PNAS RESEARCH ARTICLE | BIOCHEMISTRY OPEN ACCESS

Cancer-associated SF3B1 mutation K700E causes widespread changes in U2/branchpoint recognition without altering splicing

Andrey Damianov[a,1], Chia-Ho Lin[a], Jian Zhang[b], James L. Manley[b], and Douglas L. Black[a,1]

Edited by Donald Rio, University of California Berkeley, Berkeley, CA; received November 17, 2024; accepted February 21, 2025

Myelodysplastic syndromes and other cancers are often associated with mutations in the U2 snRNP protein SF3B1. Common SF3B1 mutations, including K700E, disrupt SF3B1 interaction with the protein SUGP1 and induce aberrant activation of alternative 3′ splice sites (ss), presumably resulting from aberrant U2/branch site (BS) recognition by the mutant spliceosome. Here, we apply a method of U2 IP-seq to profile BS binding across the transcriptome of K562 leukemia cells carrying the *SF3B1* K700E mutation. For alternative 3′ ss activated by K700E, we identify their associated BSs and show that they are indeed shifted from the WT sites. Unexpectedly, we also identify thousands of additional changes in BS binding in the mutant cells that do not alter splicing. These new BSs are usually very close to the natural sites, occur upstream or downstream, and either exhibit stronger base-pairing potential with U2 snRNA or are adjacent to stronger polypyrimidine tracts than the WT sites. The widespread imprecision in BS recognition induced by K700E with limited changes in 3′ ss selection expands the physiological consequences of this oncogenic mutation.

pre-mRNA splicing | myelodysplastic syndrome | U2 snRNP | intron branchpoint

Pre-mRNA splicing is an essential step in gene expression, and mutations in general splicing factors cause many forms of human disease (1–3). These mutations are mostly missense, usually cause slight alteration of function, and when assessed across the transcriptome may induce only modest changes in splicing. The core spliceosomal protein SF3B1 is commonly mutated in cancers (1, 4), with recurrent point mutations affecting residues of its HEAT repeat domains H4–H12 (1, 4–7). *SF3B1* mutations are found in approximately 30% of patients with myelodysplastic syndrome (MDS) (1, 3), with one mutation, K700E, accounting for more than 50% of these (5, 8). K700E and other *SF3B1* mutations, which are always heterozygous, induce recognition of cryptic 3′ splice sites (ss) in a small fraction of introns across the transcriptome. These new 3′ ss are often located upstream of the wild-type (WT) 3′ ss and are associated with an alternative upstream branch site (BS) (7, 9–12).

Recognition of the BS occurs in several steps during early spliceosome assembly. Initially, the proteins SF1, U2AF2, and U2AF1 recognize the BS, the polypyrimidine tract, and the AG dinucleotide of the 3′ ss, respectively. These early factors recruit the U2 snRNP (U2), which after structural rearrangements forms the prespliceosomal A-complex with the BS base-paired to the U2 snRNA, leaving the future branchpoint (BP) adenosine bulged from the U2/BS helix (13–16). In both the A complex and an earlier complex prior to BS pairing, SF3B1 is in contact with U2AF and with intron sequences flanking the BS, thereby stabilizing U2 binding to the BS (17). There is substantial interest in how mutations in SF3B1 alter BS recognition, both for their relevance to cancer and for gaining insight into SF3B1 function within the spliceosome.

The K700E mutation has been shown to disrupt the interaction of SF3B1 with the splicing factor SUGP1, whose depletion or mutation recapitulates to a significant degree the splicing changes observed in SF3B1-K700E cells (10, 18, 19). SUGP1 is a G patch-containing protein that can bind to and activate DHX15, a DEAH-box RNA helicase proposed to function in several steps of spliceosome assembly and disassembly (20–23). Supporting the significance of the SUGP1–DHX15 interaction with respect to mutant SF3B1 splicing defects, a DHX15-SUGP1 G-patch fusion protein was found to rescue several aberrant splicing events (20). SUGP1 is thought to act early in BS selection where it directly contacts SF3B1, and likely also SF1 and/or U2AF2 (10). By one model, SUGP1 then stabilizes prespliceosomes prior to BS recognition by activating DHX15 to displace SF1, thereby allowing the future BS to base-pair with U2 snRNA. In the absence of SUGP1, SF1 may not be efficiently displaced diverting assembly onto a cryptic BS if available (24).

Significance

Myelodysplastic syndrome (MDS) and certain cancers are associated with mutations in the protein SF3B1, which is a subunit of the U2 snRNP. The U2 snRNP plays an important role in the pre-mRNA splicing process, which is essential to the expression of eukaryotic genes. Using a method of whole genome analysis, this study shows that one common MDS mutation in SF3B1 causes more widespread disruption of U2 interactions than previously recognized. These results will be important both for understanding the role of SF3B1 in U2 function and how its disruption causes cancer.

Author affiliations: [a]Department of Microbiology, Immunology, and Molecular Genetics, Molecular Biology Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; and [b]Department of Biological Sciences, Columbia University, New York, NY 10027

Author contributions: A.D. and D.L.B. designed research; A.D. performed research; A.D., J.Z., and J.L.M. contributed new reagents/analytic tools; A.D., C.-H.L., and D.L.B. analyzed data; and A.D., J.Z., J.L.M., and D.L.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: damianov@microbio.ucla.edu or dougb@microbio.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2423776122/-/DCSupplemental.

Published March 26, 2025.

Other proteins affecting U2 assembly onto the pre-mRNA have also been implicated in aberrant BS selection by SF3B1 mutants. The G-patch protein GPATCH8 was also shown to modulate the activity of DHX15 and to functionally oppose the action of SUGP1 on approximately 30% of the splicing events altered in SF3B1-K700E mutant cells. This was suggested to reflect a competition between SUGP1 and GPATCH8 for DHX15, such that upon GPATCH8 depletion, the excess free DHX15 could associate with mutant SF3B1 and partially rescue its splicing defects (25). Two other RNA helicases, the DEAD-box proteins DDX42 and DDX46 that play roles in initial assembly of the U2 on the pre-mRNA, were also proposed to affect BS selection by mutant SF3B1, although this has not been established (26–28). Thus, it is likely that K700E and other SF3B1 mutations affect the interactions of multiple spliceosomal proteins across the early steps of spliceosome assembly.

It has so far been difficult to directly assess BS use in the presence of *SF3B1* mutations (7, 11). Mammalian BSs are highly degenerate and vary in their distance from the 3′ ss, making them difficult to predict with certainty. Introns can also have multiple BPs that are used at different frequencies. Large-scale approaches to BP identification have used sequencing of the lariat intermediates or products of splicing, or applied machine-learning algorithms trained on large lariat sequencing datasets (29–33). These approaches have identified hundreds of thousands of BPs in human cells and tissues; however, they lack the capacity to quantify differential BS use between samples (30).

We recently developed U2 IP-seq as a whole-transcriptome method for isolating and sequencing BS's base-paired with U2 snRNA in the precatalytic spliceosome (34). This method yields many unique BS RNA fragments per intron whose mapping and quantification allows comparison of BS use across different samples. To examine comprehensively how mutant SF3B1 alters BS selection, we applied IP-seq to map sites bound by U2 in K562 cells expressing SF3B1 K700E and compared these maps to those for isogenic WT SF3B1 cells. We find that the K700E mutation leads to far more widespread changes in U2/BS interactions than previously anticipated, with a large majority of these changes not affecting 3′ ss choice. By comparing sites showing differential U2 binding, we identify features that distinguish BSs favored by mutant SF3B1. Our findings provide unexpected insights into both the normal functioning of the spliceosome and how *SF3B1* mutations alter this function.

Results

U2 IP-Seq Identifies Numerous Shifts in Branch Site Recognition in MDS Mutant Cells. To obtain pre-mRNA-bound U2 complexes containing WT or mutant SF3B1, we isolated nuclei from two CRISPR-engineered sublines of K562 myelogenous leukemia cells. In each of these cell lines, one of the *SF3B1* alleles was gene-edited with either the WT sequence carrying synonymous mutations or with the K700E sequence, each fused to an N-terminal His6-FLAG epitope tag (10). Following the protocol previously described for HEK293 cells, the high-molecular-weight (HMW) nuclear fraction containing chromatin was separated from the soluble nucleoplasm, and chromatin-bound complexes were extracted with a combination of DNAse and RNAse (34). The nucleoplasmic fraction was treated in parallel. We immunopurified U2 complexes from both fractions with either anti-Flag or an antibody binding a C-terminal peptide of SF3A2 (Fig. 1 and *SI Appendix*, Fig. S1 *A* and *B*). The protein profiles of complexes purified from the HMW fraction of WT or SF3B1-K700E cells were similar (Fig. 1 and *SI Appendix*, Fig. S1 *A* and *B*, *Right*), and resembled those previously isolated from 293Flp-in cells (34). As seen previously, the SF3A and SF3B subunits were both abundant in the complexes isolated from the HMW fraction, as were the U2 snRNP-associated proteins DHX15 and RBM17, while the Sm core and U2 A′/B″ proteins were lost after nuclease digestion. PRPF8 and some other U5 components of the tri-snRNP were observed at lower stoichiometry in the immunopurified complexes. We did not detect SUGP1 protein in any of the complexes (*SI Appendix*, Fig. S1*A*). As seen previously, RBM5 and RBM10, two other G-patch motif-containing proteins associated with U2 were detected at substoichiometric levels (34). These two proteins showed equal association with U2 complexes containing either WT or mutant SF3B1 (*SI Appendix*, Fig. S1*B*), although the RBM5/10 isolation was more efficient with anti-Flag antibody targeting the epitope-tagged SF3B1 than with the anti-SF3A2 IgG. Interestingly, immunopurified complexes isolated from the nucleoplasm contained only SF3A or SF3B subunits but not both. This may indicate that the nuclease treatment causes the disassociation of SF3A and SF3B from the U2 snRNP in this fraction, which is likely in the precatalytic 17S form or another complex unbound to pre-mRNA (35) (Fig. 1 and *SI Appendix*, Fig. S1 *A* and *B*, *Left*).

We isolated RNA from the HMW-extracted complexes and the U2 snRNA was removed from these samples by antisense oligonucleotide-directed RNase H degradation. The remaining population of protected RNA fragments were of similar size to those observed previously (*SI Appendix*, Fig. S2). These fragments were converted to cDNA, sequenced, and mapped to the human genome as described (34). As expected, the majority of the reads mapped in clusters near the 3′ ends of introns, corresponding to BSs protected by precatalytic spliceosomes (*SI Appendix*, Table S1 and Fig. S3*A*).

Methods that physically map the BP nucleotide (nt) after the first catalytic step have limited sensitivity that prevent them from being comprehensive (29, 30, 32, 33). Approaches using machine learning to predict BP nucleotides provide more complete maps but may miss secondary BPs adjacent to the strongest prediction (31, 32). To identify the putative BP nucleotides within the reads of a U2 IP-seq cluster, we previously developed an algorithm that scores the match to consensus and the position within each protected RNA fragment (34). These predicted future BP nucleotides from isolated precatalytic BSs strongly correlate with BPs identified in studies of lariat intermediates, particularly for BPs detected with high frequency within the BS clusters (34). This correlation can be seen by comparing the tracks of the IP-seq predicted BPs with tracks of previously identified BPs in Fig. 2 and *SI Appendix*, Fig. S3. Note that these BP nucleotides are predicted from RNA fragments isolated prior to catalysis. It is possible that some catalysis events occur a few nucleotides distant from the prediction, but given the alignment of the predicted BPs with the BPs mapped after the first step, this is likely a small fraction. IP-seq produces homogeneous peaks of reads at each BS with a low background, and analysis of individual reads within each cluster allows identification of multiple BPs in many introns. The relative use of these alternative BPs was calculated as the ratio of reads supporting a particular BP to the total number of reads within the cluster. Detected BPs were graphically represented in browser tracks as vertical bars with the bar height proportional to this ratio (Fig. 2 and *SI Appendix*, Figs. S3 *A–F* and S4). IP-seq with the anti-SF3A2 antibody was performed in triplicate in both the WT and K700E cells and yielded highly reproducible results (Fig. 2 and *SI Appendix*, Fig. S5*A*). For comparison, we also isolated U2 via the Flag-tag on SF3B1 from both WT and K700E cells. This yielded very similar peak

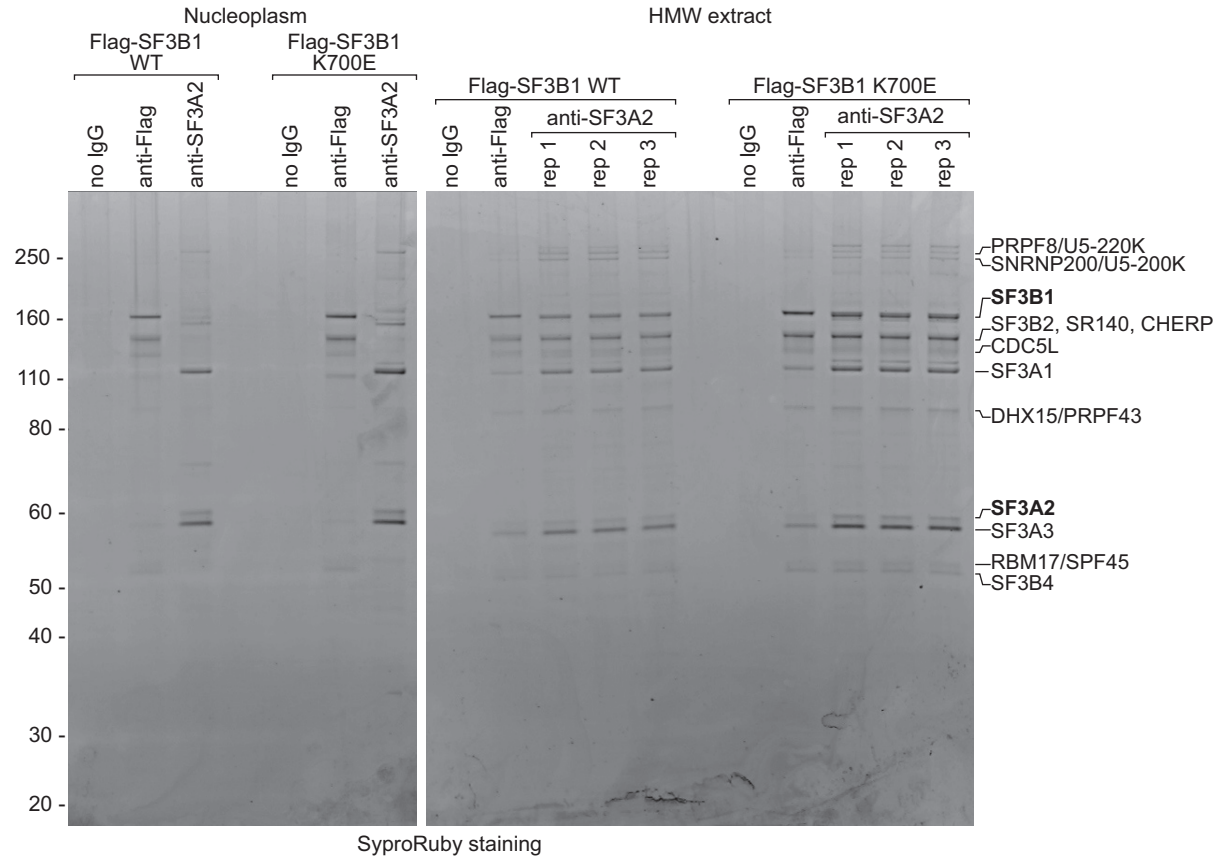

Fig. 1. SF3B1 WT and K700E mutant proteins copurify with the same sets of U2 snRNP proteins in RNase-resistant complexes from the HMW nuclear fraction. Protein profiles of U2 complexes purified from soluble nucleoplasm (*Left*) or HMW nuclear material (*Right*) following extraction with RNase and Benzonase. Complexes containing SF3B1-Flag protein or SF3A2 were purified from WT and K700E cells with the indicated antibodies. Incubation with agarose beads without antibody served as negative control (no IgG lanes). The major coprecipitating proteins detected by protein staining are indicated on the *Right*.

profiles across the transcriptome (Fig. 2 and *SI Appendix*, Fig. S3*A*).

BP positions across the transcriptome were found to be very similar between WT and SF3B1-K700E cells (*SI Appendix*, Fig. S5*B*). To specifically examine BPs upstream of differentially utilized 3′ splice sites (36), alternative splicing events were profiled using rMATS (37). This identified 531 alternative 3′ ss events that changed between WT and K700E (Dataset S1). To compare these events with the IP-seq data, we removed events with low read coverage, or where the alternative 3′ ss co-occurred with other splicing patterns, such as exon skipping or intron retention. This produced a list of 156 events where both splicing choice and BP position could be confidently determined (Fig. 3 *A* and *B*, *SI Appendix*, Fig. S6, and Dataset S2). For the majority of alternative 3′ ss events (142), we detected a shift in BP selection between the WT and K700E cells. Most of these splicing events involved the mutant protein shifting to an upstream BS and activating an upstream 3′ss (see examples with positive ΔPSI values in Fig. 3*A*, *SI Appendix*, Fig. S6, and Dataset S2). There were a minority of events (42) where the K700E protein shifted to a downstream BS and activated a downstream 3′ ss (negative ΔPSI values in Fig. 3*A*, *SI Appendix*, Fig. S6, and Dataset S2). For 3′ splice sites activated in mutant cells, we did not find a notable difference in the average BP to AG dinucleotide distance (Fig. 3*A* and *SI Appendix*, Fig. S6). These results are consistent with previous RNAseq analyses of splicing changes induced by mutant SF3B1, where new 3′ splice sites used in mutant cells were usually upstream (7, 9–12). In most cases, the distance between the WT and mutant BP ranged from a few to less than 100 nt. There were a few exceptions with WT to mutant BP distances over 500 nt that may involve different mechanisms for recruitment of U2 to the new site (*SI Appendix*, Figs. S3*F* and S6). The distances between the two 3′ss correlated strongly with these BP distances (*SI Appendix*, Fig. S6). This is predicted if the first AG downstream from the activated BP is chosen for splicing in the second catalytic step, as shown in earlier work (38, 39). The exception to this correlation was for BP to BP distances shorter than 20 nt, where the AG to AG distances remained clustered around 20 nt (Fig. 3*B* and *SI Appendix*, Fig. S6). This presumably reflects the exclusion of AGs from the polypyrimidine tracts upstream from the WT AG. In these cases, the new AG is often within the protected fragment of the WT SF3B1 (Fig. 2 and *SI Appendix*, Fig. S3*C*).

We also identified 14 alternative 3′ ss events where no alteration of BP use was detected despite the splicing shift (Dataset S2, example shown in *SI Appendix*, Fig. S4). However, examination of the BS clusters for some of these events indicated that reads for alternative BPs were present but were not sufficiently abundant to be reported as significant in our BP calling protocol (34).

It should be noted that in the cells expressing the mutant SF3B1-K700E allele, the SF3A2 antibody will isolate U2 complexes containing the products of both the mutant and WT *SF3B1* alleles. In contrast, IP-seq with the anti-Flag antibody will isolate only BSs bound by either the monoallelically tagged mutant or WT SF3B1 depending on the gene-edited cell line used. When the new BP was sufficiently distant to generate a new read cluster, the anti-Flag IP-seq usually generated a taller peak (*SI Appendix*, Fig. S3 *B* and *D*). However, even when K700E BP reads overlapped the WT reads within the same cluster, the new predicted BPs were called in both the anti-Flag and anti-SF3A2 data (Fig. 2

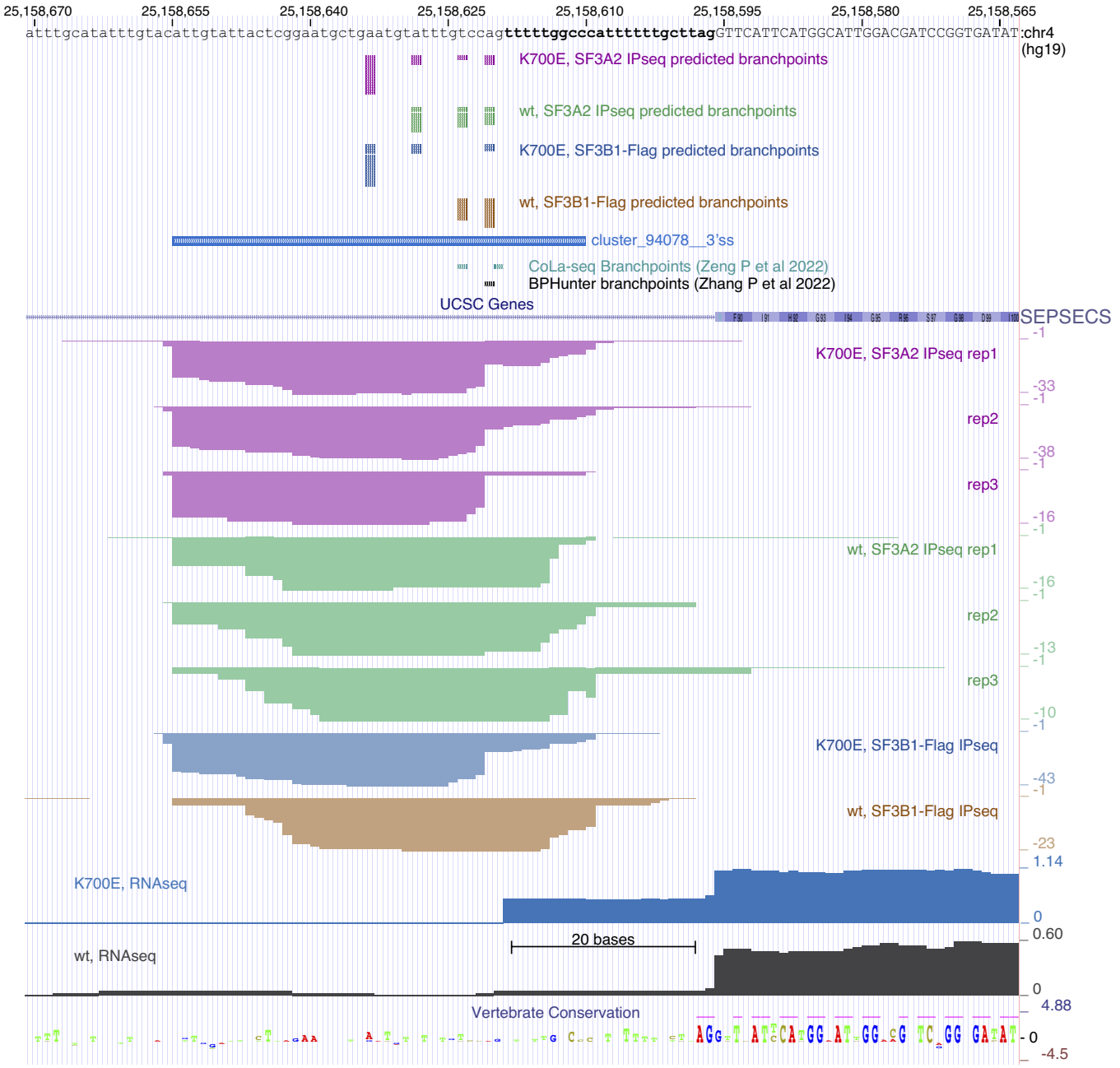

Fig. 2. A branch site is selectively engaged with U2 snRNP in K562 SF3B1-K700E mutant cells, upstream of a cryptic 3′ splice site activated by this mutation. UCSC Genome browser view of the SEPSECS exon 4 and the upstream intron. Total read coverage is shown below the gene diagram. Individual IP-seq tracks show RNA recovered from the SF3B1-Flag WT or K700E mutant complexes, or from SF3A2-containing complexes from cells expressing the WT or mutant SF3B1 protein, each in triplicate. The significant branch site cluster of overlapping reads is marked in a track above the peaks, as are tracks of branch points identified by others (29, 32, 33). Precatalytic branch points predicted from individual IP-seq reads are shown above as vertical bars with the bar height proportional to the read numbers within the cluster as described in ref. 34. Tracks of RNAseq data for the mutant and WT cells are at the bottom showing alternative 3′ splice site use. Other gene annotation and conservation tracks are shown as indicated.

and *SI Appendix*, Fig. S3 *C* and *E*). The identification of the K700E-dependent BPs by SF2A2 IP-seq highlights the sensitivity of this method. MDS mutations are always heterozygous. The ability of the SF3A2 antibody to detect allele-specific BPs bound by endogenous U2 indicates that the method should allow assays in heterozygous cell systems without epitope tagging of U2 components.

In addition to altered 3′ ss choices, the K700E mutation induced many other splicing changes, including cassette exons and retained introns (Dataset S1). One interesting example reported previously was activation of poison exon 14a in the *BRD9* gene, observed in various cancers carrying the SF3B1-K700E mutation (40). This exon was included at low levels in WT K562 cells and induced in the K562 mutant (*SI Appendix*, Fig. S7*A* and Dataset S1). The BPs upstream of exons 14a and 15 previously identified in a MEL270 uveal melanoma cell line expressing SF3B1-K700E were also observed in the K562 K700E cells (40) (*SI Appendix*, Fig. S7 *A* and *B*). How mutant SF3B1 might alter cassette exon splicing is poorly understood and such effects could be indirect. However, we noted an additional exon 14a BP specific for the WT cells downstream of that used in K700E cells (*SI Appendix*, Fig. S7*B*). This WT BP likely activates a downstream 3′ss, resulting in a shorter exon14a, which also contains a stop codon. Although it did not pass the significance threshold in

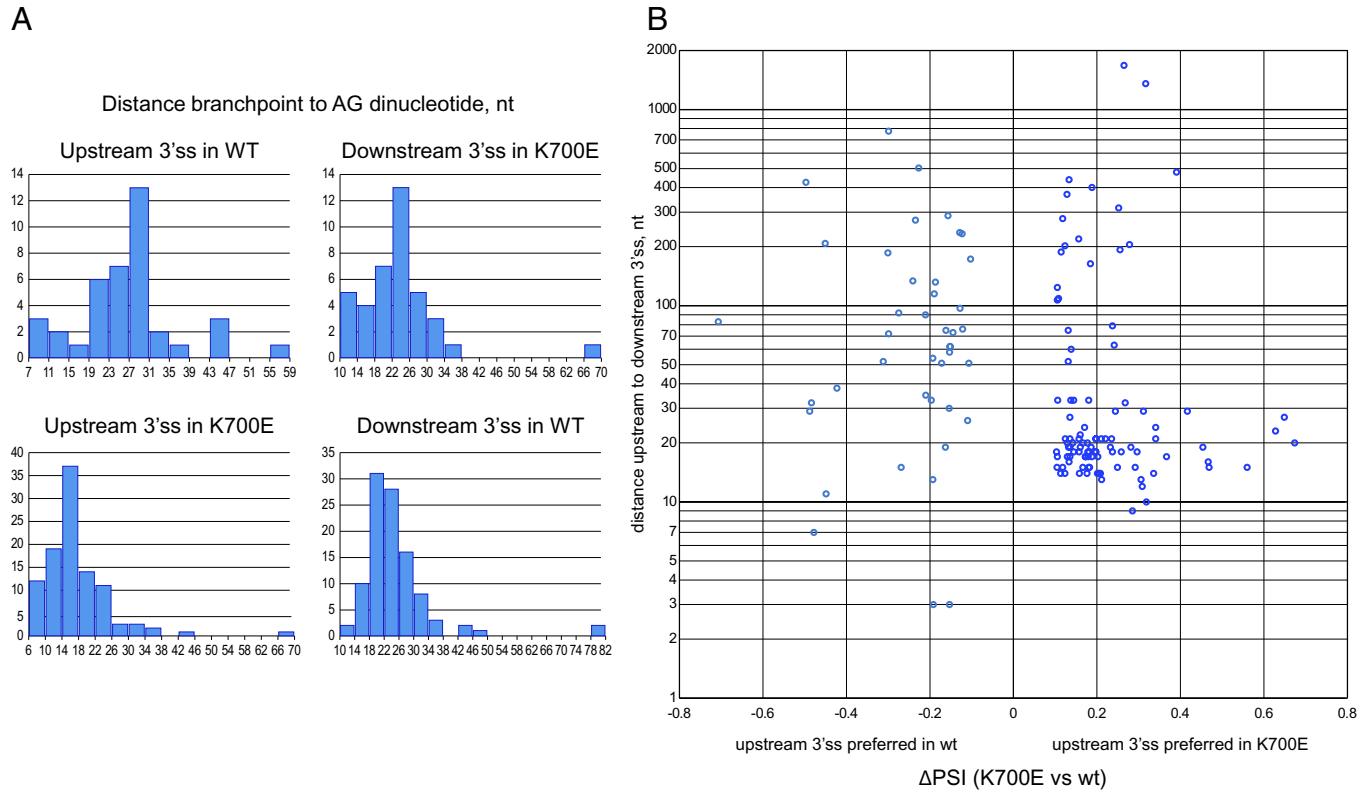

Fig. 3. Distances between 3′ splice sites and associated branchpoints differentially used in K562 SF3B1 WT and K700E mutant cells. (*A*) Distribution of the nucleotide distances between the preferred branchpoint and the AG dinucleotide at the end of the intron is plotted for each alternative 3′ splice site used. (*B*) Distances between differential 3′ss are plotted against the upstream 3′ss inclusion levels, expressed in ΔPSI (percent spliced-in).

rMATS analysis, splicing to this 3′ ss was detectable in the RNAseq tracks of the WT cells (*SI Appendix*, Fig. S7*B*). Thus, splicing of *BRD9* exon 14a may be activated via altered BP recognition by the SF3B1-K700E mutant, similar to events involving cryptic 3′ splice sites, rather than an indirect effect of other factors.

The K700E Mutation Induces Thousands of Shifts in Branchpoints without a Change in the 3′ Splice Site. Most alternative BPs found in SF3B1-K700E cells have been identified through analysis of RNAseq data that identified changes in 3′ ss usage. We next examined the IP-seq data for BP changes that may occur without altering ss choice. To search for such sites, we applied stringent criteria to create comparison sets of WT and K700E BPs. We identified the most frequently used BP in each IP-seq cluster of the K700E mutant cells. We then removed clusters where the most frequent BP in the mutant cells coincided with the most frequent BP in WT cells. For the remaining clusters, if a BP was present in WT cells at the same position as the most frequent one in mutant cells, we required that its frequency be less than half that seen in SF3B1-K700E. Strikingly, this analysis identified nearly four thousand BPs whose use in mutant cells was greater than in WT cells (Dataset S3). Only 45 (1.14%) of these altered BPs were in close proximity to an alternative 3′ ss activated in K700E cells (within 300 nt upstream). Similarly, only 1.5% or fewer of these BPs were associated with cassette exons or other alternative splicing events (*SI Appendix*, Fig. S8 and Dataset S3). The vast majority of these differentially used BPs thus did not appear to alter splicing (see *SI Appendix*, Fig. S9 *A* and *B* for examples).

Within this set of stringently defined branch site clusters, we measured the nt distances from each preferred mutant BP to that preferred in WT. These BP pairs were more narrowly spaced than were those associated with alternative 3′ splice sites (compare Fig. 4 and Dataset S2 column R), with the majority of mutant-preferred BPs only 1 to 4 nucleotides from the major BP of WT cells (Fig. 4). At this distance, mutant-preferred sites were most frequently, but not exclusively, found downstream of the WT preferred sites. With this short separation, the nt regions pairing with U2 snRNA at each site partially overlap, constraining the allowable nt sequences between the two sites. To identify features of BSs preferred in the SF3B1-K700E cells, we therefore focused on those sufficiently distant from the preferred WT BP that the base-paired segments did not overlap. We selected 338 BP pairs where the mutant-preferred site was 7 to 14 nucleotides upstream of the WT BP, and 269 pairs where the mutant-preferred site was 7 to 14 nucleotides downstream of the WT (Fig. 4).

The U2 snRNA sequence GUAGUA can bind BSs in several base-pairing modes that differ in the positions of individual base pairs and the number of bulged nucleotides (30). To assess the relative use of different pairing modes in K562 cells, we examined a subset of K562 BSs that were independently mapped in two previous studies (29, 33). Comparing the base-pairing potentials of these sites in each of the described pairing modes, we found possible examples of all six modes but at relative frequencies that differed from those observed previously (30) (*SI Appendix*, Fig. S10 and Dataset S4). In particular, there were many fewer BPs predicted to use a C nucleotide instead of A than reported in the previous study. The canonical BS that uses U2 snRNA pairing mode 1 was favored over the other modes in K562 cells (*SI Appendix*, Fig. S10). The BPs preferred in alternative 3′ splice sites induced by mutant SF3B1 had similar distributions of modes for BPs both upstream and downstream from the WT site (*SI Appendix*, Fig. S11).

BPs that were closely spaced favored particular combinations of pairing modes, presumably due to sequence constraints on the nucleotides separating them. For BP pairs separated by 4 nucleotides, the most frequent pairing mode combinations were 1-1,

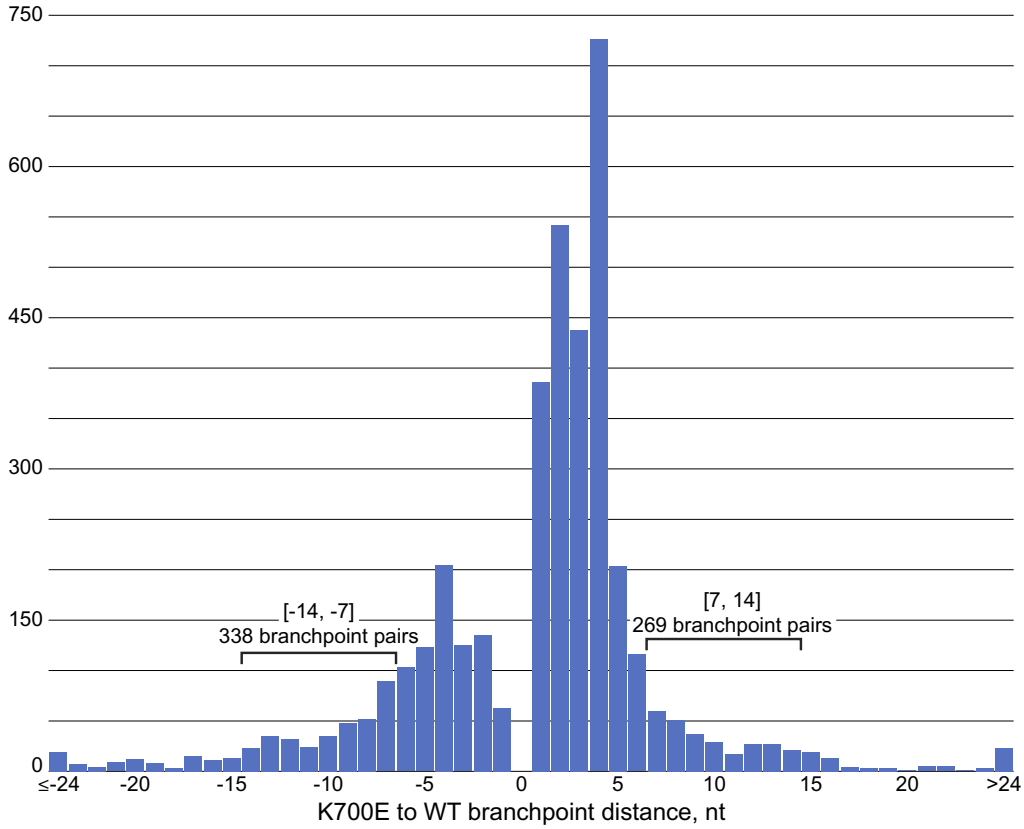

Fig. 4. Branch sites selected in SF3B1-K700E mutant cells can be upstream or downstream of the site used in WT cells. Distribution of distances between branchpoints preferred in K562 SF3B1 mutant cells and those preferred in WT. Negative distances indicate mutant sites located upstream of WT, positive distances indicate mutant sites located downstream of WT. Only branchpoints located within the same branch site cluster are analyzed. The ranges of distances for branch site pairs selected for downstream analyses are indicated by brackets.

4-1, and 5-1. However, the sites preferred by the mutant protein did not show a bias for a particular pairing mode. Similarly, the consensus sequences of these closely spaced BS pairs did not show strong differences between the upstream and downstream K700E sites (*SI Appendix*, Fig. S12).

K700E SF3B1 Favors Branchpoints with More Stable Pairing to U2 snRNA or with Stronger Polypyrimidine Tracts. We next assessed what intronic features, if any, might distinguish the BSs preferred by mutant SF3B1. To compare the U2 snRNA base-pairing potential of WT and mutant BSs separated by 7 to 14 nucleotides, we grouped them by pairing mode (*SI Appendix*, Fig. S13) and generated consensus sequences. For canonical mode 1 sites, upstream sites preferred in K700E cells exhibited more potential base-paired nucleotides with U2 snRNA than the downstream WT sites (Fig. 5 *A*, *Upper* row). Interestingly, the mutant-preferred sites downstream from WT did not necessarily exhibit greater base pairing than the WT site. Instead, these mutant-preferred downstream sites were adjacent to more prominent U-rich polypyrimidine tracts (Fig. 5 *A*, *Lower* row). Similar findings were made for the mutant and WT BS pairs associated with alternative 3′ splice sites (Fig. 5*B*). BSs assigned to other pairing modes often showed similar trends for the mutant sites, but these enrichments were based on fewer examples and were not as statistically significant (*SI Appendix*, Fig. S14).

Overall, we have found that the *SF3B1* K700E mutation results in thousands of shifts to alternative BSs by the mutant U2 snRNP. These widespread changes in BSs occur independently of changes in 3′ ss choice. They reflect the shift of U2 to more strongly interacting sites, either through enhanced base pairing with U2 snRNA or through a more U-rich polypyrimidine tract that may exhibit stronger U2AF2 binding.

Discussion

We previously isolated U2 particles from the chromatin fraction of 293Flp-In cells and showed that they copurified with protected intron BSs across the transcriptome. We now show that U2 particles can also be isolated from K562 leukemia cells carrying either *SF3B1* WT or the K700E allele found in MDS and other cancers. U2 complexes with WT or mutant SF3B1 were stably bound to intron BSs across the transcribed RNA and contained the same set of core U2 proteins, as well as the RBM5 and RBM10 proteins. The strong protection of intron BPs by the U2 particles indicates formation of the U2/BS helix, while the lack of branch formation indicates that the particles are likely derived from A- or B- complex spliceosomes. Characterizing the RNA fragments protected from nuclease in these particles allowed us to compare U2/BS interactions in WT and mutant cells.

The distribution and number of reads recovered from a particular BS are consistent across samples (34), allowing detection of changes in BS preference between cell populations. This was validated by the identification of differentially engaged BSs for a large fraction of the alternative 3′ splice sites identified as activated by SF3B1 K700E (36). It is notable that many of these alternative 3′ splice sites were associated with multiple upstream BPs. In some cases, a new site not seen in WT was strongly preferred in mutant cells (Fig. 2). In other cases, most or all the BSs were observed in both WT and mutant cells, but their distribution changed. Our finding of new BPs for 3′ splice sites activated by K700E indicates that changes in spliceosome

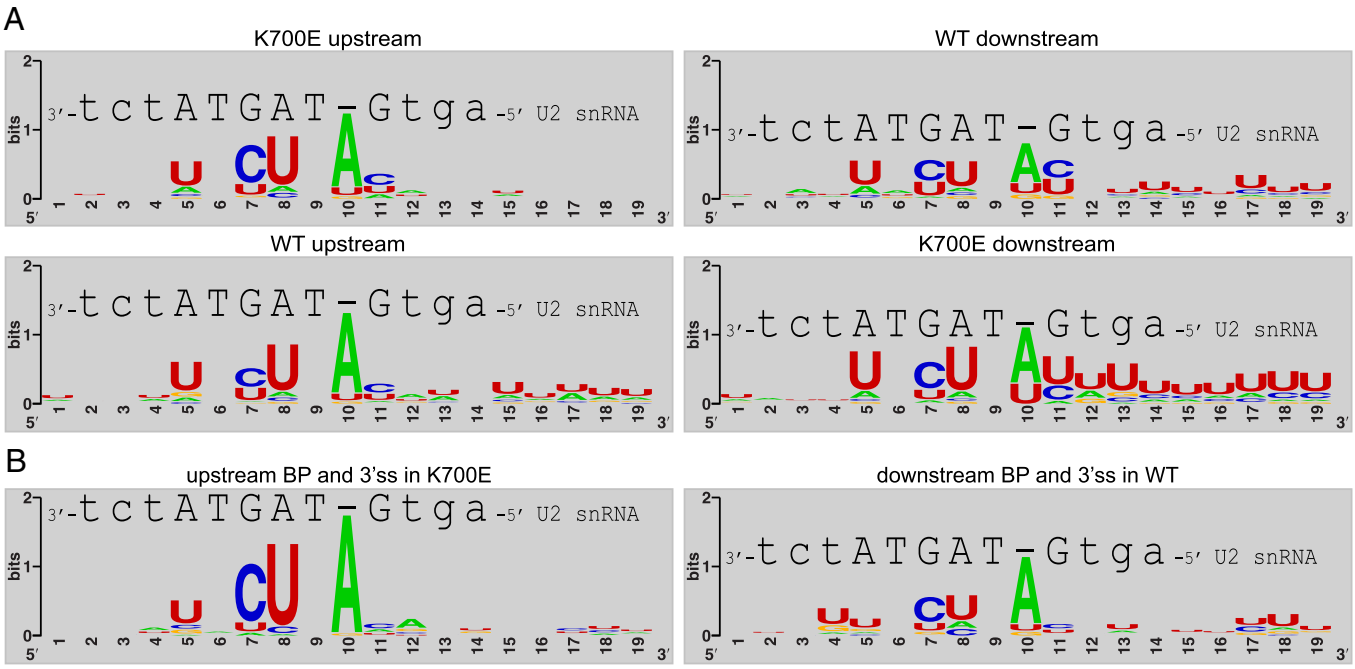

Fig. 5. Upstream branch sites preferred in K700E mutant cells exhibited stronger matches to the consensus than the downstream WT site, whereas mutant-preferred downstream sites are adjacent to more U-rich polypyrimidine tracts. (*A*) Consensus sequences derived from differentially used branch sites separated by 7 to 14 nt and their base-pairing with U2 snRNA in the canonical mode 1 are shown as in *SI Appendix*, Fig. S10*A*. (*B*) Consensus sequences for mode 1 branch sites, associated with alternative 3′ splice sites. Only sites from introns where the upstream 3′ss is selected in mutant cells are shown.

assembly have already occurred prior to the first catalytic step of splicing, although changes in later stages may also occur.

Previous studies demonstrated that multiple *SF3B1* mutations, including K700E, destabilize SF3B1 interaction with the protein SUGP1 and reduce its ability to activate the helicase DHX15 (8, 14, 20). SUGP1 has not been observed in early-stage spliceosomes so far characterized and likely acts transiently prior to BS recognition. These data are consistent with a model in which SUGP1 mediates interactions between U2, SF1, and/or U2AF2 at the BS/3′ SS (24). By this model, the G patch of SUGP1 then activates DHX15 ATPase activity to displace SF1 from the pre-mRNA, allowing U2 to associate with the BS (24). In the absence of SUGP1, DHX15 is not activated, SF1 is not effectively removed from the BS, and U2 may shift to a cryptic BS if available. Notably, the U2 particles described here have already undergone BP assembly and lack SUGP1 (34). The observed binding of U2 to alternative BSs in the SF3B1 mutant cells indicates that assembly onto these sites is stable after the proposed early SUGP1 step and persists into mature spliceosomes.

An alternative model for how loss of SUGP1 leads to cryptic BS/3′ ss usage posits a quality control function whereby SUGP1 and DHX15 induce prespliceosome disassembly when an erroneous BS is bound (21, 23). This model is consistent with the known role of DHX15, and its yeast homologue Prp43, in driving disassembly of the intron lariat complex in the final step of the splicing cycle (41–43). Our data indicate that such an activity in the early spliceosome would involve WT U2 snRNP recognizing thousands of alternative BSs, often in close juxtaposition to authentic BSs, and then inducing SUGP1 recruitment and DHX15 activation to selectively disassemble the alternative BS complexes. Although neither model is ruled out, we favor the idea that SUGP1/DHX15 facilitates recognition of the correct BS rather than preventing use of incorrect BSs. Our finding that alternative BSs used in the mutant cells have an overall better match to the consensus than the WT sites seems consistent with a positive role for SUGP1/DHX15 in facilitating recognition of authentic, sometimes weaker, BSs. It is not clear in either model what features of a BS might allow SUGP1 to determine its use, or how loss of GPATCH8 can compensate for loss of SUGP1 on some sites. We need to learn more about the steps of U2 snRNP assembly where these proteins act to fully understand their role in BP selection.

A remarkable finding from our data is that the mutant spliceosome is binding to thousands of alternative BSs and not just those associated with alternative 3′ splice sites. It is likely that reducing the stringency of our search criteria would identify additional BSs that are used at lower frequency. Note that of the ~124,000 mapped intronic BS clusters, only ~4,000 showed prominent changes in BP choice (~3.2%). Although there are extensive changes in BPs in the mutant cells, the majority of splicing across the transcriptome is occurring normally.

It is well established that the same 3′ ss can be chosen in the second splicing step after branching to multiple BPs. Early in vitro studies indicated that BP choices can be altered by regulatory factors and can affect the rate of splicing (44–46). The use of BSs additional in MDS mutant cells that are not associated with changes in splicing has implications for the disease phenotype. For example, it is not understood whether alternate BPs might differ in performance during the later steps of spliceosome assembly and catalysis. BP changes could have complex effects on mRNA levels by altering the kinetics of splicing for certain introns, or the overall efficiency of product formation. Thus, the altered BS selection resulting from mutation of SF3B1 may affect cellular mRNA metabolism and gene function in more subtle ways than the identified changes in mRNA isoforms.

The BSs preferred in K700E mutant cells over WT showed several common features. Mutant-preferred sites upstream of the WT BS on average exhibited stronger base pairing potential with U2 snRNA (Fig. 4). In contrast, mutant-preferred BSs downstream of WT sites were not markedly stronger in their U2 base-pairing potential than the WT sites but were instead adjacent to polypyrimidine tracts with greater enrichment for U residues. A previous study of SF3B1 mutations at R625 and

K666 (7), which also disrupt interaction with SUGP1, made similar observations that potential BSs associated with activated upstream 3′ splice sites were stronger than the downstream WT BSs (8). Notably, a study of the K700E mutation in chronic lymphocytic leukemia cells found that 3′ splice sites and their associated BSs activated in the mutant cells tended to be weaker than the WT 3′ss (12). This difference from our findings may point to interesting mechanistic differences and factor dependencies for early spliceosome assembly between myeloid and lymphoid cells. It is also possible that the different bioinformatic tools and approaches used in the two studies could lead to the different observations.

Our findings fit with a model in which K700E mutant SF3B1, lacking interaction with SUGP1, requires additional stabilization for U2 to engage a BS (24). This stabilization may come from either stronger pairing with U2 snRNA or a more U-rich and more closely adjacent polypyrimidine tract, with higher U2AF2 occupancy (10, 47). A closer match to the BS consensus could also act to enhance SF1 binding (48). The BSs preferred in K700E and WT cells were most frequently in close proximity and partially overlapping (Fig. 3). It may be that U2 snRNA can shift its pairing along the pre-mRNA to a more favorable interaction close by, or that different sites of SF1 binding can be favored prior to engagement of the BS by the incoming U2. We did not find evidence for mutant SF3B1 preferring particular modes of U2 snRNA base pairing. Canonical pairing [Mode 1; (30)] was the most common in both WT and mutant cells, and there were no clear preferences for a new pairing mode when the binding shifted. Examples of BSs using noncanonical pairing modes of U2 pairing were more limited, and it is difficult to compare the relative stabilities of these interactions. Analyses of larger datasets combined with better assessment of the binding affinities for different RNA pairings may elucidate additional features of SF3B1 function in BS selection.

The K562 cells carrying the K700E mutation provide a valuable means for determining how an MDS mutation affects both splicing and BS recognition. However the disease phenotype is manifested in hematopoietic stem and progenitor cells, and it may be that different BP alterations occur in these cells. It should thus be informative to perform U2 IP-seq in MDS patient tissue samples or in cells derived from patient iPSCs. Our study demonstrates the feasibility of such analyses in its sensitive detection of new BPs in cells carrying a heterozygous *SF3B1* allele, and in using a U2 antibody to assay endogenous protein without an epitope tag. It will also be interesting to assay changes in U2 assembly resulting from other mutations in SF3B1 and in cells of other tumor types carrying splicing factor mutations.

In summary, we have identified transcriptome-wide alterations in branch site recognition caused by a common cancer-associated mutation in the U2 snRNP protein SF3B1. Our findings give insight into both the mechanisms controlling spliceosome assembly and the role of splicing misregulation in neoplastic disease.

Methods

Cell Lines. K562 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FBS in a 5% CO2 incubator (10). SF3B1-K700E mutant cells were inoculated at twice the density of the WT cells and grown for 72 h.

Antibodies. Primary antibodies used for immunoblot assays: SUGP1 (Thermo Fisher Scientific, A304-675A). SF3A1, SF3B1, and U1-70 K are described in ref. 49, RBM5 and RBM10 are described in ref. 34.

SF3A2 antibody: rabbits were immunized with the human SF3A2 C-terminal peptide C _ MLRPPLPSEGPGNIP, conjugated to Keyhole Limpet Hemocyanin. Following three boosts, IgG clones cross-reacting with the SF3A2 peptide were obtained by single B cell sequencing (GenScript). These were coupled to Protein G-agarose (Pierce), crosslinked using DMP, and further screened for specific immunoprecipitation and efficient elution of U2 snRNP with excess MLRPPLPSEGPGNIP (SF3A2-CT) peptide.

RNP Sample Preparation. Cells were harvested and washed in phosphate-buffered saline. Nucleoplasm and HMW material were extracted as described (34). Extracts were incubated overnight at 4 °C with 5 to 7.5 μL packed-volume of either M2 FLAG agarose beads (Millipore-Sigma), anti-SF3A2 Protein G-agarose beads, or Protein G-agarose crosslinked in absence of IgG. Beads were then washed four times with wash buffer (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, and 0.05% Triton-X100), and eluted for 2 h at 4 °C in the presence of 150 ng/μL of either 3xFLAG (Millipore-Sigma) or SF3A2-CT peptide. Proteins analyses were performed by staining with SYPRO Ruby (Thermo Fisher Scientific) and immunoblotting. Aliquots of RNA obtained following deproteinization were subjected to 5′ radiolabeling and separated by Urea-PAGE or were sequenced using Illumina technology as described (34). Raw data is deposited in the Gene Expression Omnibus, National Center for Biotechnology Information (NCBI GEO) (50).

IPseq and RNAseq Data Analysis. Removal of PCR duplicates, genome alignment, analysis of clustered reads, and branchpoint prediction were performed as described (34). Branchpoint use within a defined 3′ splice site cluster was expressed as the ratio of the number of reads supporting a particular branchpoint to the total number of reads in the cluster. Branchpoint preference was calculated as the ratio of branchpoint use in WT vs K700E mutant samples. All sequences shown were aligned to the human genome hg19.

Alternative splicing was analyzed by rMATS (37) and expressed as changes in percent-spliced-in values (ΔPSI). 3′ splice site events showing a splicing change between control K700E and WT cells were considered regulatory targets (|ΔPSI|>10 with FDR less than 0.05). Differentially used branchpoints at these alternative 3′ splice sites were selected manually. Only one pair of branchpoints was chosen, even though multiple WT or K700E branchpoints were observed at some of these events.

Differentially used branchpoints not necessarily associated with alternative 3′ splice sites were selected from 3′ splice site and intronic IPseq clusters. K700E rank1 branchpoints were preferred at least two-fold compared to WT and were selected along with the rank1 branchpoint used in WT. Branchpoint pairs supported by fewer than 25 SF3A2 IPseq K700E and WT reads combined and not ranked in SF3B1 IPseq data were excluded.

Branch site base pairing with the U2 sequence GUAGUA in each of the six described modes (30) was assessed by awarding four points per GC pair, three points per AU pair, and two points per GU pair. Mismatches were not issued a penalty. The pairing mode with highest score, calculated as the sum of points, was selected as the most likely for each branch site.

Data, Materials, and Software Availability. Sequence data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) (GSE280659) (50).

ACKNOWLEDGMENTS. We thank Luke Buerer and William Fairbrother (Brown University, RI) for sharing code and helpful discussion. This work was supported by awards from the Jonsson Comprehensive Cancer Center at University of California, Los Angeles to D.L.B. and by NIH Grants R01GM127473 (A.D.), R21HG012624 (D.L.B.), R35GM136426 (D.L.B.), and R35GM118136 (J.L.M.). Sequencing data generated in this study have been deposited at NCBI GEO under the accession number GEO: GSE280659 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE280659).

1. K. Yoshida *et al.*, Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
2. R. K. Bradley, O. Anczuków, RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* **23**, 135–155 (2023).
3. Q. Zhang, Y. Ai, O. Abdel-Wahab, Molecular impact of mutations in RNA splicing factors in cancer. *Mol. Cell* **84**, 3667–3680 (2024).
4. S.C.-W. Lee *et al.*, Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. *Cancer cell* **34**, 225 (2018).

5. L. Malcovati *et al.*, SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood* **126**, 233–241 (2015).
6. V. Quesada *et al.*, Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet* **44**, 47–52 (2012).
7. S. Alsafadi *et al.*, Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
8. S. Huber *et al.*, SF3B1 mutated MDS: Blast count, genetic co-abnormalities and their impact on classification and prognosis. *Leukemia* **36**, 2894–2902 (2022).
9. R. B. Darman *et al.*, Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
10. J. Zhang *et al.*, Disease-causing mutations in SF3B1 alter splicing by disrupting interaction with SUGP1. *Mol. Cell* **76**, 82–95.e7 (2019).
11. C. DeBoever *et al.*, Transcriptome sequencing reveals potential mechanism of cryptic 3′ splice site selection in SF3B1-mutated cancers. *PLOS Comput. Biol.* **11**, e1004105 (2015).
12. I. López-Oreja *et al.*, SF3B1 mutation–mediated sensitization to H3B-8800 splicing inhibitor in chronic lymphocytic leukemia. *Life Sci. Alliance* **6**, e202301955 (2023).
13. M. E. Wilkinson, C. Charenton, K. Nagai, RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
14. C. van der Feltz, A. A. Hoskins, Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Crit Rev. Biochem. Mol. Biol.* **54**, 443–465 (2019).
15. Z. Zhang *et al.*, Structural insights into how Prp5 proofreads the pre-mRNA branch site. *Nature* **596**, 296–300 (2021).
16. Z. Zhang *et al.*, Structural insights into the cross-exon to cross-intron spliceosome switch. *Nature* **630**, 1012–1019 (2024).
17. O. Gozani, J. Potashkin, R. Reed, A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell Biol.* **18**, 4752–4760 (1998).
18. Z. Liu *et al.*, Pan-cancer analysis identifies mutations in SUGP1 that recapitulate mutant SF3B1 splicing dysregulation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10305–10312 (2020).
19. S. Alsafadi *et al.*, Genetic alterations of SUGP1 mimic mutant-SF3B1 splice pattern in lung adenocarcinoma and other cancers. *Oncogene* **40**, 85–96 (2021).
20. J. Zhang *et al.*, DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2216712119 (2022).
21. I. Beusch *et al.*, Targeted high-throughput mutagenesis of the human spliceosome reveals its in vivo operating principles. *Mol. Cell* **83**, 2578–2594.e9 (2023).
22. H. M. Maul-Newby *et al.*, A model for DHX15 mediated disassembly of A-complex spliceosomes. *RNA* **28**, 583–595 (2022).
23. Q. Feng, K. Krick, J. Chu, C. B. Burge, Splicing quality control mediated by DHX15 and its G-patch activator SUGP1. *Cell Rep.* **42** (2023).
24. J. Zhang *et al.*, Characterization of the SF3B1–SUGP1 interface reveals how numerous cancer mutations cause mRNA missplicing. *Genes Dev.* **37**, 968–983 (2023).
25. S. Benbarche *et al.*, GPATCH8 modulates mutant SF3B1 mis-splicing and pathogenicity in hematologic malignancies. *Mol. Cell* **84**, 1886–1903.e10 (2024).
26. Q. Tang *et al.*, SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. *Genes Dev.* **30**, 2710–2723 (2016).
27. B. Zhao *et al.*, Cancer-associated mutations in SF3B1 disrupt the interaction between SF3B1 and DDX42. *J. Biochem.* **172**, 117–126 (2022).
28. F. Yang *et al.*, Mechanisms of the RNA helicases DDX42 and DDX46 in human U2 snRNP assembly. *Nat. Commun.* **14**, 897 (2023).
29. T. R. Mercer *et al.*, Genome-wide discovery of human splicing branchpoints. *Genome Res.* **25**, 290–303 (2015).
30. A. J. Taggart *et al.*, Large-scale analysis of branchpoint usage across species and cell lines. *Genome Res.* **27**, 639–649 (2017).
31. B. Signal, B. S. Gloss, M. E. Dinger, T. R. Mercer, Machine learning annotation of human branchpoints. *Bioinformatics* **34**, 920–927 (2018).
32. P. Zhang *et al.*, Genome-wide detection of human variants that disrupt intronic branchpoints. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2211194119 (2022).
33. Y. Zeng *et al.*, Profiling lariat intermediates reveals genetic determinants of early and late co-transcriptional splicing. *Mol. Cell* **82**, 4681–4699.e8 (2022).
34. A. Damianov *et al.*, The splicing regulators RBM5 and RBM10 are subunits of the U2 snRNP engaged with intron branch sites on chromatin. *Mol. Cell* **84**, 1496–1511.e7 (2024).
35. Z. Zhang *et al.*, Molecular architecture of the human 17S U2 snRNP. *Nature* **583**, 310–313 (2020).
36. Y. K. Lieu *et al.*, SF3B1 mutant-induced missplicing of MAP3K7 causes anemia in myelodysplastic syndromes. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111703119 (2022).
37. S. Shen *et al.*, rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5593–E5601 (2014).
38. C. W. J. Smith, E. B. Porro, J. G. Patton, B. Nadal-Ginard, Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature* **342**, 243–247 (1989).
39. C. W. J. Smith, T. T. Chu, B. Nadal-Ginard, Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns. *Mol. Cell Biol.* **13**, 4939–4952 (1993).
40. D. Inoue *et al.*, Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019).
41. J.-B. Fourmann *et al.*, The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP-intron interaction. *eLife* **5**, e15564 (2016).
42. R. Yoshimoto, N. Kataoka, K. Okawa, M. Ohno, Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Res.* **37**, 891–902 (2009).
43. R. Toroney, K. H. Nielsen, J. P. Staley, Termination of pre-mRNA splicing requires that the ATPase and RNA unwindase Prp43p acts on the catalytic snRNA U6. *Genes Dev.* **33**, 1555–1574 (2019).
44. J. C. S. Noble, Z.-Q. Pan, C. Prives, J. L. Manley, Splicing of SV40 early pre-mRNA to large T and small t mRNAs utilizes different patterns of lariat branch sites. *Cell* **50**, 227–236 (1987).
45. R. A. Padgett *et al.*, Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8349–8353 (1985).
46. B. Ruskin, J. M. Greene, M. R. Green, Cryptic branch point activation allows accurate in vitro splicing of human β-globin intron mutants. *Cell* **41**, 833–844 (1985).
47. E. Glasser *et al.*, Pre-mRNA splicing factor U2AF2 recognizes distinct conformations of nucleotide variants at the center of the pre-mRNA splice site signal. *Nucleic Acids Res.* **50**, 5299–5312 (2022).
48. M. Corioni, N. Antih, G. Tanackovic, M. Zavolan, A. Krämer, Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res.* **39**, 1868–1879 (2011).
49. S. Sharma, S. P. Wongpalee, A. Vashisht, J. A. Wohlschlegel, D. L. Black, Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* **28**, 2518–2531 (2014).
50. A. Damianov *et al.*, U2 IP-seq from K562 cells expressing SF3B1-3xFlag WT or SF3B1-3xFlag K700E. NCBI GEO. https://doi.org/10.22033/ESGF/CMIP6.7673. Deposited 30 October 2024.