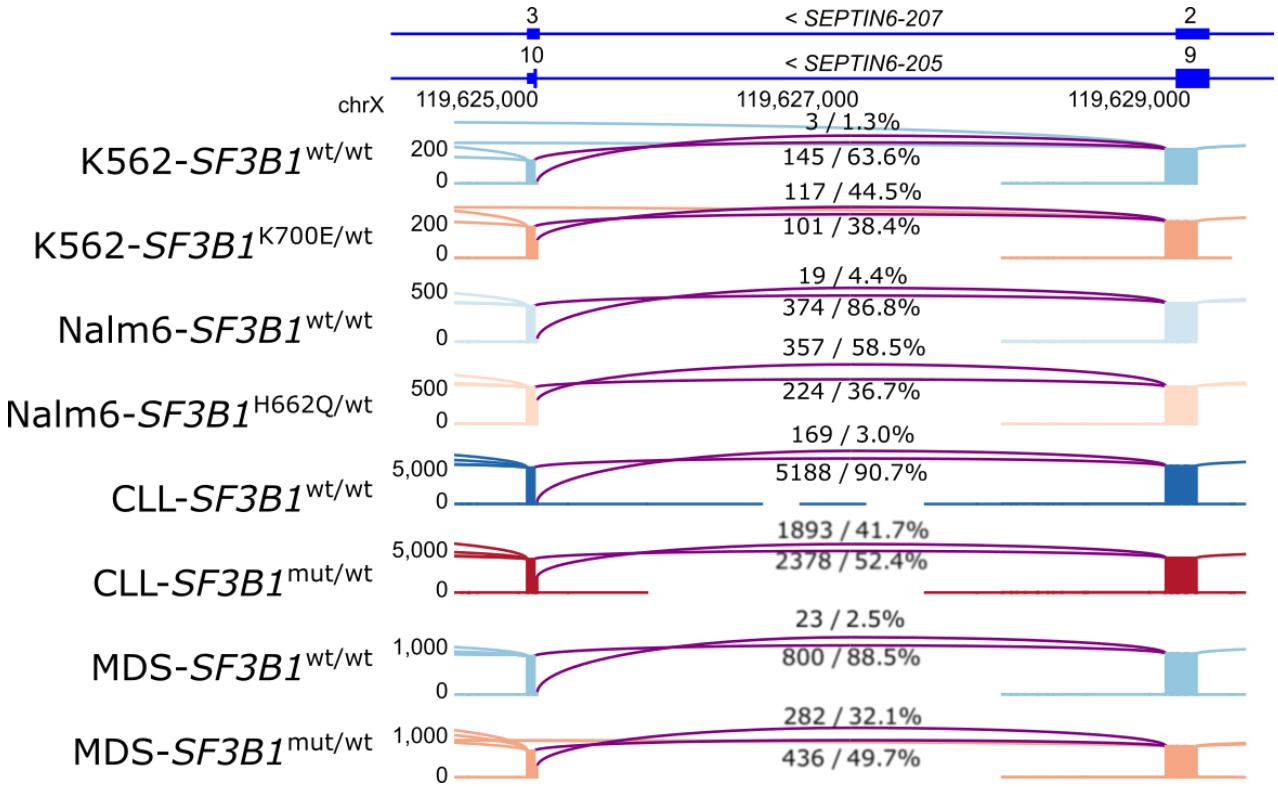
Supplemental Fig. S13. A subset of the most significant alternative splicing events (p.adjusted < 0.01) affected by *SF3B1* mutations in leukemia cell lines as well as CLL and MDS patients show increased 3' alternative splice sites usage and decreased intron retention in samples with *SF3B1* mutation.



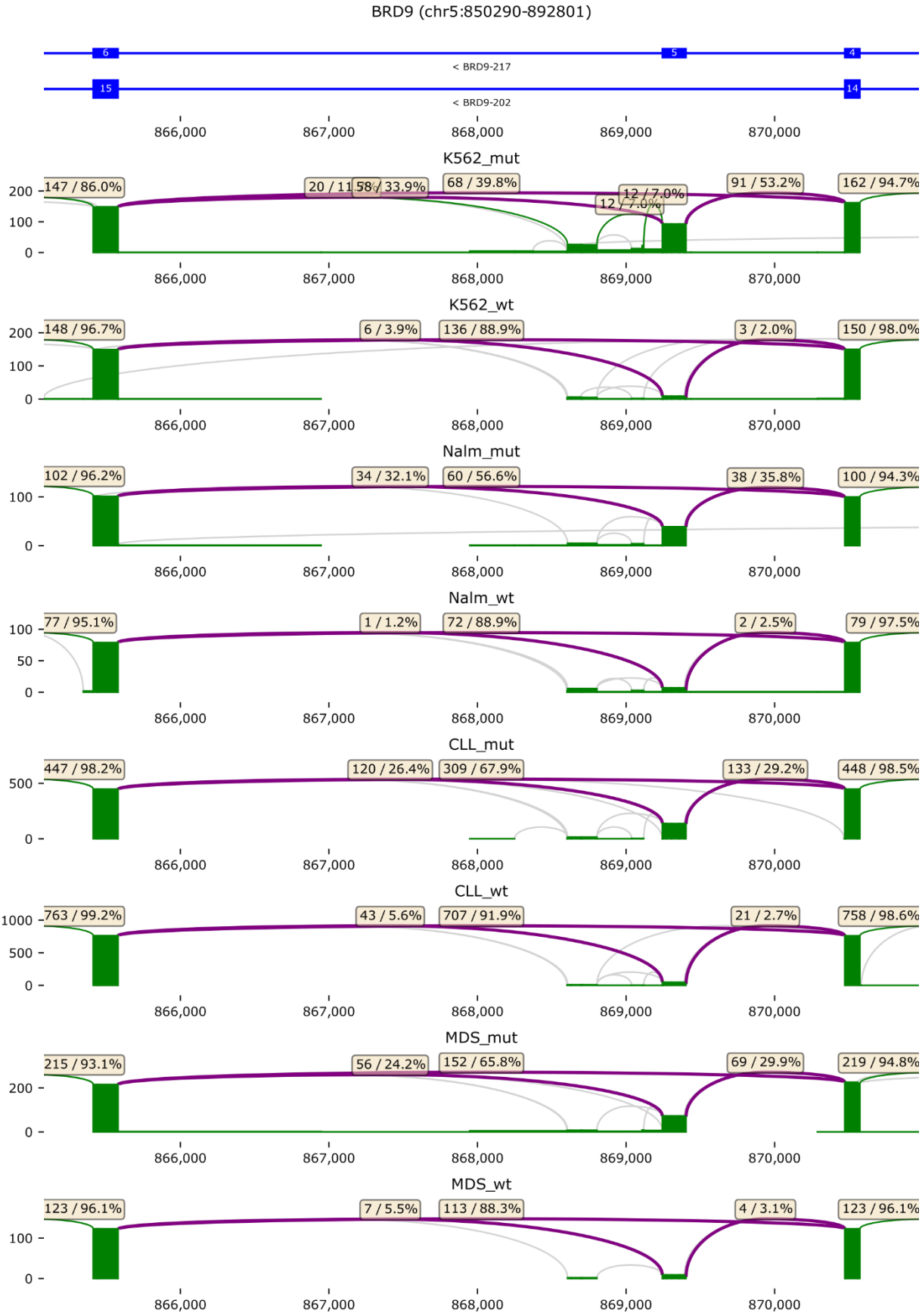
Supplemental Fig. S14. Principal component (PC) analysis based on the isoform usage of CLL (***A***), MDS (***B***), or cell lines samples (***C***) using all or each of the splice event types separately.



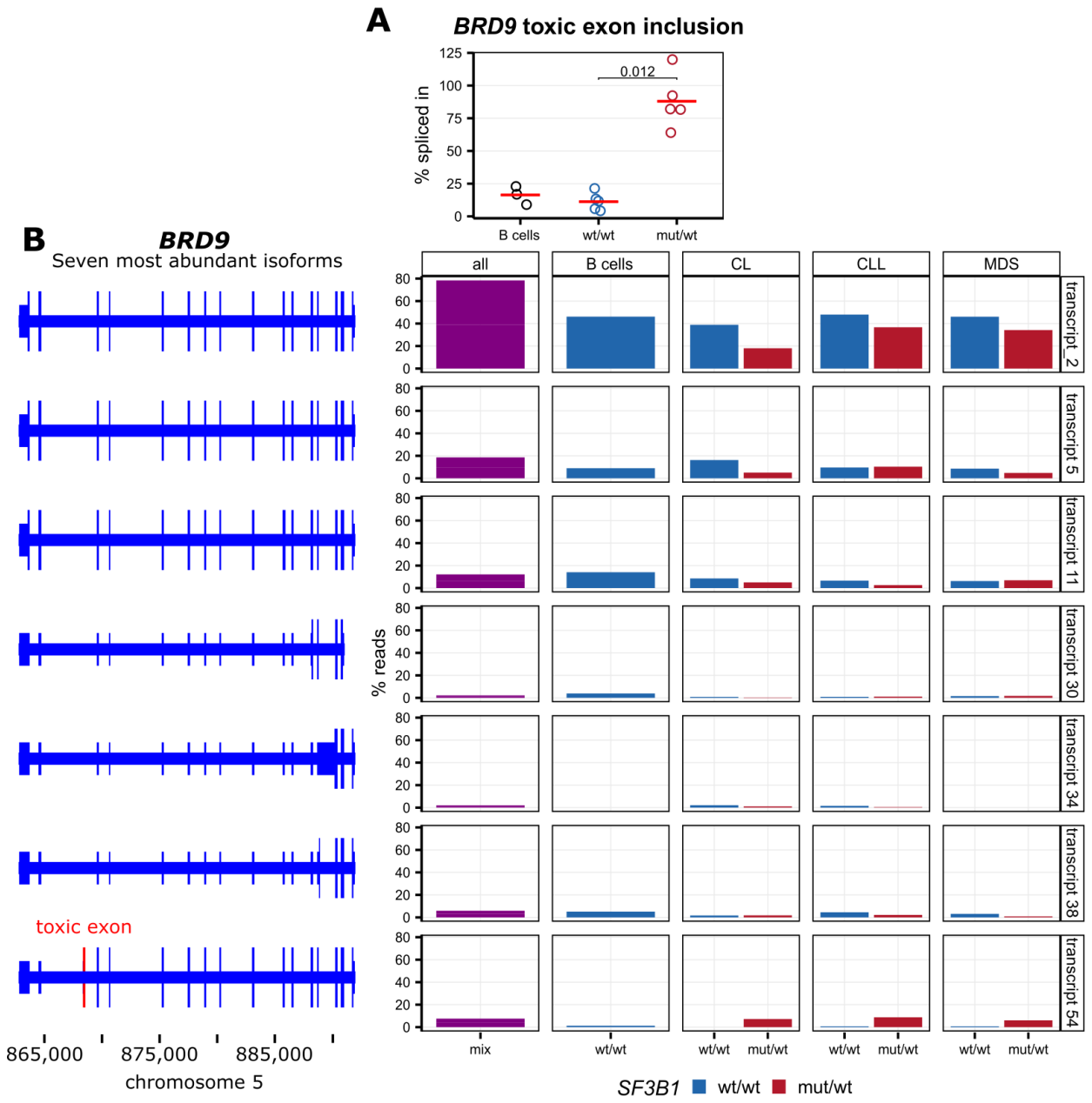
Supplemental Fig. S15. Principal component (PC) analysis based on the isoform usage of all (***A***) or single type splicing events (***B***–***F***).



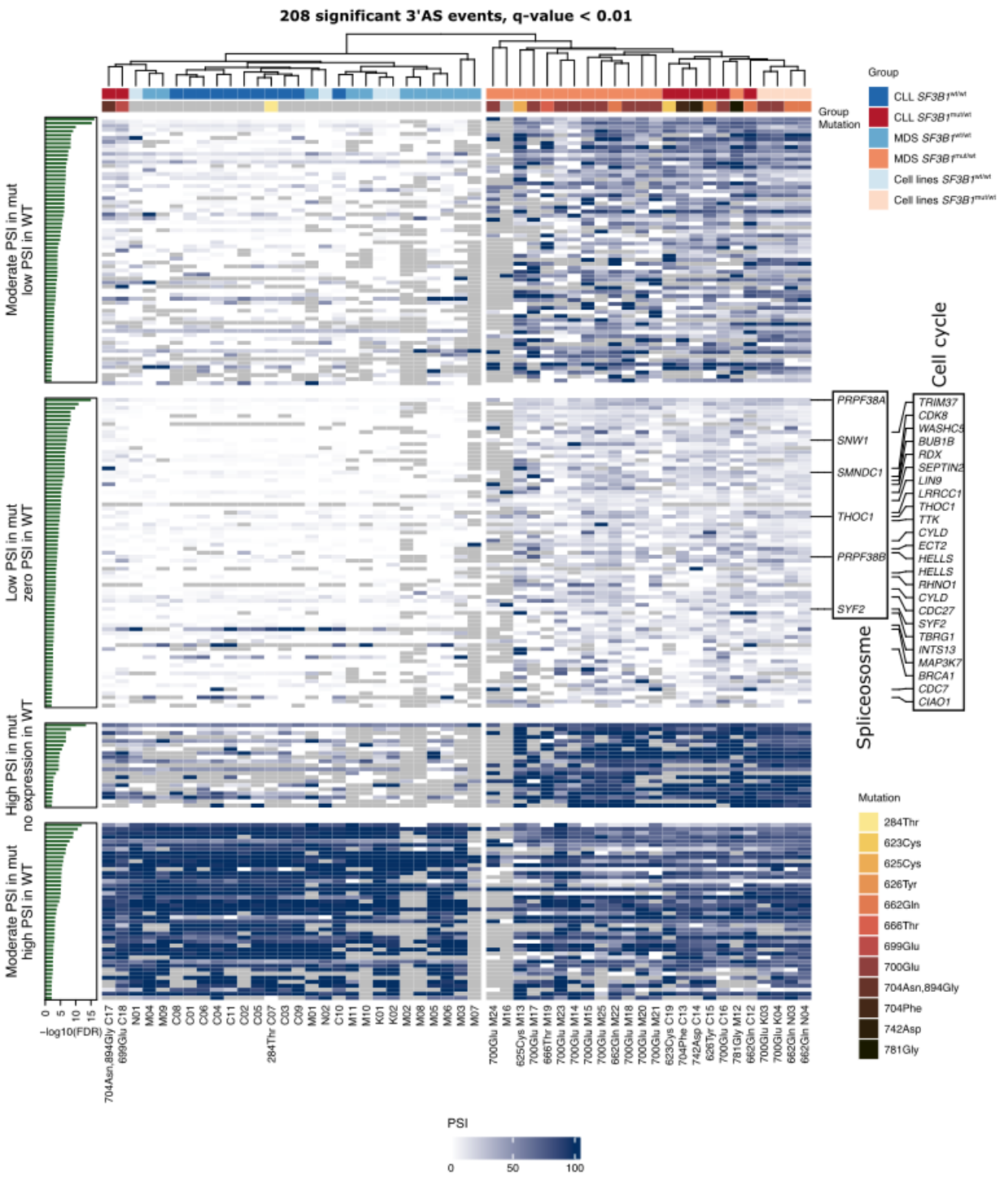
Supplemental Fig. S16. *SEPTIN6* isoforms with already reported 3'AS event in *SF3B1*$^{mut/wt}$. Each track shows sum of the reads identified in each group. The number of reads supporting each junction and percentage is shown. The width of the lines correlates with significance. Significant differences are coloured in purple.


BRD9 (chr5:850290-892801)
6
5
4
< BRD9-217
15
14
< BRD9-202
866,000
867,000
868,000
869,000
870,000
K562_mut
147 / 86.0%
20 / 11.7%
58 / 33.9%
68 / 39.8%
12 / 7.0%
12 / 7.0%
91 / 53.2%
162 / 94.7%
K562_wt
148 / 96.7%
6 / 3.9%
136 / 88.9%
3 / 2.0%
150 / 98.0%
Nalm_mut
102 / 96.2%
34 / 32.1%
60 / 56.6%
38 / 35.8%
100 / 94.3%
Nalm_wt
77 / 95.1%
1 / 1.2%
72 / 88.9%
2 / 2.5%
79 / 97.5%
CLL_mut
447 / 98.2%
120 / 26.4%
309 / 67.9%
133 / 29.2%
448 / 98.5%
CLL_wt
763 / 99.2%
43 / 5.6%
707 / 91.9%
21 / 2.7%
758 / 98.6%
MDS_mut
215 / 93.1%
56 / 24.2%
152 / 65.8%
69 / 29.9%
219 / 94.8%
MDS_wt
123 / 96.1%
7 / 5.5%
113 / 88.3%
4 / 3.1%
123 / 96.1%

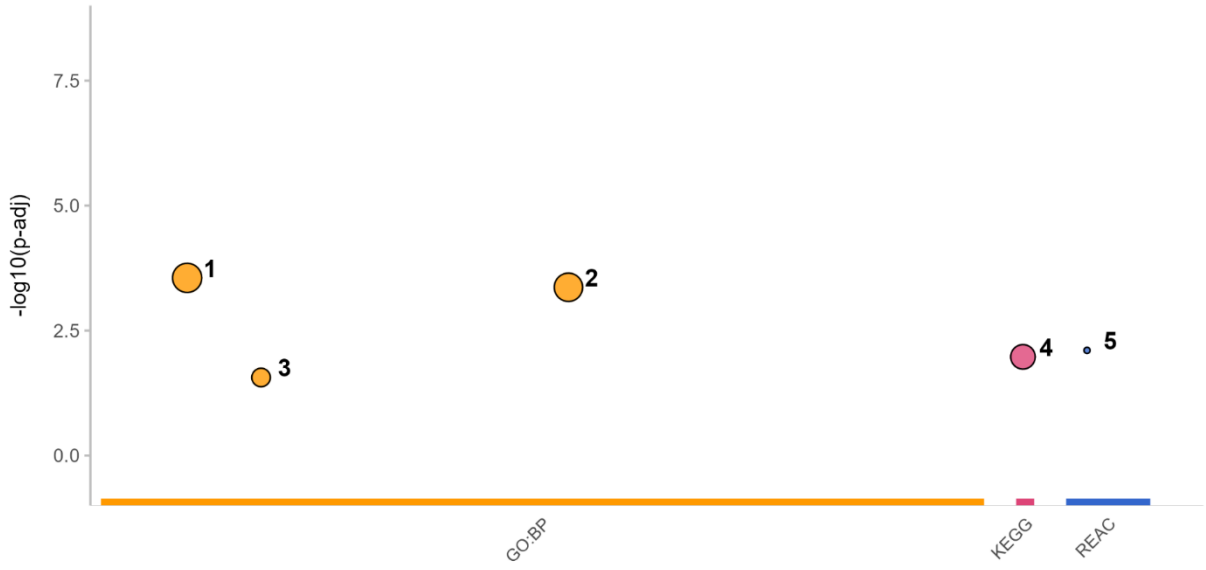
Supplemental Fig. S17. *BRD9* isoforms with already reported *BRD9* toxic exon inclusion in *SF3B1*$^{mut/wt}$. Each track shows sum of the reads identified in each group. The number of reads supporting each junction and percentage is shown. Significantly different splicing event are shown as purple arcs whereas non-significantly different splicing events supported by > 5% reads are shown as green arcs, and events supported by > 0.1% reads as grey arcs.



Supplemental Fig. S18. *BRD9* exon inclusion in *SF3B1* mutated samples. (***A***) Proportion of the exon inclusion based on qPCR using a subset of CLL and healthy B cell samples used in Iso-Seq. (***B***) Most abundant isoforms of *BRD9* identified with Iso-Seq. (***C***) Expression shown as percentage of total reads mapped to *BRD9* gene. The toxic exon is marked in red (*B*) and was identified in transcript 54 isoform (*C*), which had a higher proportional expression in samples with an *SF3B1* mutation.



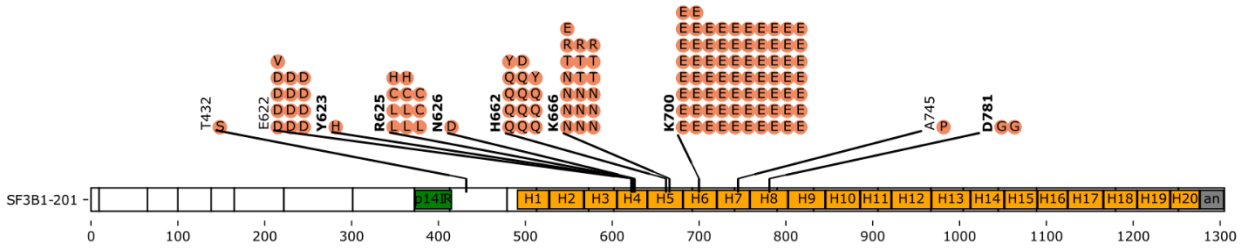
Supplemental Fig. S19. Clusters of 3' alternative splice sites strongly affected by *SF3B1* mutations (p.adjusted < 0.01) in leukaemia cell lines as well as CLL and MDS patients show enrichment in spliceosome and cell cycle pathways among events with low occurrence in *SF3B1*$^{mut/wt}$ and very low to none in *SF3B1*$^{wt/wt}$. The optimal number of clusters was found with the silhouette method.



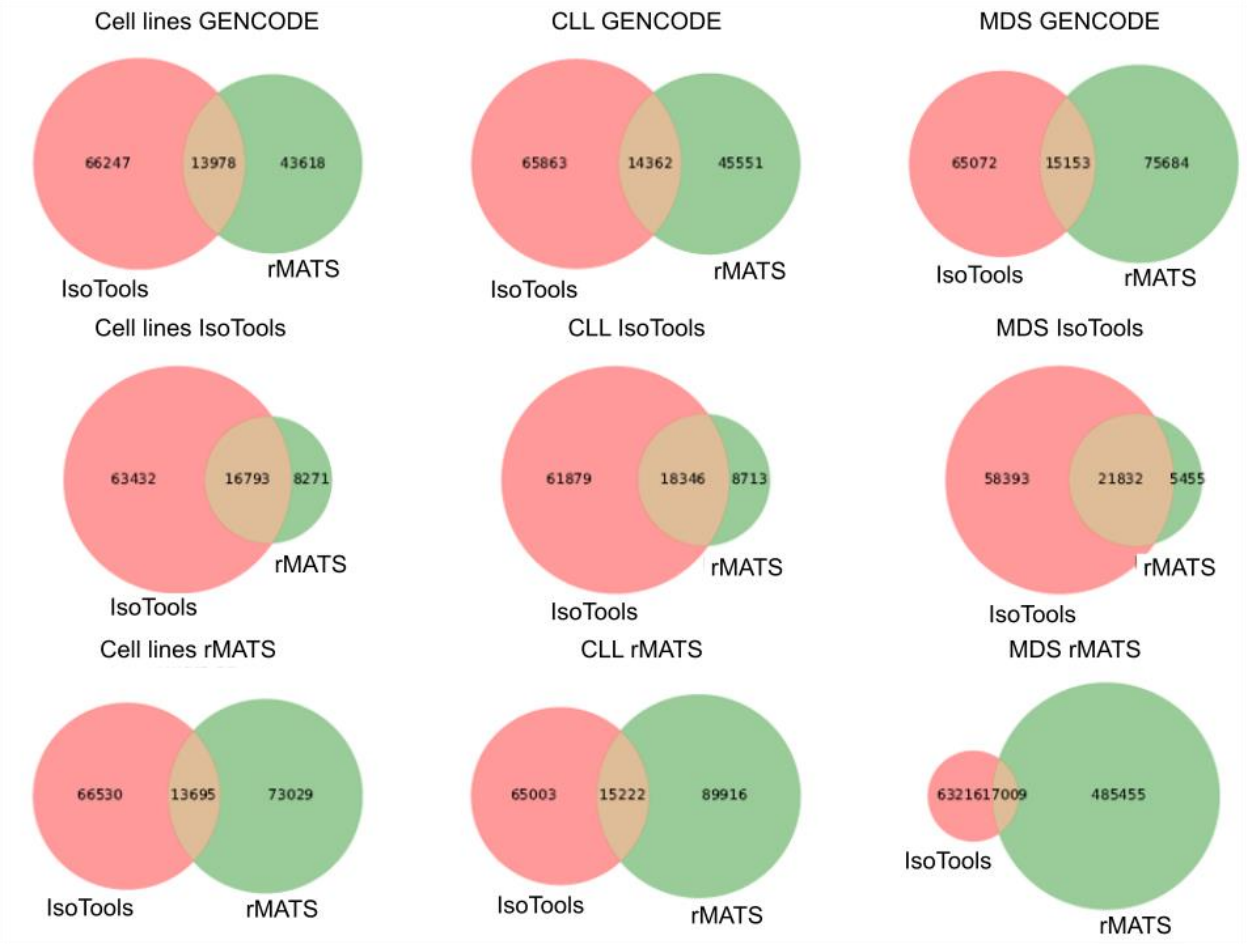
id	source	term id	term name	term size		q-value
1	GO:BP	GO:0007049	cell cycle	1811		2.8×10^{-4}
2	GO:BP	GO:0051726	regulation of cell cycle	1112		4.3×10^{-4}
3	GO:BP	GO:0015803	branched-chain amino acid transport	14		2.7×10^{-2}
4	KEGG	KEGG:03040	Spliceosome	150		1.1×10^{-2}
5	REAC	REAC:R-HSA-5660862	Defective SLC7A7 causes lysinuric protein intolerance (LPI)	2		7.9×10^{-3}

g:Profiler (biit.cs.ut.ee/gprofiler)

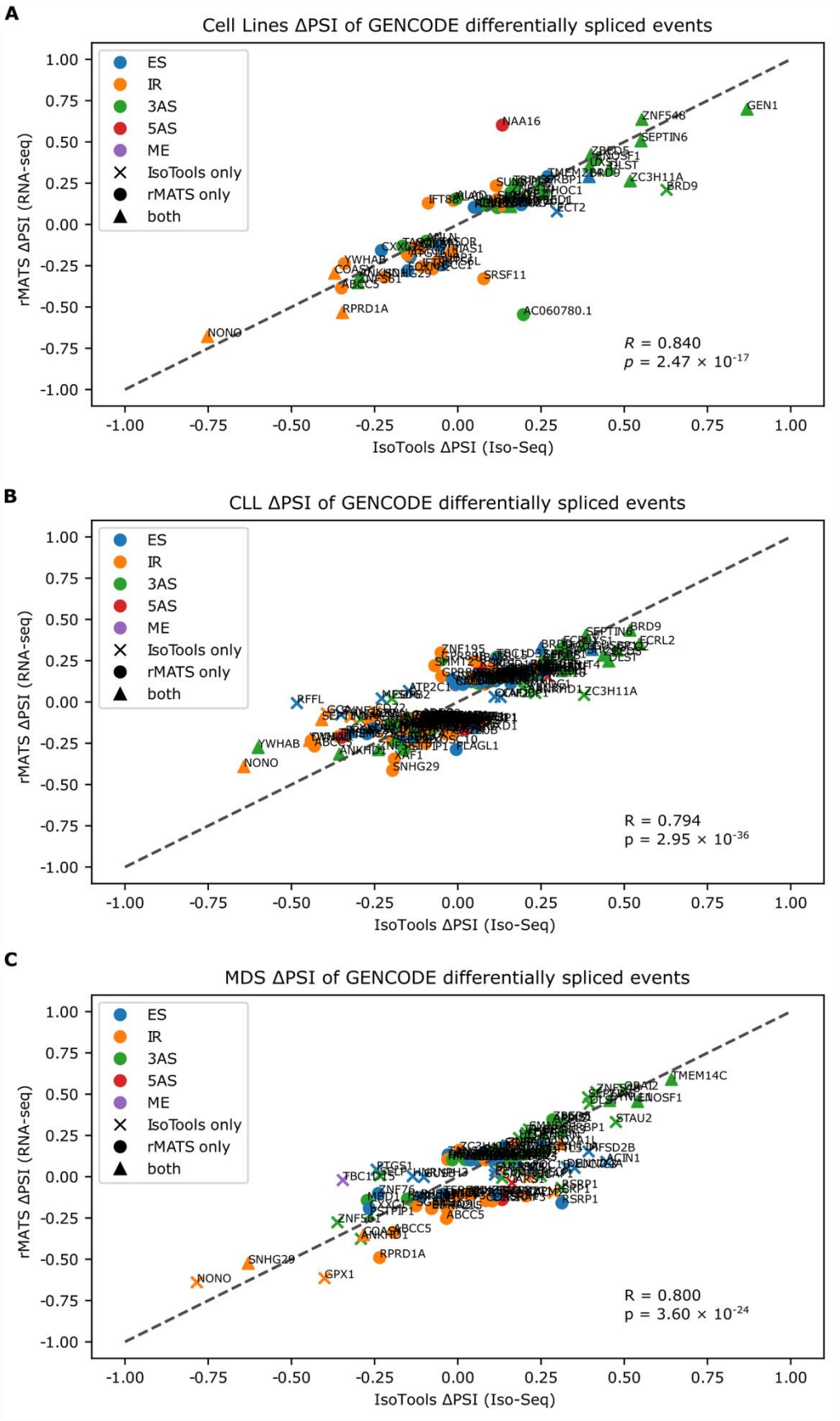
Supplemental Fig. S20. Overrepresentation analysis of the genes with 3'AS belonging to "Low in mut, zero in WT" cluster from Supplemental Fig. S19. "q-value" – adjusted p-value.



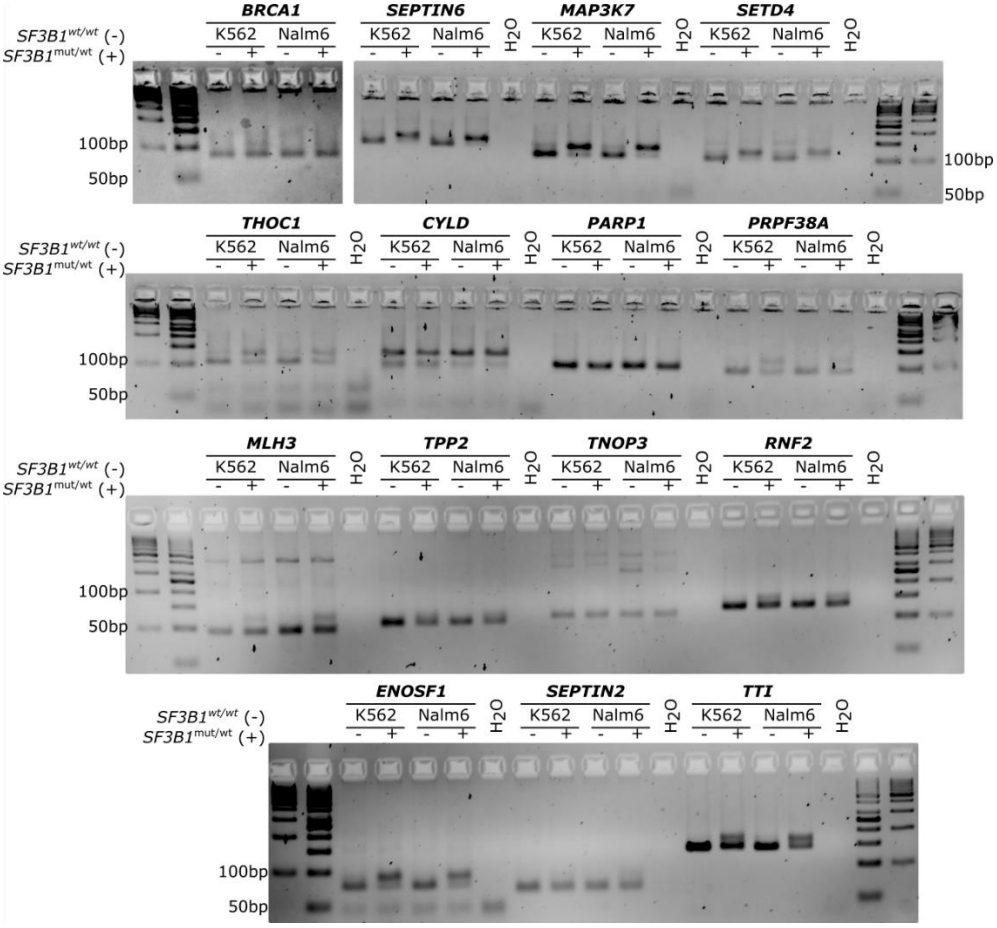
Supplemental Fig. S21. *SF3B1* mutations detected in the samples used in this and publicly available MDS patients' RNA-seq data (see Supplemental Table S1 for details).



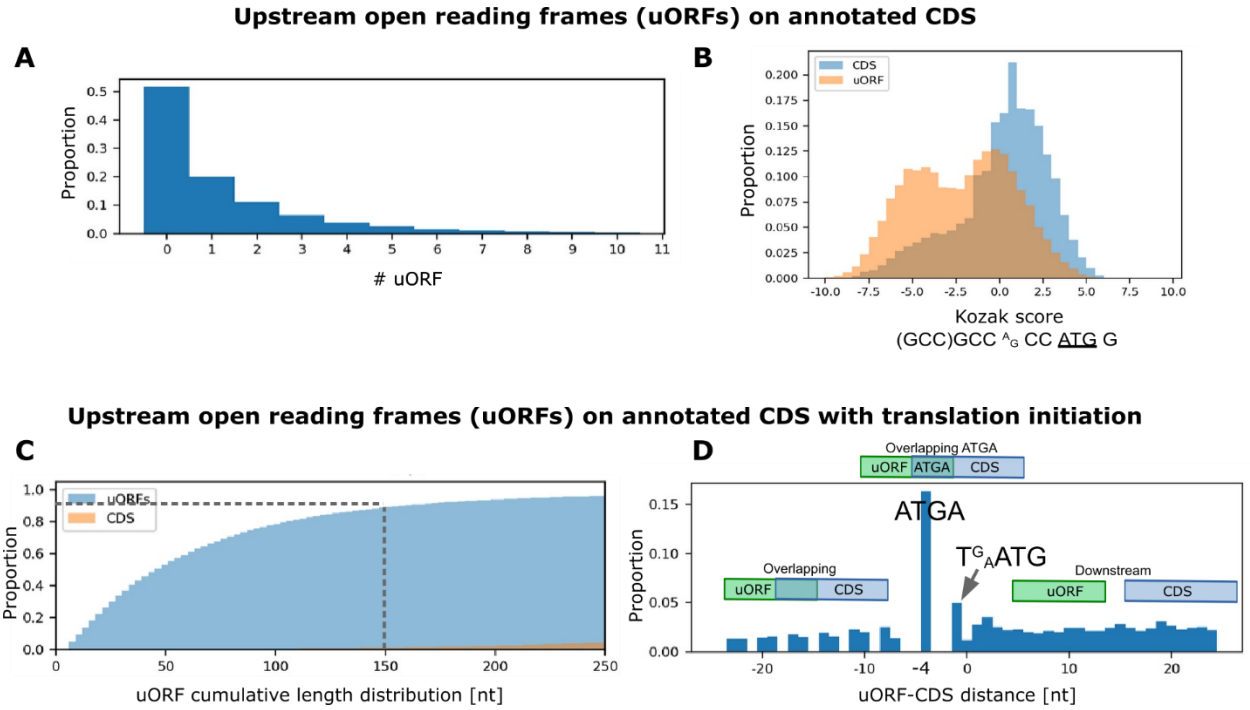
Supplemental Fig. S22. Overlap of the splicing events detected with long Iso-Seq reads analyzed with IsoTools and short RNA-seq analyzed with rMATS using gene structures from GENCODE, derived from IsoTools, and rMATS. The overlaps in each dataset (cell lines (CL), CLL, MDS) are shown.


A
Cell Lines ΔPSI of GENCODE differentially spliced events
ES
IR
3AS
5AS
ME
IsoTools only
rMATS only
both
rMATS ΔPSI (RNA-seq)
IsoTools ΔPSI (Iso-Seq)
R = 0.840
p = 2.47 × 10-17
NAA16
ZNF548
GEN1
SEPTIN6
ZC3H11A
BRD9
SRSF11
RPRD1A
AC060780.1
NONO
B
CLL ΔPSI of GENCODE differentially spliced events
R = 0.794
p = 2.95 × 10-36
C
MDS ΔPSI of GENCODE differentially spliced events
R = 0.800
p = 3.60 × 10-24
TMEM14C
STAU2
GPX1
SNHG29

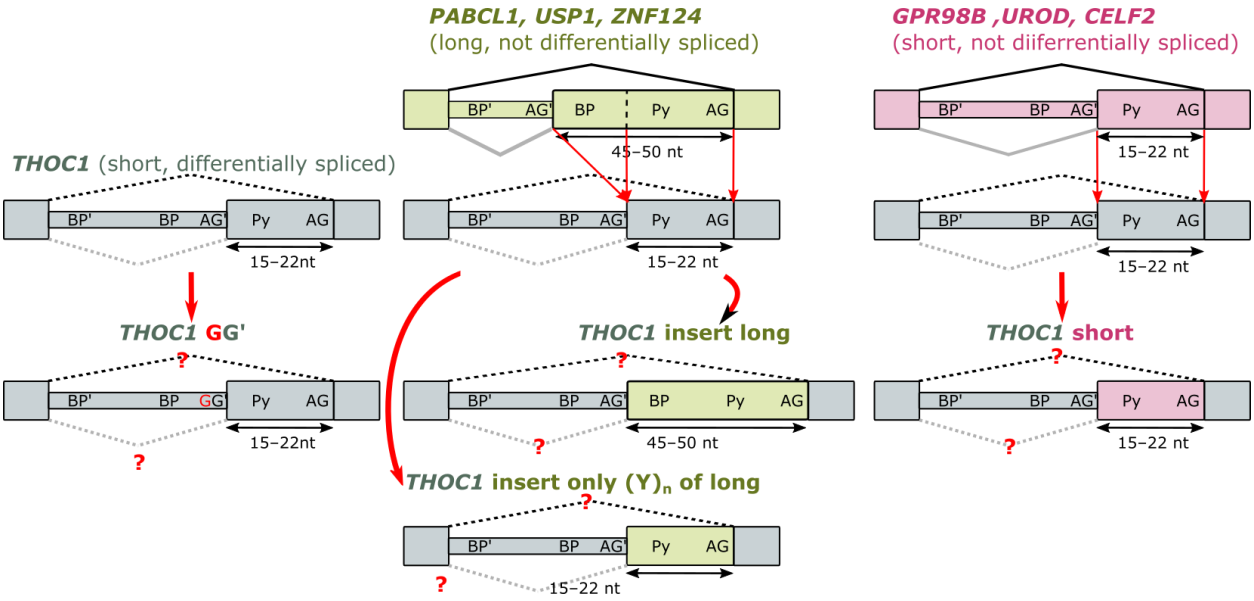
Supplemental Fig. S23. Comparison of the *SF3B1* mutations effect on splicing using Iso-Seq coupled with IsoTools and RNA-seq coupled with rMATS analyses in (***A***) K562 and Nalm6 cell lines, (***B***) CLL, and (***C***) MDS patients. The colour codes for splicing event type, whereas the point shape indicates significance detected by each of the method used. Only events affecting known isoforms annotated in GENCODE v36 were taken into consideration. (*B*) CLL samples: for the RNA-seq, eight additional patients' samples were sequenced whereas in (*C*) MDS Iso-Seq samples are from this study and RNA-seq samples are derived from publicly available datasets (see Supplemental Table S1 and Supplemental Fig. S2 for details).



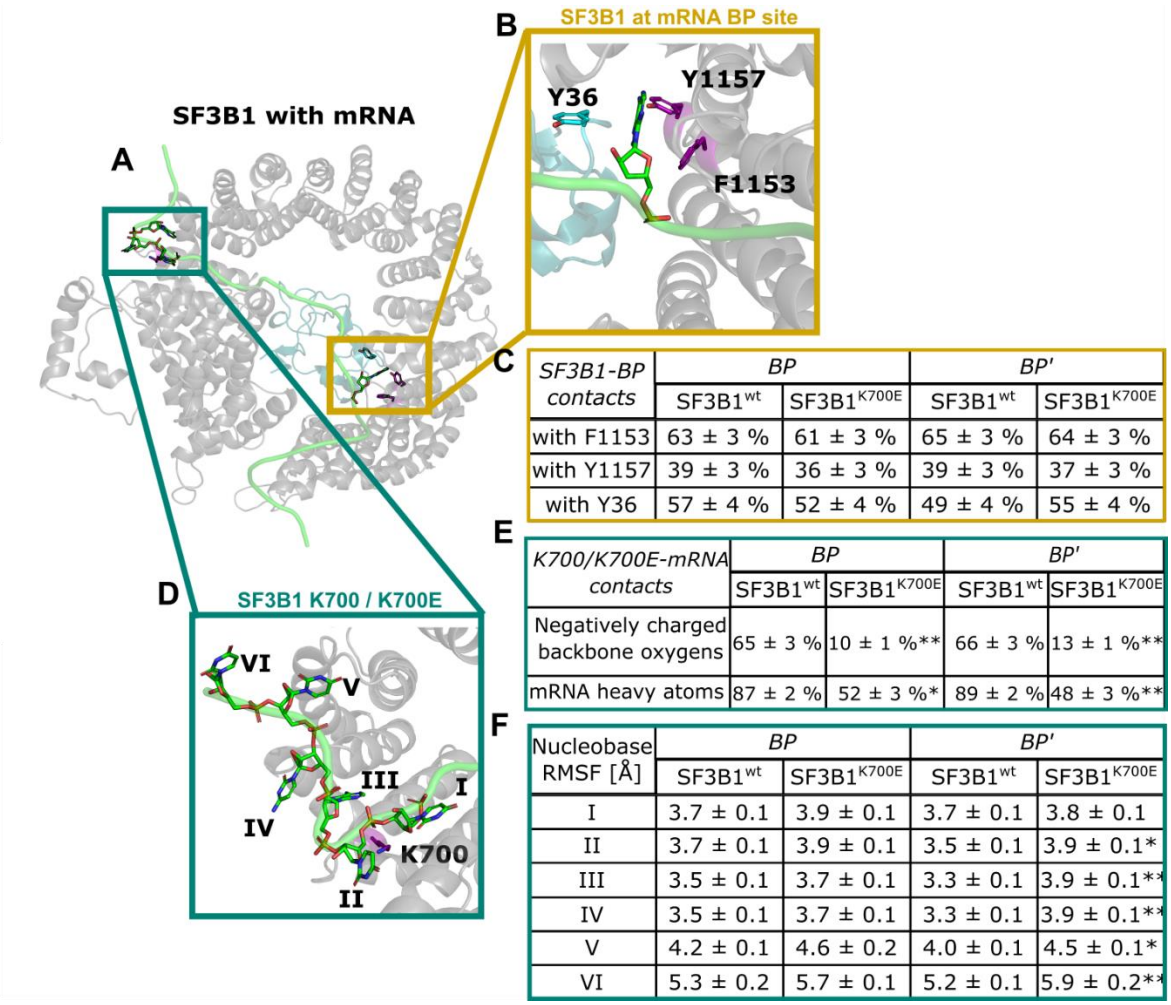
Supplemental Fig. S24. Validation of alternative splicing detected with Iso-Seq. Agarose gels showing alternative splicing resulting in longer fragments in the *SF3B1*$^{mut/wt}$ cell lines.



Supplemental Fig. S25. Investigation of upstream open reading frames (ORFs) on annotated coding DNA sequences (CDS). (***A***) The distribution of the number of uORFs per CDS showing > 50% protein coding transcript have upstream start codons. (***B***) – The distribution of Kozak sequence(Kozak 1987) scores on annotated CDS. (***C***) uORF length on annotated CDS showing 99% of CDSs > 150 bases, whereas around 99% of uORFs were < 150 bases (Embree et al. 2022). (***D***) The distribution of the distance between uORF and CDS. Negative values denote overlapping sequences. The most common distance observed was four bases overlap and these bases were ATGA. The overlapping uORF-CDS were observed up to 50 bases. The downstream instances were most common and dropped down slowly at a distance around 300 bases.



Supplemental Fig. S26. Design of the minigene assay shown in Fig. 5B. Briefly, the intron fragment between AG' and AG from differentially spliced *THOC1* was replaced with a long intron AG'-AG fragment from a not differentially spliced gene (*PABCL1*, *USP1*, or *ZNF124* shown in green), or a short intron AG'-AG fragment from a not differentially spliced gene (*GRP98B*, *UROD*, or *CELF2* shown in pink). Additionally, the long fragments were cut to match shorter fragment length but with polypyrimidine tract (Py) preserved.



C

SF3B1-BP contacts	*BP*		*BP'*	
	SF3B1^{wt}	SF3B1^{K700E}	SF3B1^{wt}	SF3B1^{K700E}
with F1153	63 ± 3 %	61 ± 3 %	65 ± 3 %	64 ± 3 %
with Y1157	39 ± 3 %	36 ± 3 %	39 ± 3 %	37 ± 3 %
with Y36	57 ± 4 %	52 ± 4 %	49 ± 4 %	55 ± 4 %

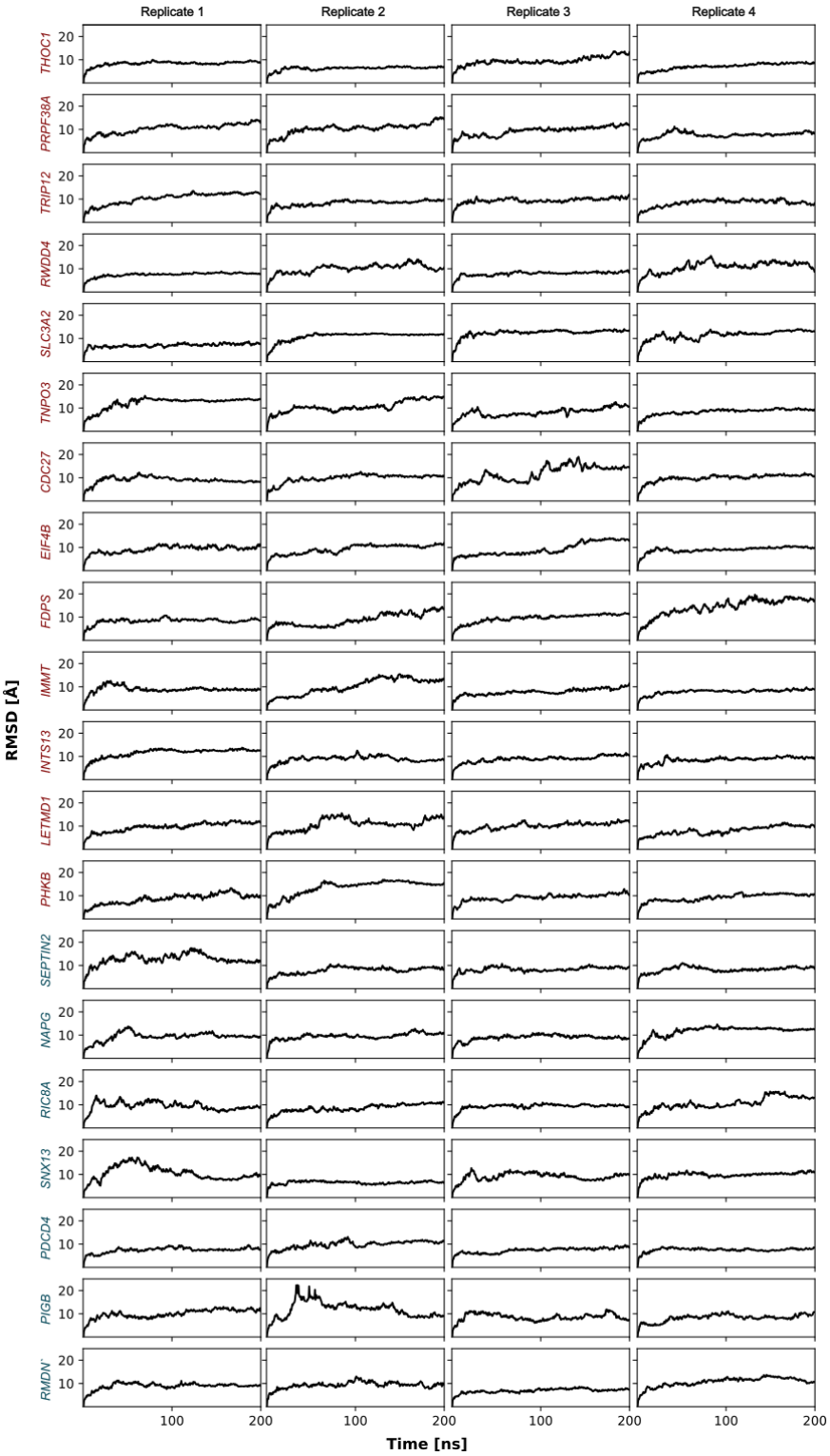
E

K700/K700E-mRNA contacts	*BP*		*BP'*	
	SF3B1^{wt}	SF3B1^{K700E}	SF3B1^{wt}	SF3B1^{K700E}
Negatively charged backbone oxygens	65 ± 3 %	10 ± 1 %**	66 ± 3 %	13 ± 1 %**
mRNA heavy atoms	87 ± 2 %	52 ± 3 %*	89 ± 2 %	48 ± 3 %**

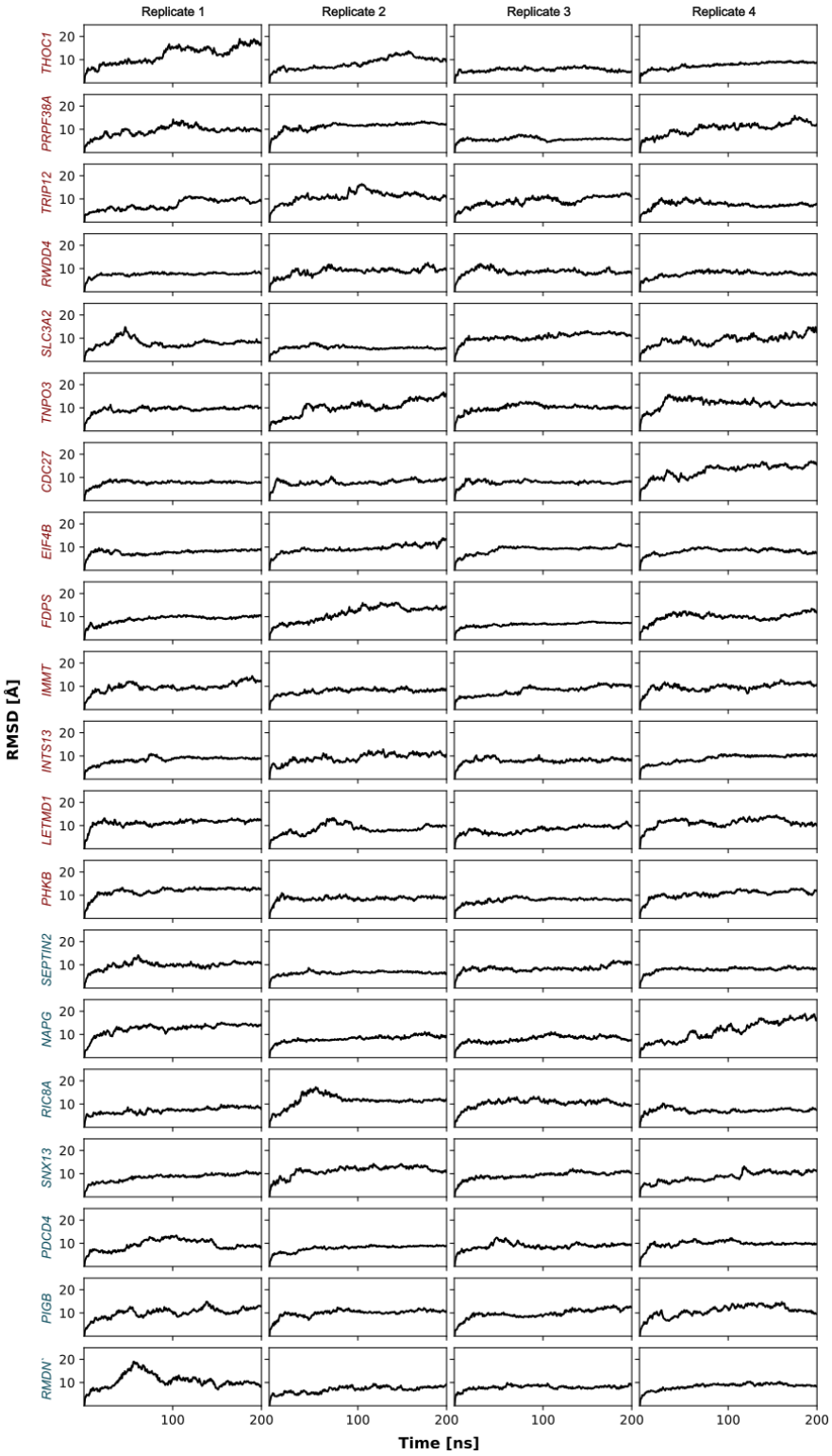
F

Nucleobase RMSF [Å]	*BP*		*BP'*	
	SF3B1^{wt}	SF3B1^{K700E}	SF3B1^{wt}	SF3B1^{K700E}
I	3.7 ± 0.1	3.9 ± 0.1	3.7 ± 0.1	3.8 ± 0.1
II	3.7 ± 0.1	3.9 ± 0.1	3.5 ± 0.1	3.9 ± 0.1*
III	3.5 ± 0.1	3.7 ± 0.1	3.3 ± 0.1	3.9 ± 0.1**
IV	3.5 ± 0.1	3.7 ± 0.1	3.3 ± 0.1	3.9 ± 0.1**
V	4.2 ± 0.1	4.6 ± 0.2	4.0 ± 0.1	4.5 ± 0.1*
VI	5.3 ± 0.2	5.7 ± 0.1	5.2 ± 0.1	5.9 ± 0.2**

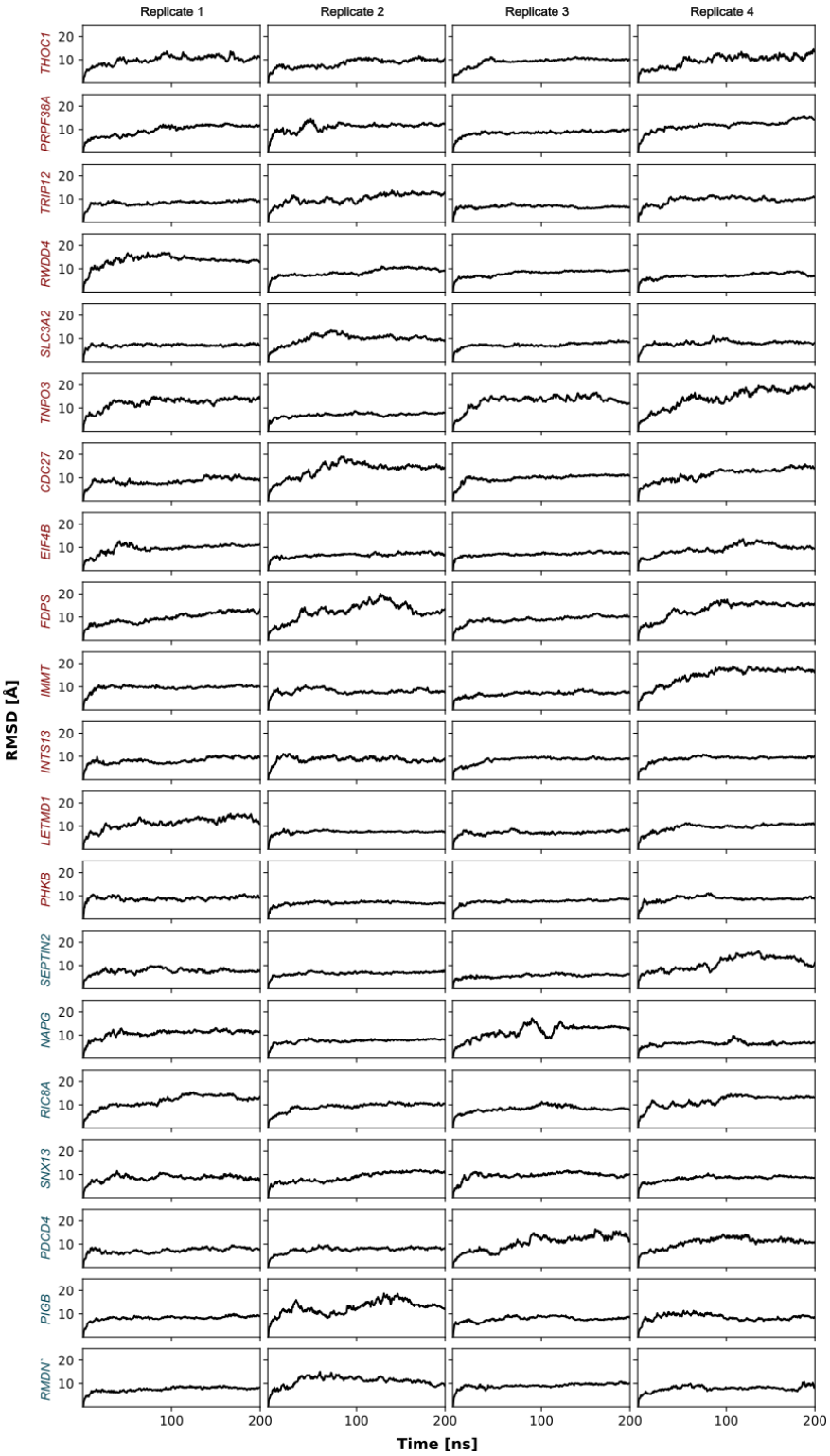
Supplemental Fig. S27. K700E mutation results in a destabilization of mRNA binding at the second pre-mRNA binding pocket around position K700. (***A***) simulated structure of SF3B1-mRNA: mRNA (green) bound to SF3B1 (grey) and the PHD finger-like domain-containing protein 5A (PHF5A, light blue). (***B***) Close-up of the BP recognition site. The branch point nucleobase (green) and aromatic amino acids surrounding the nucleobase (purple for SF3B1 and cyan for PHF5A) are shown. (***C***) Contact frequency of heavy atoms of the BP/BP' nucleobase with heavy atoms of aromatic amino acids Y36, Y1157, and F1153 in the surroundings. (***D***) Close-up of the K700 binding site. K700 (purple) and surrounding nucleobases of the pre-mRNA (green) are shown exemplarily for the *NAPG* pre-mRNA. Nucleobases are numbered with Roman numbers such that K700 is positioned between nucleobase I and II in the starting structure of MD simulations. (***E***) Interaction frequency of the functional group in the side chain of K700/E700 to the negatively charged oxygen atoms in the pre-mRNA backbone and contact frequency of heavy atoms of K700/E700 with heavy atoms of the pre-mRNA. (***F***) RMSF of nucleobases as shown in panel (*D*). Significant changes between SF3B1 and SF3B1^{K700E} are denoted by "*" for p-values < 0.05 and "**" for p-values < 0.01 (unpaired, two-tailed Student's *t*-test).



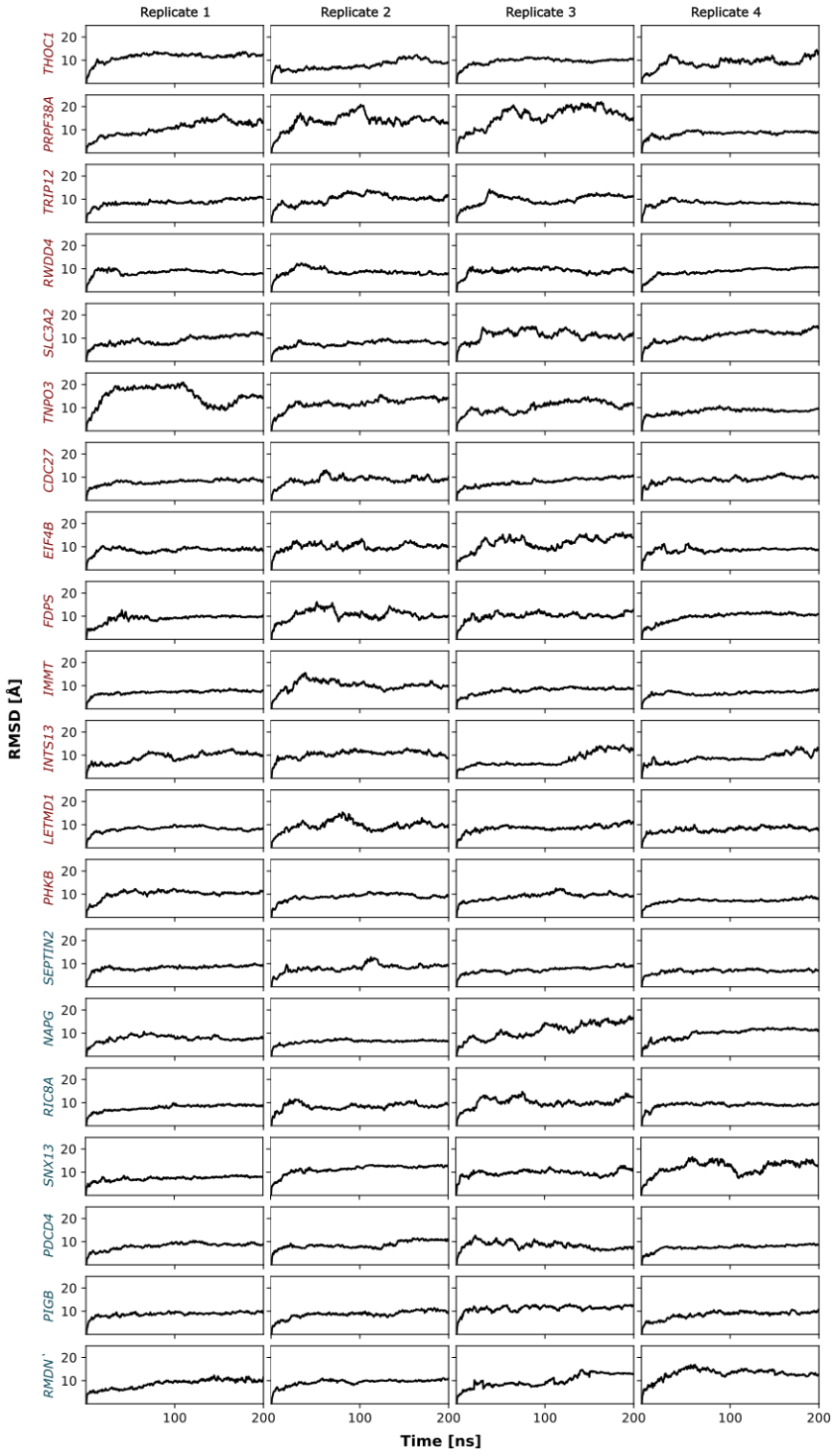
Supplemental Fig. S28. Root mean square deviation (RMSD) of the SF3B1 backbone (CA, C, N) for molecular dynamics simulations of the downstream BP bound to SF3B1^{wt}. mRNAs with 3' alternative splice site differentially spliced between SF3B1$^{mut/wt}$ and SF3B1$^{wt/wt}$ mRNAs are shown in red and non-differentially spliced are shown in blue.



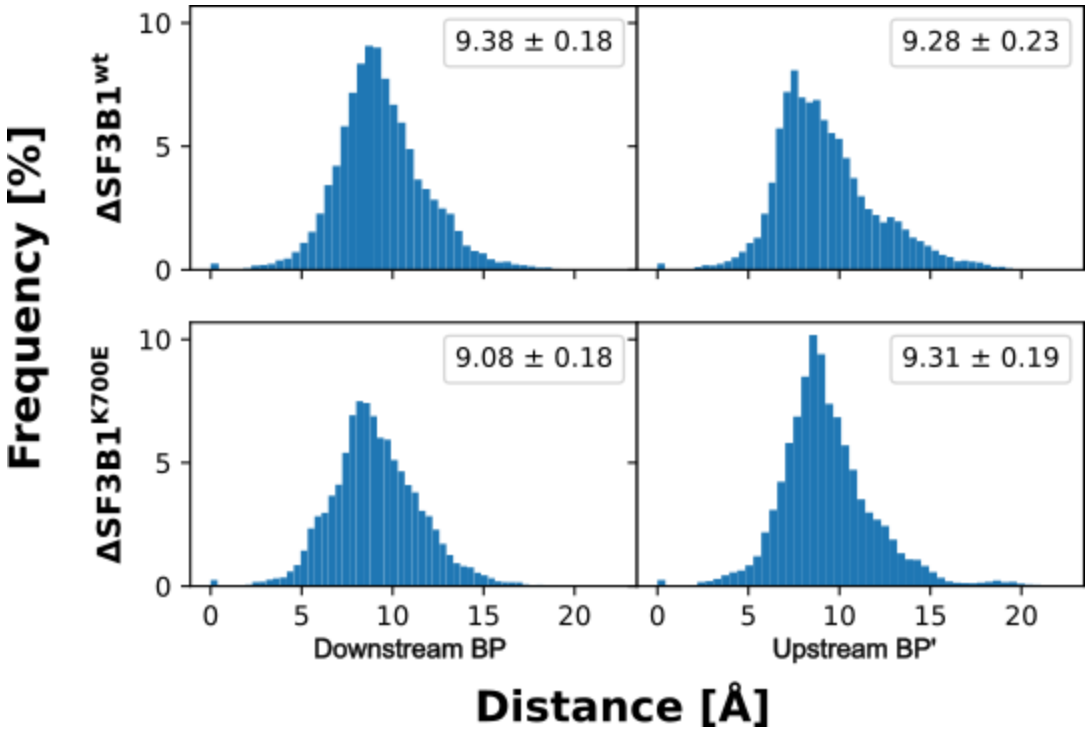
Supplemental Fig. S29. Root mean square deviation (RMSD) of the SF3B1 backbone (CA, C, N) for molecular dynamics simulations of the downstream BP bound to SF3B1^{K700E}. mRNAs with 3' alternative splice site differentially spliced between SF3B1$^{mut/wt}$ and SF3B1$^{wt/wt}$ mRNAs are shown in red and non-differentially spliced are shown in blue.



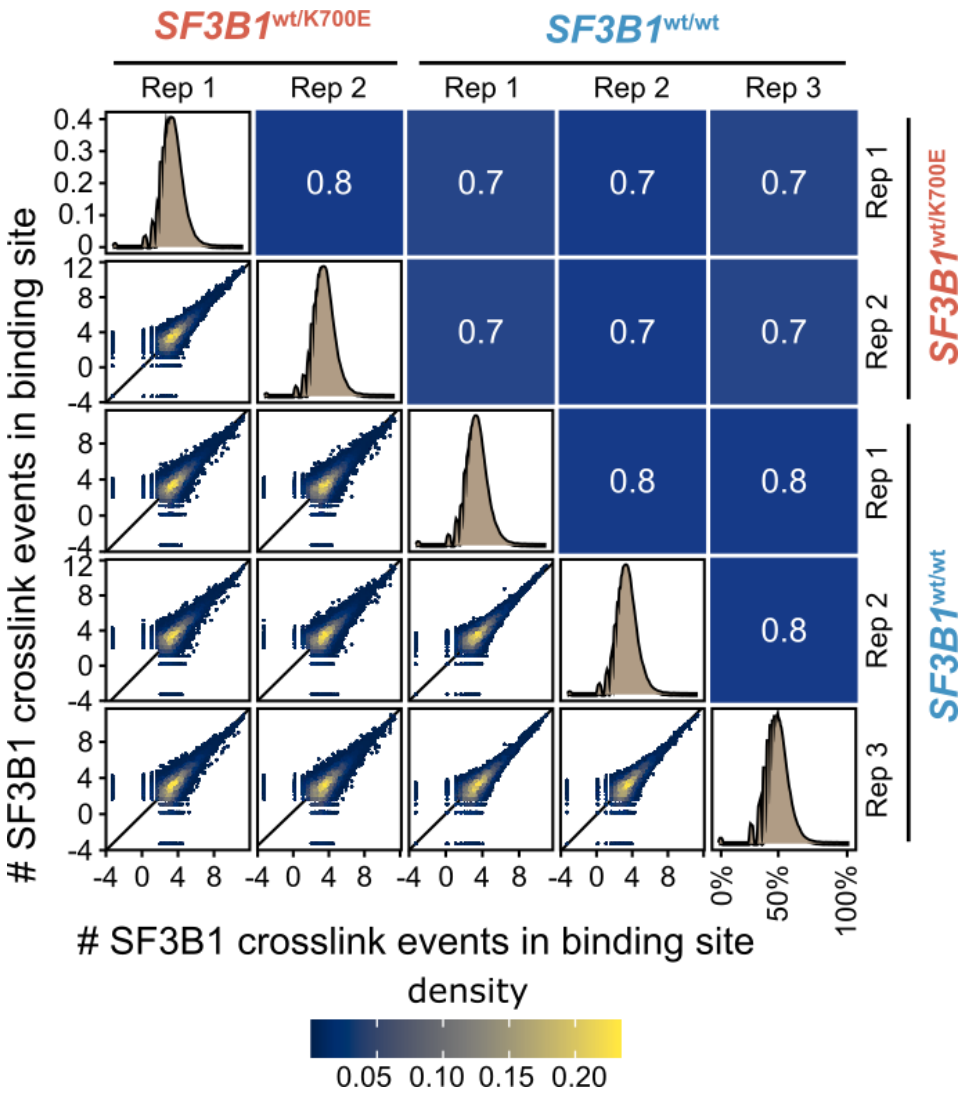
Supplemental Fig. S30. Root mean square deviation (RMSD) of the SF3B1 backbone (CA, C, N) for molecular dynamics simulations of the alternative upstream BP (BP') bound to SF3B1^{wt}. mRNAs with 3' alternative splice site differentially spliced between SF3B1$^{mut/wt}$ and SF3B1$^{wt/wt}$ mRNAs are shown in red and non-differentially spliced are shown in blue.



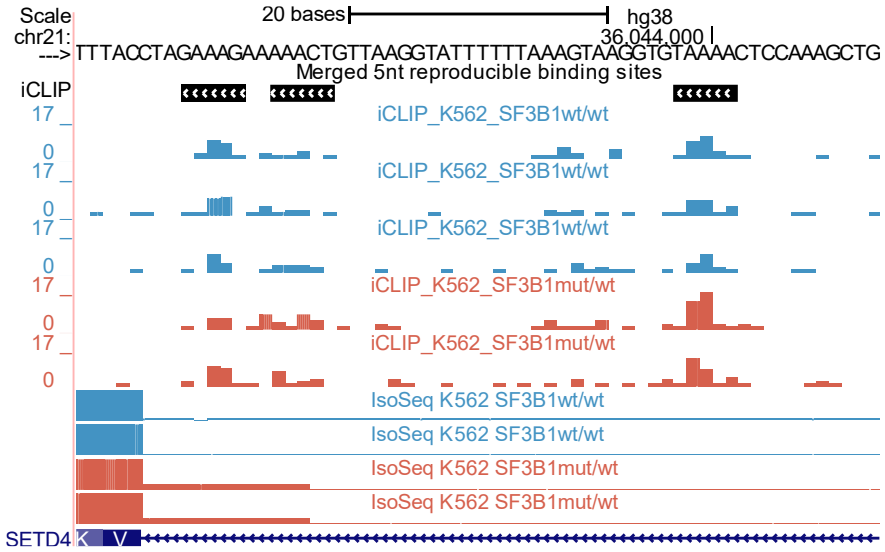
Supplemental Fig. S31. Root mean square deviation (RMSD) of the SF3B1 backbone (CA, C, N) for molecular dynamics simulations of the alternative upstream BP (BP') bound to SF3B1^{K700E}. mRNAs with 3' alternative splice site differentially spliced between SF3B1$^{mut/wt}$ and SF3B1$^{wt/wt}$ mRNAs are shown in red and non-differentially spliced are shown in blue.



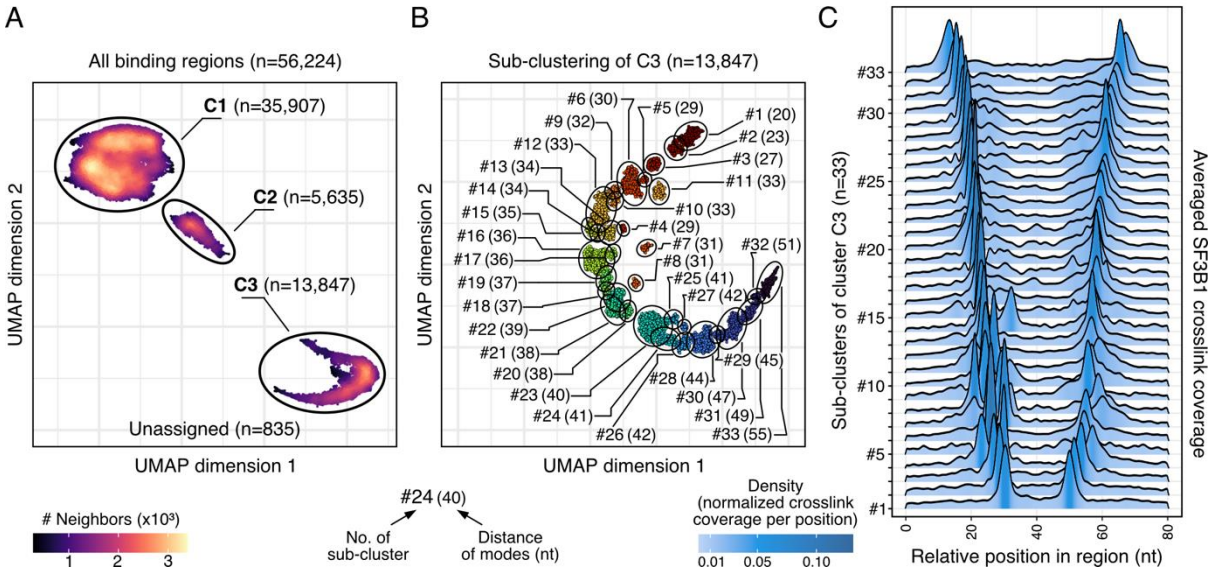
Supplemental Fig. S32. Distribution of root mean square deviation (RMSD) values of the SF3B1 backbone (CA, C, N) for molecular dynamics simulations of the downstream branch point (BP) bound to $SF3B1^{wt}$ (top left), downstream BP bound to $SF3B1^{K700E}$ (bottom left), alternative upstream BP (BP') bound to $SF3B1^{wt}$ (top right), and alternative upstream BP (BP') to $SF3B1^{K700E}$ (bottom right). Values in the boxes indicate the mean ± SEM.



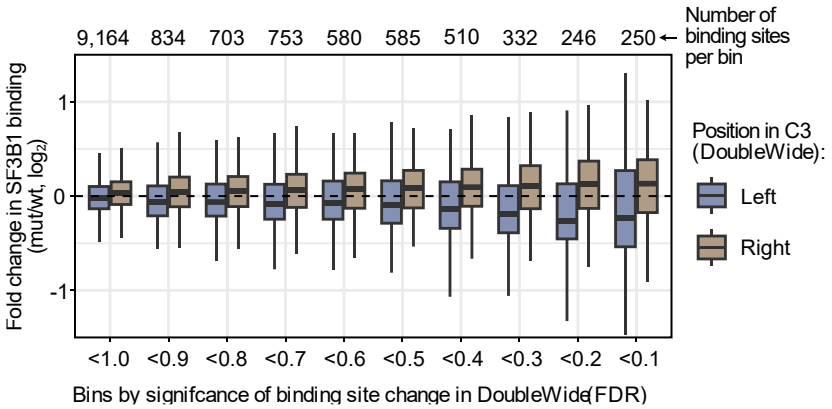
Supplemental Fig. S33. The SF3B1 iCLIP signal within the binding sites was highly reproducible between replicates. Scatter plot shows the correlation of SF3B1 crosslink events per binding site between biological replicates, with Pearson correlation coefficients given in the opposing quadrant. Distribution of crosslink events per binding sites in each replicate are shown along the diagonal.



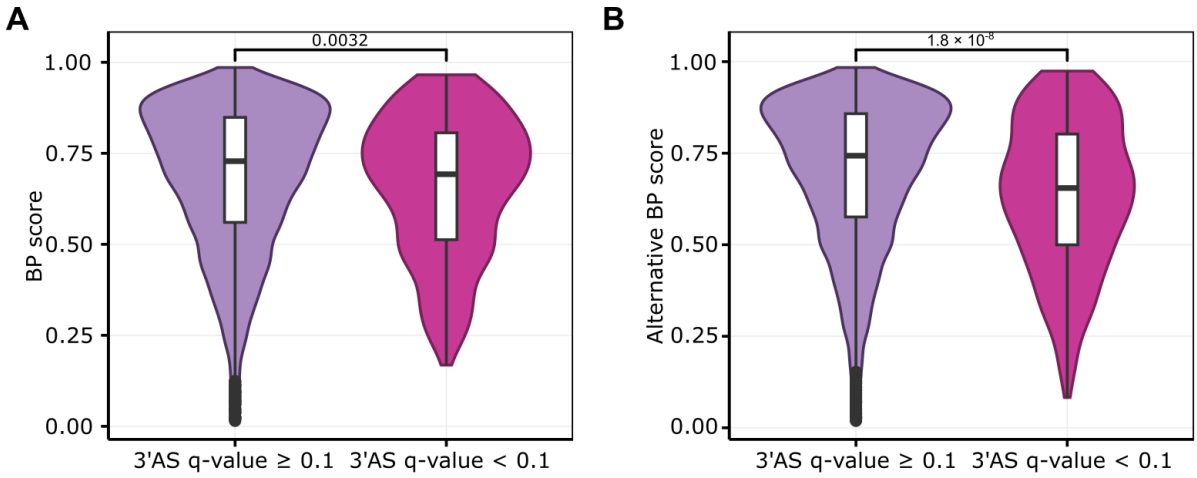
Supplemental Fig. S34. iCLIP signal within C3 DoubleWide region at *SETD4* exon/intron junction. The UCSC browser tracks show the iCLIP minus strand coverage (top) and Iso-Seq read coverage (bottom) in K562-SF3B1$^{K700E/wt}$ (red) and K562-SF3B1$^{K700E/wt}$ cells (blue).



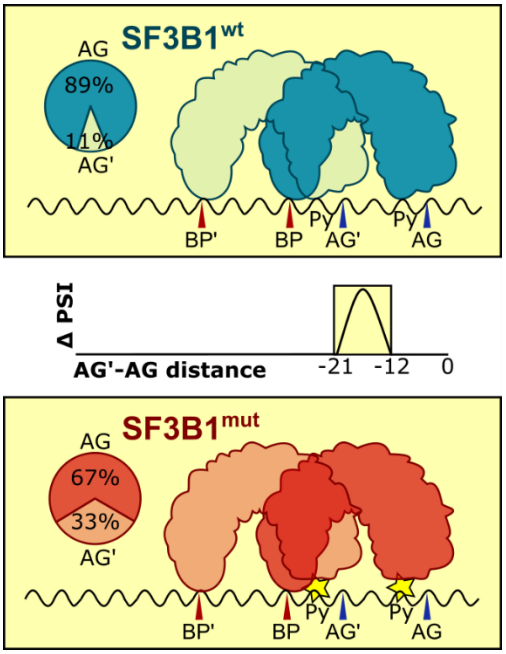
Supplemental Fig. S35. SF3B1-bound regions show distinct binding patterns. (***A***) Unsupervised clustering separates SF3B1-bound regions into three distinct clusters (C1, n = 35,907 regions; C2, n = 5,635; C3, n = 13,847). The SF3B1 crosslink coverage (sum of all replicates) in 81-nt windows around SF3B1 binding sites was subjected to min-max normalization and spline-smoothing, followed by dimension reduction using uniform manifold approximation and projection (UMAP) and density-based clustering of applications with noise (DBSCAN). Regions in cluster C0 (n = 835) could not be assigned to any of the fitted density centers and were excluded from further analysis. (***A,B***) Regions of cluster C3 show two broadly spaced modes at variable distance. (***B***) UMAP orders regions by increasing distances. Smoothing of the regions in cluster C3 was repeated with higher resolution, followed by a second round of UMAP and DBSCAN, yielding 33 clusters. Clusters were numbered based on the increasing distance between the two modes (indicated in brackets). (***C***) Ridge plot displays smoothed SF3B1 crosslink coverages, averaged over the regions within each cluster.



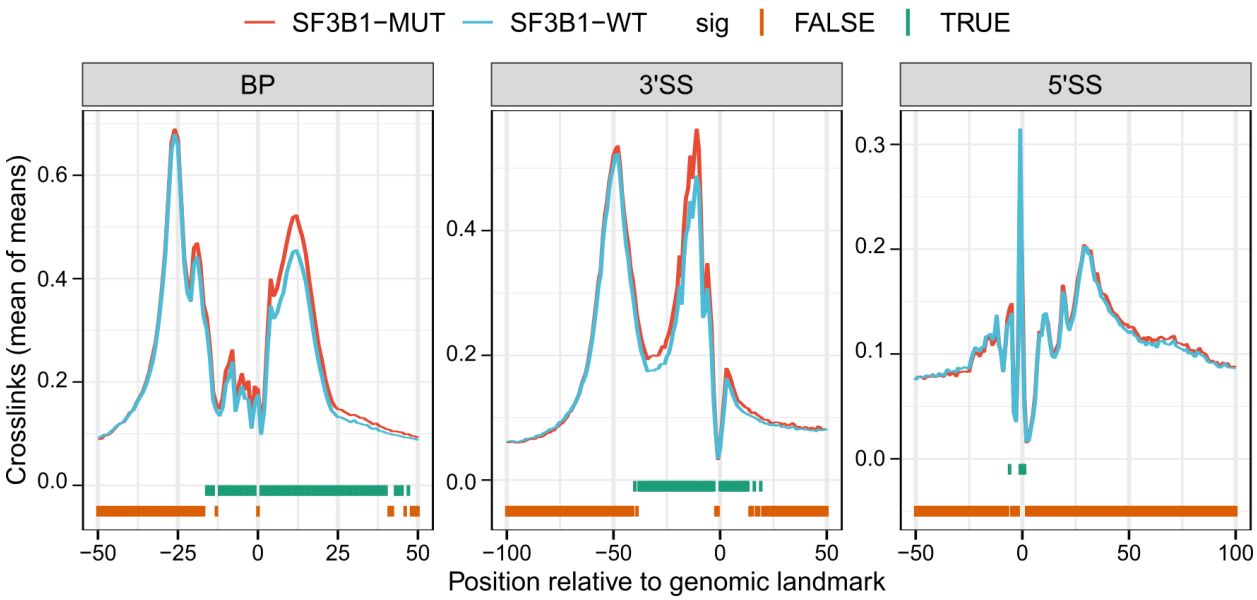
Supplemental Fig. S36. SF3B1 binding differences between K562-SF3B1$^{K700E/wt}$ and K562-SF3B1$^{wt/wt}$ among cluster C3 DoubleWide regions.



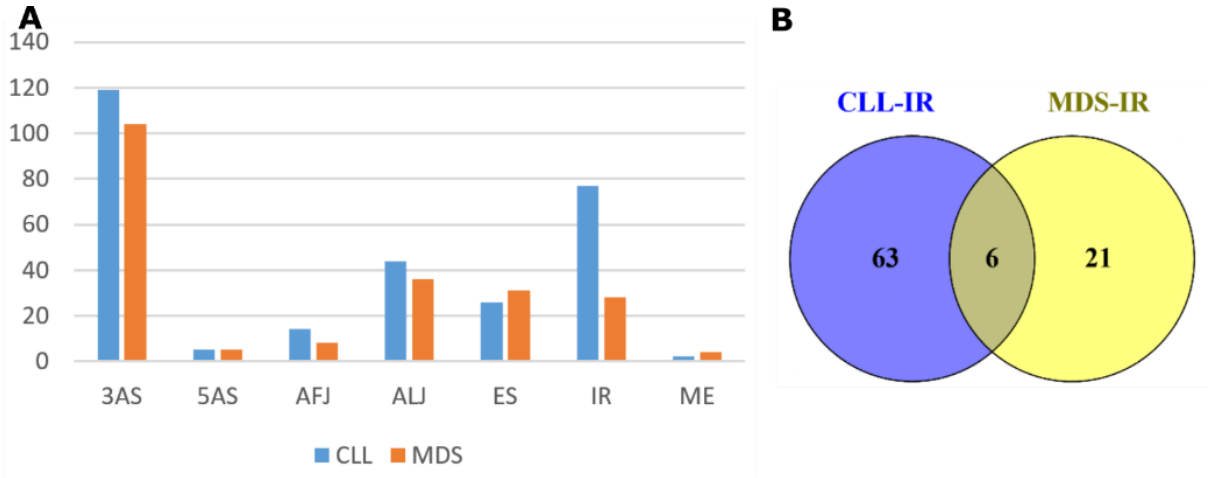
Supplemental Fig. S37. Branch points (BP) predicted for canonical (***A***) and alternative (***B***) AGs among significant and non-significant 3' alternative splicing events detected in *SF3B1*$^{mut/wt}$ vs. *SF3B1*$^{wt/wt}$ samples based on Iso-Seq data.



Supplemental Fig. S38. The model of *SF3B1* mutation effect on splicing. Within a short AG'–AG distance, in SF3B1$^{mut/wt}$ cells preferably the first (upstream) AG (AG') is used compared to SF3B1$^{wt/wt}$. This process often requires a strong downstream canonical AG. The switch between AG' and AG may be explained by changes in the second SF3B1 mRNA binding pocket that contains the most frequent K700E mutation that lower the number of contacts between mRNA and E700 of SF3B1, resulting in increased mobility of the mRNA (indicated by the star). The percentages for upstream (AG') and downstream (AG) AGs show the median PSI values calculated for all events with -12 nt ≤ AG'–AG distance ≥ -21 nt. BP – branch point, Py – polypyrimidine tract.



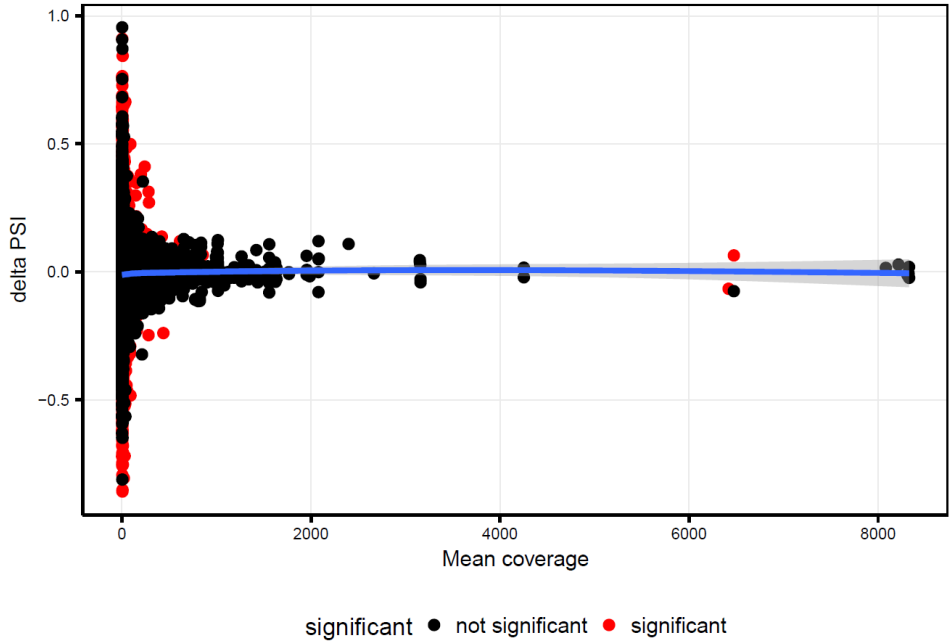
Supplemental Fig. S39. SF3B1 binding meta-profiles based on the iCLIP experiment with K562-SF3B1$^{K700E/wt}$ and parental K562-SF3B1$^{wt/wt}$ cells. The signal was aligned to branch point (BP), 3' splice site (3'SS) or 5' splice site (5'SS).



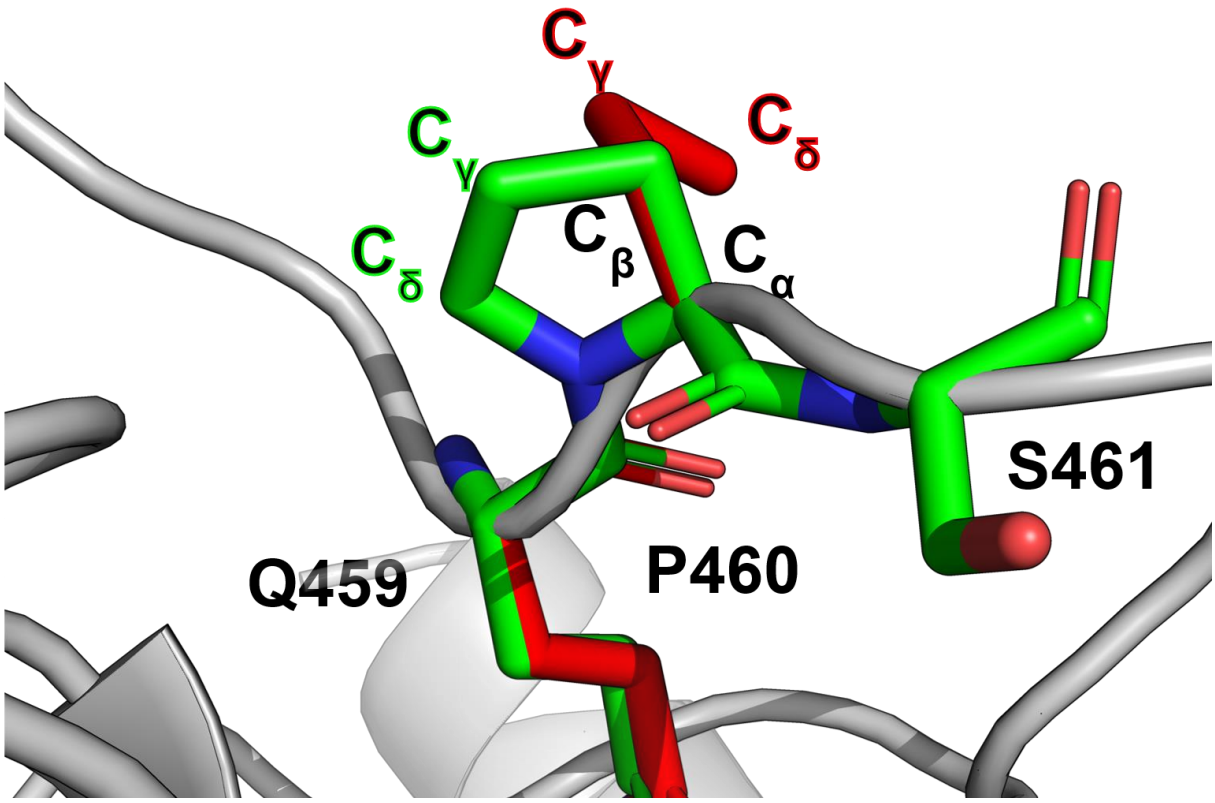
C

MDS (q<0.05)							
p-value	q-value	pathway	source	external_id	size	effective_size	overlap
1.65E-05	0.007676	Amino acid transport across the plasma membrane	Reactome	R-HSA-352230	32	32	5 (15.6%)
CLL (q<0.05)							
p-value	q-value	pathway	source	external_id	size	effective_size	overlap
4.02E-06	0.00209	Processing of Capped Intron-Containing Pre-mRNA	Reactome	R-HSA-72203	241	241	13 (5.4%)
3.66E-05	0.006214	mRNA Splicing - Major Pathway	Reactome	R-HSA-72163	177	177	10 (5.6%)
3.83E-05	0.006214	Metabolism of RNA	Reactome	R-HSA-8953854	584	583	19 (3.3%)
5.32E-05	0.006214	mRNA Splicing	Reactome	R-HSA-72172	185	185	10 (5.4%)
5.98E-05	0.006214	Spliceosome - Homo sapiens (human)	KEGG	path:hsa03040	151	151	9 (6.0%)
0.000376	0.032553	COPI-mediated anterograde transport	Reactome	R-HSA-6807878	83	83	6 (7.2%)
0.000484	0.033442	Toll Like Receptor 3 (TLR3) Cascade	Reactome	R-HSA-168164	87	87	6 (6.9%)
0.000547	0.033442	TRIF(TICAM1)-mediated TLR4 signaling	Reactome	R-HSA-937061	89	89	6 (6.7%)
0.00058	0.033442	MyD88-independent TLR4 cascade	Reactome	R-HSA-166166	90	90	6 (6.7%)
0.000643	0.033442	mRNA Processing	Wikipathways	WP411	127	127	7 (5.5%)
0.000881	0.038168	MAP kinase activation	Reactome	R-HSA-450294	65	65	5 (7.7%)
0.000881	0.038168	Interleukin-17 signaling	Reactome	R-HSA-448424	65	65	5 (7.7%)
0.000964	0.038574	ER to Golgi Anterograde Transport	Reactome	R-HSA-199977	136	136	7 (5.1%)
0.001505	0.049081	TNFalpha	NetPath	Pathway_TNFalpha	234	234	9 (3.8%)
0.00176	0.049081	Herpes simplex virus 1 infection - Homo sapiens (human)	KEGG	path:hsa05168	498	498	14 (2.8%)
0.001815	0.049081	Toll Like Receptor 4 (TLR4) Cascade	Reactome	R-HSA-166016	112	112	6 (5.4%)
0.001888	0.049081	MyD88 cascade initiated on plasma membrane	Reactome	R-HSA-975871	77	77	5 (6.5%)
0.001888	0.049081	Toll Like Receptor 10 (TLR10) Cascade	Reactome	R-HSA-168142	77	77	5 (6.5%)
0.001888	0.049081	Toll Like Receptor 5 (TLR5) Cascade	Reactome	R-HSA-168176	77	77	5 (6.5%)
0.001888	0.049081	TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation	Reactome	R-HSA-975138	77	77	5 (6.5%)
0.001999	0.049491	MyD88 dependent cascade initiated on endosome	Reactome	R-HSA-975155	78	78	5 (6.4%)
0.002114	0.049975	Toll Like Receptor 7/8 (TLR7/8) Cascade	Reactome	R-HSA-168181	79	79	5 (6.3%)

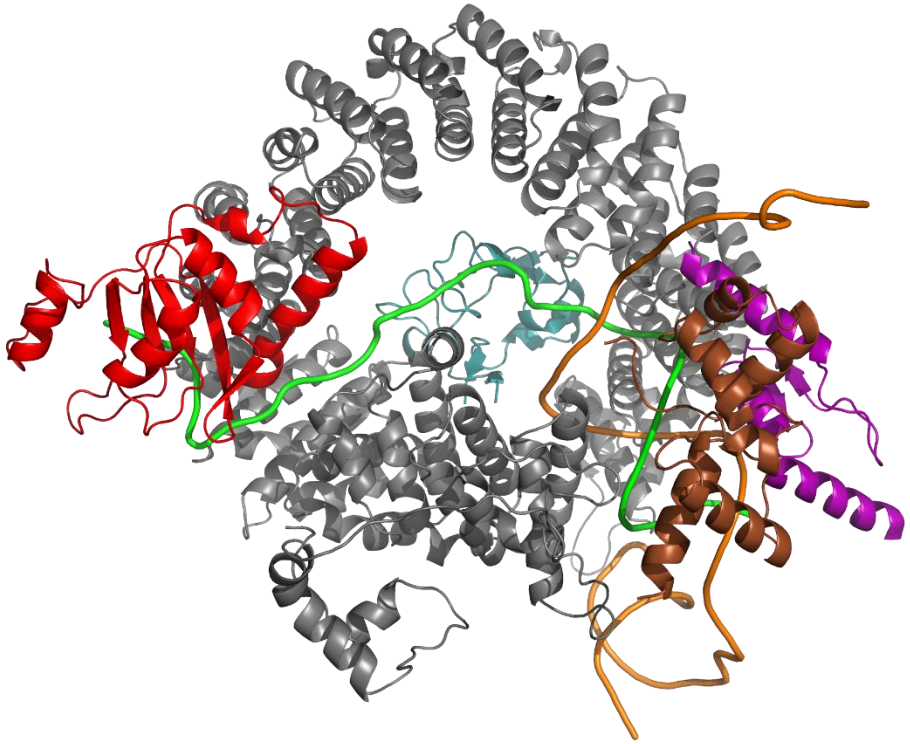
Supplemental Fig. S40. (***A***) The number of ASEs called in CLL and MDS patients separately. (***B***) the overlap between intron retention (IR) events called in CLL and MDS patients associated with SF3B1 mutation. (***C***) Over-representation of the CLL-specific and MDS-specific ASEs yields multiple pathways being affected in CLL but only a single pathway being affected in MDS ($q < 0.05$).



Supplemental Fig. S41. Delta PSI values of the splicing events detected in all three datasets shown as the function of the mean expression across all samples. The smoothed line was calculated with gam method.



Supplemental Fig. S42. Incorrect proline conformation after side chain filling using Maestro[4] (red) with a missing bond between C_δ and N. After further optimization using MOE[5] (green) C_γ is correctly bound to N. P460 of SF3B1 is shown exemplarily.



Supplemental Fig. S43. Final model used for molecular dynamics simulations featuring SF3B1 (grey), the pre-mRNA (green), the PHD finger-like domain containing protein 5A (light blue), the RNA-binding motif protein, x-linked 2 (red), the cell division cycle 5-like protein (brown), the splicing factor 3A subunit 2 (purple), and the u2-snRNA (orange).

References:

Embree CM, Abu-Alhasan R, Singh G. 2022. Features and factors that dictate if terminating ribosomes cause or counteract nonsense-mediated mRNA decay. *Journal of Biological Chemistry* **298**. https://doi.org/10.1016/j.jbc.2022.102592.

Kozak M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**: 8125–8148. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/15.20.8125.

Lienhard M, van den Beucken T, Timmermann B, Hochradel M, Börno S, Caiment F, Vingron M, Herwig R. 2023. IsoTools: a flexible workflow for long-read transcriptome sequencing analysis. *Bioinformatics* **39**: btad364. https://doi.org/10.1093/bioinformatics/btad364.

Pardo-Palacios FJ, Wang D, Reese F, Diekhans M, Carbonell-Sala S, Williams B, Loveland JE, De María M, Adams MS, Balderrama-Gutierrez G, et al. 2024. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods*. https://doi.org/10.1038/s41592-024-02298-3.