The Genetics of Splicing in Neuroblastoma

Justin Chen1,2,3, Christopher S. Hackett2,3, Shile Zhang4,5, Young K. Song5, Robert J.A. Bell1,6, Annette M. Molinaro2,3,6,7, David A. Quigley6,8, Allan Balmain9, Jun S. Song9,10, Joseph F. Costello6, W. Clay Gustafson11, Terry Van Dyke12, Pui-Yan Kwok13,14,15, Javed Khan8, and William A. Weiss2,3,11
ABSTRACT

Regulation of mRNA splicing, a critical and tightly regulated cellular function, underlies the majority of proteomic diversity and is frequently disrupted in disease. Using an integrative genomics approach, we combined both genomic data and exon-level transcriptome data in two somatic tissues (cerebella and peripheral ganglia) from a transgenic mouse model of neuroblastoma, a tumor that arises from the peripheral neural crest. Here, we describe splicing quantitative trait loci associated with differential splicing across the genome that we use to identify genes with previously unknown functions within the splicing pathway and to define de novo intronic splicing motifs that influence splicing from hundreds of bases away. Our results show that these splicing motifs represent sites for functional recurrent mutations and highlight novel candidate genes in human cancers, including childhood neuroblastoma.

SIGNIFICANCE: Somatic mutations with predictable downstream effects are largely relegated to coding regions, which comprise less than 2% of the human genome. Using an unbiased in vivo analysis of a mouse model of neuroblastoma, we have identified intronic splicing motifs that translate into sites for recurrent somatic mutations in human cancers. Cancer Discov; 5(4); 380-95. ©2015 AACR.

INTRODUCTION

Alternative splicing of mRNA is a highly conserved process that is subject to both genetic regulation and heritability (1). Some of these regulatory systems act in cis within the primary sequence of the pre-mRNA transcript, whereas others act in trans via genetically distant factors recruited to the splice site (2). Alternative splicing may be particularly important to cancer, as the unique cancer environment selects for novel splice isoforms that promote tumor growth, metastasis, or response to treatment (3). In addition, recurrent somatic mutations in known splicing factors, including U2AF1 (4) and SF3B1 (5), implicate functional contributions of this pathway in cancer and have led to interest in these factors as targets for cancer therapy (6).

Neuroblastoma is the most common cancer of infancy and the most common extracranial solid tumor of childhood. Extensive whole-genome and whole-exome studies, including sequencing analyses of over 300 tumors, have identified point mutations in genetic drivers of neuroblastoma (e.g., MYCN, ALK, PHOX2B, ARID1A, ARID1B, and ATRX) in only a minority of patients (7). Genome-wide association studies of high-risk neuroblastoma, however, have identified a robust signal at the BARD1 locus where the risk allele is associated with functional effects of BARD1 splicing (8). Coupled with the identification of differences in splicing between stage I and stage IV disease in patients with neuroblastoma (9), alternative splicing has the potential to be a major contributor to this disease.

We used an integrative genomics approach to survey alternative splicing in neuroblastoma, combining both genome and transcriptome data into a single analysis. Linkage mapping, by identifying associations between genotypes and phenotypes in a genetically controlled cohort, can identify genomic regions with functional importance. This type of approach, when adapted to high-throughput technologies and used to query expression quantitative trait loci (eQTL), represents a powerful tool to discover genetic mechanisms governing gene expression (10). We applied an extension of this concept, a splicing quantitative trait locus (sQTL) analysis (11–16), in a defined backcrossed mouse system using a genetically engineered model of neuroblastoma (17). By comparing two somatic neural tissues, our sQTL analysis uncovered a complex genome-wide splicing landscape, including the identification of novel trans-acting splicing-related genes. Coupled with available whole-genome sequencing (WGS) data of the parental strains, we also identified novel candidate intronic splicing motifs. We found that these motifs serve as sites for recurrent somatic mutations in human cancer that lead to functional changes in alternative splicing. We also identify a strain-specific triplet splicing preference within FUBP1 that leads to upregulation of MYC, with functional consequences in human neuroblastoma.
RESULTS
sQTL Distribute throughout the Genome

FVB/NJ TH-MYCN transgenic mice were backcrossed to wild-type 129/SVj mice, and the N1 generation (n = 102) was profiled on Affymetrix Exon Arrays and genotyped at 349 SNP and microsatellite markers. We identified 1,664 and 1,751 sQTL (defined here as a paired alternative splicing event associated with a marker, as markers may have multiple associations; see Methods) in cerebellum and superior cervical ganglia (SCG), representing brain- and peripheral neural crest–derived tissues, respectively (Fig. 1A and B; 5% false detection rate). The low density of our genotyping panel reflects the controlled genetic heterogeneity of our backcrossed cohort and was not intended to identify causative polymorphisms. Instead, the resulting genetic map allowed us to distinguish splicing events with local cis effects from those with distal trans effects. The majority of sQTL was within 50 Mb of the spliced transcript, and thus defined to be cis (90.3% in cerebellum and 92.5% in SCG; Supplementary Tables S1 and S2). Of these cis-sQTL, 874 reflected similar splicing events in both cerebellum and SCG, whereas only five trans-sQTL (mapping at least 50 Mb away from the spliced transcript or to a different chromosome) overlapped between tissues (Fig. 1C and Supplementary Table S3). The polymorphisms of both of these types of sQTL associate with strain-specific alternative splicing events, suggesting different possible functional roles. Trans-sQTL could represent polymorphisms within genes that influence the splicing pathway, which can then affect the splicing of other transcripts derived from a distal genomic region. Conversely, cis-sQTL represent polymorphisms at or near the gene being spliced that can directly alter canonical splice sites or create novel splicing signals. These cis-sQTL polymorphisms could also reside in functional regions where they can affect the ability of intronic splicing enhancers (ISE), intronic splicing silencers (ISS), and their exonic counterparts, ESES/ESSs, to recruit accessory splicing factors to mediate alternative splicing.

Trans-sQTL Identify Genes That Influence Splicing

Trans-sQTL have become powerful tools to identify transcription factors (18) and other functional elements (19), as these function in trans to regulate the transcription or translation of other genes. We therefore looked at putative trans-sQTL to identify novel regulators that could affect splicing of distant genes. If master regulators of splicing were subject to genetic control, we would expect to find a high degree of colocalization, whereby the majority of trans-sQTL (with alternatively spliced exons spread throughout the genome) would associate with just a few genomic loci. Instead, we found that trans-sQTL were scattered throughout the genome (Fig. 2A), with little overlap at any individual genomic locus. We observed a maximum of eight colocalizing trans-sQTL and only a few “hotspots,” or individual loci linked to several splicing events (≥4; Fig. 2B). We focused on two of these hotspots to simplify our efforts to validate genes that influence splicing. Specifically, we focused on the five trans-sQTL linked to the marker rs29347557 on chromosome 10 in SCG and the eight trans-sQTL linked to the marker rs33477935 on the X chromosome in cerebellum.

Candidate genes in these regions were identified by examining differential transcript expression between SCG and cerebellum (Fig. 2B). Given the overwhelming tissue-specific nature of the trans-sQTL, we hypothesized their origin to be from tissue-specific regulation of putative splicing-related genes in these regions. The minimal overlapping 95% confidence interval for sQTL mapping to rs29347557 in SCG encompassed the range between markers rs13480474 and rs38621064 on chromosome 10. This region spans over 13 Mb and contains 66 known genes, of which 30 were differentially expressed between cerebellum and SCG (Supplementary Table S4). Almost all of these genes had known functions. Among these, Sfh5 encodes a splicing factor subunit (20) and was the only gene known to function within the splicing pathway, supporting the idea that differentially expressed splicing machinery would reside in these loci (Fig. 2B). The sQTL mapping to rs33477935 in cerebellum possessed 95% confidence intervals that minimally overlapped from rs33478059 to rs13483805 on the X chromosome. This region spans over 77 Mb and contains 489 known genes, 123 of which were differentially expressed (Supplementary Table S5). Two of these genes, Hnmpa3 and Rbm10, encode known splicing factors (21, 22). Others, such as Rbmx, are also implicated in the splicing pathway. We identified putative candidate genes Rbmx2, which contains a RNA-binding motif, and Ddx26b, which encodes a protein containing a DEAD-box domain. Both of these genes are relatively uncharacterized (Fig. 2B).

To examine the role these genes could play within the splicing pathway and to extend these results into a human setting, we used siRNA to knockdown DDX26B and RBMX2 in HEK293T cells (Supplementary Fig. S1A) and looked for splicing differences in the orthologous human genes associated with the colocalized trans-sQTL, by reverse-transcription PCR (RT-PCR). We looked specifically for alternative products by eliminating amplification of the canonical isoform. Cloning and sequencing of RT-PCR products from five of these eight targets did not support any changes in splicing patterns (i.e., Cons2, Cdc123, Phd13, Nlufs2, and SapS2; data not shown), whereas CS (orthologous to AI182371) and FBXW12 (orthologous to Fbxw15) indicated changes in splicing following knockdown (Supplementary Table S4 and Supplementary Fig. S1B; we were unable to investigate the final target due to the lack of annotation). Although it appeared that knockdown of DDX26B and RBMX2 both affected the RT-PCR profile of CS, we examined all potential alternative products and discarded those arising from nonspecific priming. Specifically in the presence of siRNA directed toward RBMX2, we identified a novel splice isoform of CS lacking six internal exons (Supplementary Fig. S1C). Similarly, a previously undescribed splice isoform of FBXW12 lacking an internal exon was expressed only when DDX26B was knocked down (Supplementary Fig. S1D). These isoforms were undetected in the presence of nontargeting control siRNA. These data support functional roles for DDX26B and RBMX2 within the splicing pathway. We did not identify a statistically significant difference in Ddx26b or Rbmx2 expression between the parental mouse strains (Supplementary Fig. S1E), suggesting that the
Figure 1. Genomic distribution of sQTL. sQTL are distributed throughout the genome similarly in cerebellum (CB; A) and SCG (B). The x-axis indicates the location of each SNP, and the y-axis indicates the level of significance of an associated sQTL. The horizontal red line is drawn to mark an estimation of genome-wide significance at an FDR < 0.05. C, cis-sQTL were more abundant than trans-sQTL, and a majority were shared between both tissues. In contrast, trans-sQTL were largely tissue specific.
**Figure 2.** Trans-sQTL in the cerebellum (CB) and SCG. A, Circos plot of 162 cerebellum and 131 SCG trans-sQTL identified at an FDR < 0.05. Mouse chromosomes are plotted on the outer ring, with locations of SNPs with sQTL indicated in black on the first inner ring. Locations of spliced exons are identified in blue on the second inner ring. Each trans-sQTL (paired association between SNP and alternatively spliced exon) is depicted by a line that links the exon and the SNP in SCG (purple), and cerebellum (green), with sQTL shared between both tissues drawn in black. The histogram outside of the chromosomes indicates the number of sQTL (on a scale of 1–8) that colocalize at any given genomic locus. B, Circos plot of sQTL hotspots, where four or more colocalized. An additional data track between the chromosome ideograms and SNP loci indicates differentially expressed genes at the transcript level (purple = higher expression in SCG; green = higher expression in cerebellum). Five SCG-specific trans-sQTL mapped to rs29347557 on chromosome 10. This region included differential expression of Sf3b5, encoding a splicing factor subunit. Eight cerebellum-specific trans-sQTL mapped to rs33477935 on the X chromosome. Shown are the locations of the candidate genes that are differentially expressed at this locus.
difference in their abilities to affect splicing does not lie at the transcriptional level.

**Cis-sQTL Identify Strain-Specific Isoforms**

In both cerebellum and SCG, a cis-sQTL identified strain-specific isoform expression of fifth exon within the Astn2 (Astr2) gene. This was associated with a logarithm of odds (LOD) score of 65.1 at rs13477756 (Supplementary Fig. S2A and S2B), the highest observed in cerebellum. Astr2 may function in neuronal migration, and exonic deletions have been identified in patients with schizophrenia (23). RT-PCR on cDNA derived from cerebellum in both parental strains confirmed that 129/SvJ mice specifically expressed a transcript resulting from exon skipping of the fifth exon (Supplementary Fig. S2C and S2D).

We examined WGS information for parental FVB/NJ (24) and 129/SvJ (25) at the Astr2 locus, which revealed a lack of coverage throughout exon 5 specifically in 129/SvJ, leading us to hypothesize that this alternative isoform could result from structural variation between the two strains. Indeed, an approximately 30-kb deletion present in 129/SvJ, but not in FVB/NJ, colocalized with Astr2 and encompassed the entirety of exon 5 (Supplementary Fig. S2E).

We identified Neuroblastoma (data not shown; n = 134). E, Western blot analysis of human neuroblastoma cell lines transfected with GFP control, FUBP1<sup>97S</sup>, or FUBP1<sup>97-</sup>-lentivirus tagged with V5. MYC is upregulated in both SHEP and SK-N-AS lines in the presence of high FUBP1<sup>97S</sup> levels.

129/SvJ WGS, we further identified 1,324 somatic copy number variations (SCNV) in 129/SvJ mice and 1,802 SCNVs in FVB/NJ compared with the reference genome (Supplementary Fig. S3A and S3B). Fifty-five of these colocalized with cis-sQTL–identified alternative splicing events in 129/SvJ and 58 colocalized in FVB/NJ, with 38 shared between strains (Supplementary Table S6). Expression of the alternative isoform in 129/SvJ is thus a result of genomic structural variation closely linked to the rs13477756 marker. Despite this alternative mechanism, this sensitive detection of strain-specific exon usage in Astr2 by a SNP marker residing over 2 Mb away from the gene itself is a clear proof-of-principle for our sQTL method.

The highest LOD score observed in SCG (LOD = 73.9) belonged to the FUSE-binding protein 1 (Fubp1) gene (Fig. 3A). FUBP1 is a transcription factor that binds the Far Upstream Element (FUSE), thereby regulating transcription of Myc (26, 27). Given the identification of cancer-associated somatic mutations in FUBP1 (28) and the known importance of MYC proteins in neuroblastoma (29), we further explored Fubp1. Although array data indicated a strain-specific splicing difference at the fifth exon (Fig. 3B), complete skipping of this exon was not detected by RT-PCR (Supplementary Fig. S4). We therefore explored other models through which...
Splicing could modulate FUBP1 activity. Exon 5 within FUBP1 is a known site of triplet NAGNAG splicing (30), which occurs at the terminal end of 3 splice sites. This repetition includes the highly conserved AG dinucleotide that defines the intron–exon boundary, resulting in alternative 3 splice sites that differ by three bases and are referred to by their respective intron lengths. For FUBP1, this splicing event, which is conserved between mouse and human, incorporates a serine at position 97 in the proximal isoform (shorter intron, FUBP1<sup>97−</sup>) that is absent in the distal isoform (longer intron, FUBP1<sup>97S</sup>−; Fig. 3C). Expression analyses of these isoforms across several tissue types suggest that this splicing event is not a stochastic process, but rather subject to an as-yet-undefined regulatory mechanism.

Gene-level expression datasets supported tumor-suppressive effects of FUBP1 in human neuroblastoma, as higher expression was correlated with better survival in all stages (P = 1.48 × 10<sup>−4</sup>; FDR = 2.70 × 10<sup>−3</sup>; Supplementary Fig. SSA; ref. 31) and event-free survival in a MYCN nonamplified cohort (P = 7.48 × 10<sup>−6</sup>; FDR = 6.51 × 10<sup>−6</sup>; Supplementary Fig. SSB; ref. 32). Because these data do not have sufficient resolution to evaluate triplet splicing of exon 5, we next analyzed RNA-seq data to quantify FUBP1<sup>97−</sup>:FUBP1<sup>97S</sup> ratios in human neuroblastoma. Given the inverse relationship between MYC and MYCN (33), the poor survival of neuroblastoma patients expressing MYC (34), and the known interaction between FUBP1 and MYC, we looked specifically at MYCN-nonamplified patients. Median event-free survival was 669 days in patients who expressed higher FUBP1<sup>97−</sup>:FUBP1<sup>97S</sup> ratios as compared with 966 days in those with low ratios (P = 0.0486; Fig. 3D). This indicated that higher FUBP1<sup>97−</sup> expression relative to FUBP1<sup>97S</sup> was correlated with poor event-free survival in MYCN-nonamplified disease, consistent with recent evidence that triplet splicing is subject to regulation and has functional consequences (30, 35). We next examined isoform-specific effects in human MYCN-nonamplified neuroblastoma cell lines. We transduced lentivirus with V5-tagged constructs of FUBP1<sup>97−</sup> or FUBP1<sup>97S</sup> cDNA into SHEP and SKN-AS cell lines. Overexpression of FUBP1<sup>97−</sup>, but not FUBP1<sup>97S</sup>, led to higher MYC levels (Fig. 3E), consistent with deregulation of triplet splicing within exon 5 of FUBP1 having a functional effect.

**Splicing Motifs Serve as Sites of Recurrent Mutations in Cancer**

Our genotyping marker set was not designed to identify functional polymorphisms. Rather, markers were selected to map chromosomal regions to either parental strain. As cis-sQTL represent differences in splicing driven by local genetic polymorphisms, we compared WGS information for FVB/NJ (24) and 129/SvJ (25) surrounding alternatively spliced exons possessing cis-sQTL in both tissues, identifying recurrent intronic decamer motifs that spanned SNPs. Enrichment analysis of these sequences identified 22 unique motifs (E value < 0.05; Supplementary Fig. S6). Seven were similar to previously described motifs, with six known to bind splicing factors (Supplementary Fig. S7; FDR < 0.05; ref. 36). Thus, our method identified motifs associated with regulation of alternative splicing.

Collectively, these motifs matched 0.07% of all possible decamer sequence permutations. Despite this low rate, our position-agnostic approach identified an abundance of splicing motifs in both mouse and human introns (94.7% of murine introns and 95.2% of human introns at least one of these splicing motifs), leading us to hypothesize that these splicing motifs could be functionally conserved in humans regardless of their relative intronic positions. It has also been estimated that as much as 87% of genes have conserved splicing patterns between mouse and human orthologs (37). To see whether these splicing motifs could represent functional sites commonly mutated in human cancer, we queried WGS of 40 neuroblastoma tumors with matched normal sequence (38). Three candidate genes, Dynlrb1, Cgb7, and Moxd2p, were identified as having an enrichment of somatic mutations compared with germline variation in intronic splicing motifs, each occurring in two samples (5% of the cohort; Fig. 4A–C, respectively). For Dynlrb1 and the chr7:14194354 splicing motif mutation in Moxd2p, these mutations resulted in a stronger match to a splicing motif by replacing a base found infrequently (or not at all) with a base more commonly associated with the motif. In the case of Cgb7 and the chr7:141944631 splicing motif mutation in Moxd2p, the motif is effectively destroyed by substituting in a base that is not normally found.

The low incidence of recurrent genes with splicing motif mutations was unsurprising, given the low mutation rate in neuroblastoma (7). We similarly analyzed WGS from 42 high-risk glioblastoma (GBM) tumors, each also with a paired normal sample (39, 40), and found a higher rate of recurrence. We identified three candidate genes, LOC100505811, Tram2-As1, and Npipa1, with enrichment of somatic mutations compared with germline variants in intronic splicing motifs in at least three samples (7.1%–11.1% of the cohort; Supplementary Fig. S8A–S8C, respectively). Eleven additional genes were recurrent in at least two samples (4.7% of the cohort; Table 1).

**Splicing Motif Mutations Lead to Functional Changes in Alternative Splicing**

With RNA-seq data available for 25 of the 42 samples, we analyzed The Cancer Genome Atlas (TCGA) GBM dataset further. Each of the recurrent mutations identified in GBM occurred at highly conserved positions within the splicing motif and either substituted a nucleotide that was not normally observed at that position (Loc100505811 and Npipa1), or replaced a nucleotide with the conserved nucleotide (Tram2-As1), creating a match to the splicing motif. To see whether these mutations changed the splicing of these genes, we profiled normalized exon expression for the exons immediately proximal to the introns with splicing motif mutations. Although the trend in differences of normalized exon expression between the two exons in Loc100505811 was not statistically significant (Supplementary Fig. S8A), the splicing motif mutations in Tram2-As1 and Npipa1 were associated with a significant decrease in normalized exon expression for the exons downstream of the intron (reduction in means by 12.87%, P = 0.02248 and 27.62%, P = 0.03037; Supplementary Fig. S8B and S8C, respectively). Thus, these mutations are located in sites correlated with changes in alternative splicing.

We continued to examine the neuroblastoma splicing motif mutations in Moxd2p and Cgb7, along with the Npipa1...
The Genetics of Splicing in Neuroblastoma

RESEARCH ARTICLE

Figure 4. Recurrent somatic mutations occur in intronic splicing motifs in neuroblastoma. Analysis of 40 neuroblastoma samples reveals enrichment of somatic mutations in splicing motifs in DYNLRB1 (A), CGB7 (B), and MOXD2P (C). Germline variants (GV) and somatic mutations (SM) were identified by MuTect. MAST was used to examine the sequence immediately surrounding both intronic GV and intronic SMs for a match to any of the 22 splicing motifs (sequence-level match, \( P < 0.0001 \)). The sequence logo of the splicing motif is drawn linked to the physical genomic location of the somatic mutation (black line). The total height of each nucleotide position is the information content in bits and represents the level of conservation for that position. The height of each nucleotide letter represents the ratio that they are found to occupy that position. The gene structures of known isoforms are depicted in red with arrows indicating the direction of transcription. The reference sequence (black) is given directly beneath the sequence logo with the position and nucleotide of the mutant allele shown in red. Red silhouettes indicate the number of tumor samples with that particular mutation. A black silhouette indicates the allele was found in a normal sample.

A, two neuroblastoma samples (5%) had a recurrent C>T mutation within DYNLRB1 at chr20:33,121,706. This mutation created a stronger match to a splicing motif by substituting a base found infrequently to a base more commonly associated with the motif.

B, two tumor samples had a recurrent G>C mutation at chr19:49,558,080 within CGB7; another neuroblastoma germline sample was also found to be G/C heterozygous at the same position. The splicing motif matched here was the same as observed in the DYNLRB1 mutation, and when analyzed in the orientation of the corresponding transcript, this mutation effectively destroyed the match to the motif.

C, two neuroblastoma samples were identified with distinct splicing motif mutations at chr7:141,943,454 (A>C) and chr7:141,944,631 (C>T) in MOXD2P. When analyzed in the orientation of the corresponding transcript, the mutation at chr7:141,943,454 destroyed a splicing motif match, whereas the chr7:141,944,631 mutation resulted in creating a match to the splicing motif.

(Chr16:15040359) and TRAM2-AS1 splicing motif mutations identified in GBM. We cloned these introns containing the mutations into the pGINT splicing reporter where the introns separated the eGFP coding sequence into two exons, and we used quantitative RT-PCR to empirically assess canonical splicing efficiencies in SHEP human neuroblastoma cells after a 16-hour transfection. We compared introns containing the wild-type allele, the splicing motif mutation, and a control mutation that converted the AG dinucleotide at the 3’ splice site to a GG dinucleotide. We hypothesized that introduction of the mutant allele in the splicing motif would alter the ability of the cells to efficiently splice the eGFP exons, with the control mutation eliminating canonical splicing altogether.

The MOXD2P splicing motif mutations at chr7:141943454 (A>C) and chr7:141944631 (C>T) are located 47 bp and 149 bp from the nearest intron–exon boundary, respectively.
These mutant alleles led to 3- and 2-fold increases in splicing efficiency, respectively, when compared with the wild-type allele (P = 0.0317 and P = 0.0015; Fig. 5A and B). The CGB7 splicing motif mutation at chr19:49558080 (G>C), however, did not lead to a statistically significant change in splicing efficiency, despite residing only 18 bp away from the nearest intron–exon boundary (Fig. 5C).

The G>A mutation at chr16:15040359 within NPIPA1 resides 520 bp away from the nearest intron–exon boundary in an intron approximately 3 kb in total length. This mutation led to a 14% increase of canonical splicing efficiency, respectively, when compared with the wild-type allele (P = 0.0478; Fig. 5D). Similarly, the T>G mutation at chr6:52445058 within TRAM2-AS1 is located 650 bp away from the nearest intron–exon boundary in a 3.3-kb intron. The mutant allele led to a decrease in splicing efficiency, resulting in a 56% of the splicing efficiency of wild-type allele (P = 0.0375; Fig. 5E). All splicing reporters with a splice-site mutation had undetectable eGFP expression. These results confirm that splicing motifs, even when located beyond immediate exon–intron boundaries, can function to influence alternative splicing. It should be noted that enhancement or reduction in splicing efficiency was not associated directionally with either gain-of-function or loss-of-function mutations; this can be attributed to the creation of a match to a splicing motif that could increase the affinity for either splicing silencers or splicing enhancers. Similarly, a mutation that destroys a match to a splicing motif could decrease the affinity for either silencers or enhancers, leading to unpredictable splicing effects.

Recurrence Splicing Motif Mutations Identify Novel Genes of Interest in Neuroblastoma

Given the genome-wide somatic mutation rates in these WGS cohorts (2.685/Mb and 6.082/Mb for neuroblastoma and GBM, respectively), the nucleotide-level recurrence of splicing motif mutations suggests a high degree of selection, whereby these altered splicing profiles contribute to the development of disease. Three genes (NPIPA1, DYNLRB1, and AIMPI) had enriched splicing motif mutations in both neuroblastoma and GBM (Table 1), but evidence establishing their involvement in either disease is lacking, highlighting the need for further investigation.

Although the function of nuclear pore complex interacting protein family–member A1 (NPIPA1) is unknown, both DYNLRB1 and AIMPI encode proteins implicated in the pathogenesis of other cancers. Dynemin light chain roadblock-type I (DYNLRB1) interacts with TGFβ signaling and is mutated in human ovarian, colorectal, and gastric cancers (41). Interestingly, in ovarian cancer, novel splice isoforms have been reported (42). In contrast, the function of nuclear pore complex interacting protein family–member A1 (NPIPA1) is unknown, both DYNLRB1 and AIMPI encode proteins implicated in the pathogenesis of other cancers.

Table 1. Recurrent genes with enriched splicing motif mutations

<table>
<thead>
<tr>
<th>Gene name</th>
<th>GBM splicing motif GVs</th>
<th>GBM splicing motif SMs</th>
<th>Neuroblastoma splicing motif GVs</th>
<th>Neuroblastoma splicing motif SMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOXD2P</td>
<td>0</td>
<td>0</td>
<td>1-chr7:141944631 C&gt;T</td>
<td>1-chr7:141943454 A&gt;C</td>
</tr>
<tr>
<td>CGB7</td>
<td>0</td>
<td>0</td>
<td>1-chr19:49558080 G/C</td>
<td>2-chr19:49558080 G&gt;C</td>
</tr>
<tr>
<td>DYNLRB1</td>
<td>1-chr20:33104296 C/G</td>
<td>2-chr20:33121706 C&gt;T</td>
<td>2-chr20:33121706 C&gt;T</td>
<td>1-chr:16:15040359 G&gt;A</td>
</tr>
<tr>
<td>NPIPA1</td>
<td>1-chr16:15040359 G/A</td>
<td>4-chr16:15040359 G&gt;A</td>
<td>1-chr16:15031765 G&gt;C</td>
<td>1-chr16:15040359 G&gt;A</td>
</tr>
<tr>
<td>LOC10050581</td>
<td>0</td>
<td>3 chr5:117618623 C&gt;T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TRAM2-AS1</td>
<td>0</td>
<td>3 chr6:52445058 T&gt;G</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AIMPI-1</td>
<td>0</td>
<td>2-chr1:107260426 G&gt;T</td>
<td>1 chr:107260426 G&gt;T</td>
<td>0</td>
</tr>
<tr>
<td>PRDM12</td>
<td>0</td>
<td>2-chr9:13354135 G&gt;A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TIGD7</td>
<td>0</td>
<td>2-chr16:3353982 A&gt;G</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FBXO45</td>
<td>0</td>
<td>2-chr3:196310237 C&gt;T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SLC4A1</td>
<td>0</td>
<td>2-chr17:42342896 C&gt;A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARHGAP29</td>
<td>1-chr1:97667742 G/C</td>
<td>2-chr1:94678897 G&gt;A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LOC100128006</td>
<td>1-chr17:12681207 G/A</td>
<td>1-chr17:12679656 T&gt;G</td>
<td>1-chr17:12673734 C&gt;T</td>
<td>0</td>
</tr>
<tr>
<td>LOC440896</td>
<td>1-chr9:69180408 G/A</td>
<td>1-chr9:69179798 T&gt;C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STXBPS</td>
<td>1-chr6:14746822 G/C</td>
<td>2-chr6:147693778 T&gt;C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TNII2</td>
<td>1-chr11:1860346 T/G</td>
<td>1-chr11:1860346 T&gt;G</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE: Enrichment was identified by taking the ratio of samples with somatic mutations (SM) within splicing motifs to samples with germline variants (GV) within splicing motifs.
Figure 5. Recurrent somatic mutations in intronic splicing motifs affect splicing efficiency. A, the mutant splicing motif allele within MOXD2P at chr7:141,943,454 led to a 3-fold increase in splicing efficiency (213.8%) compared with the wild-type allele (P = 0.0317). B, the mutant splicing motif allele within MOXD2P at chr7:141,944,631 led to an increase in splicing efficiency (93.9%) compared with the wild-type allele (P = 0.0015). C, no changes were detected between the mutant and wild-type splicing motif allele in the CGB7 intron. D, the mutant splicing motif allele within NPIPA1 at chr16:15,040,958 led to a 14% increase in splicing efficiency compared with the wild-type allele (P = 0.0478). E, the mutant splicing motif allele at chr6:52,444,408 within TRAM2-AS1 resulted in a decrease to 56% splicing efficiency compared with the wild-type allele (P = 0.0375). Splicing efficiency was measured as the relative expression of the canonical eGFP transcript compared with the expression of plasmid Neomycin resistance marker. Shown are representative means from at least three independent experiments. Error bars represent SEM. N.D., not detected.

TGFβ signaling (42). Functional studies also implicate a role for DYNLRB1 in cell migration. Aminoacyl tRNA synthetase complex-interacting multifunctional protein 1 (AIMP1) is an intracellular protein that is subject to proteolytic cleavage when present at the cell surface. Upon cleavage, the released 22-kDa peptide induces apoptosis, antagonizes angiogenesis, and can stimulate an inflammatory response in the tumor microenvironment. In contrast, at high intracellular concentrations, AIMP1 induces cell migration (43).

Although we can examine the genomic effects of these mutations with regard to splicing motifs, we cannot categorically define these as gain-of-function or loss-of-function at the gene or protein level. As the results from our splicing reporter assays suggest, both the creation and destruction of splicing motifs could facilitate and impair splicing, depending on the binding of splicing enhancers or splicing silencers. The introns containing the splicing motif mutations in DYNLRB1 and AIMP1 spanned over 8 kb and 10 kb, respectively, making them unamenable to in vitro splicing reporter assays. In addition, any specific effects on splicing are not necessarily predictive of downstream gene function, as alternatively spliced variants could have no function or are immediately degraded by nonsense-mediated decay (possibly leading to no effect or haploinsufficiency), act as a dominant negative, or even show constitutive activity.

To investigate the relevance of these genes in neuroblastoma, we examined gene expression and survival across two publicly available neuroblastoma datasets containing either 87 untreated primary neuroblastoma tumors across all stages (31) or 102 untreated MYCN-nonamplified primary neuroblastoma tumors (32). Across both datasets, higher DYNLRB1 expression correlated with better overall survival and...
event-free survival ($P = 8.40 \times 10^{-3}$, FDR = 9.97 $\times 10^{-3}$, and $P = 4.49 \times 10^{-3}$, FDR = 6.83 $\times 10^{-4}$, respectively; Fig. 6A and B), suggesting a tumor-suppressive effect, whereas higher AIMP1 expression correlated with worse overall survival and event-free survival ($P = 5.62 \times 10^{-4}$, FDR = 0.0157, and $P = 3.16 \times 10^{-4}$, FDR = 2.49 $\times 10^{-3}$, respectively; Fig. 6C and D), indicating an oncogenic effect. Higher NPIPA1 expression was associated with better event-free survival in neuroblastoma as well, statistically significant only in the MYCN-nonamplified dataset ($P = 2.10 \times 10^{-3}$, FDR = 3.49 $\times 10^{-3}$, Fig. 6E; $P = $ N.S., Fig. 6F). These data suggest that splicing motif mutations undergo selection in genes with clinical relevance and can be useful to identify novel targets in disease.

**DISCUSSION**

In this study, we leveraged a mouse genetic system to decode and map genetic regulation of splicing. We identified over 2,500 putative sQTL, consistent with alternative splicing being subject to a greater degree of genetic control (16) than indicated by the earliest studies (<400 sQTL; refs. 11–14). Furthermore, by using the TH-MYCN neuroblastoma mouse model, we were able to examine splicing in somatic tissue during the course of neuroblastoma development and oncogenic overexpression. Although several splicing factors are known direct targets of MYC, we have observed no changes in transgene expression between the parental strains, suggesting genetic
differences between strains are the primary drivers of the sQTL. As the conservation of splicing between mouse and human has been estimated to be up to 87% (37), we looked to see whether our results could have relevance in a human setting.

With the growing appreciation that alternative splicing can play functional roles in cancer, splicing factors have garnered considerable interest as potential targets for therapy (6). In this regard, our analysis enabled us to focus on only a few genes in each region. Although this analysis was limited by relatively large 95% confidence intervals stemming from the backcrossed nature of our cohort, we were nevertheless able to use this approach to identify critical loci harboring genes previously not known to affect splicing. We did not find a high degree of trans-sQTL colocalization, sites that might be indicative of splicing master regulators. This could reflect a rarity of such trans-sQTL overall, or might alternatively be attributed to a limitation within our system, as polymorphisms in such a critical component of the splicing process might not be tolerated in either strain. Our analysis reveals an abundance of trans-acting genes that influence splicing and act in a limited gene-to-gene scope in a highly tissue-specific fashion. Although the genes we have identified could act as bona fide splicing factors and bind directly to RNA or the spliceosome, they could also act indirectly by affecting the expression or activity of other splicing factors. Nevertheless, these data suggest a greater complexity to the splicing pathway than what has already been defined by mass spectrometry experiments focused on the spliceosome (44). It should be noted that our trans-sQTL analysis is not comprehensive; by specifically examining genes within the confidence interval that are differentially expressed between tissues, we were not able to identify genes with activities that are modified by posttranslational modifications and microRNAs that could regulate the expression of other splicing factors.

FUBP1 has been reported to be a tumor suppressor (28) and an oncprotein (45) with the ability to both activate and repress MYC transcription (46). The strain-specific splicing we observe suggests distinct functions for alternatively spliced FUBP1 isoforms. Of the two triplet splicing products, FUBP177 led to an increase of MYC, whereas FUBP1975 reduced MYC levels. Our RNA-seq-based survival analysis (Fig. 3D) represented an interim analysis of the TARGET program and is currently under-powered to examine the relationship between isoform expression and disease stage. These data raise the possibility that FUBP1 mutations observed in cancers, such as oligodendrogloma, are not strictly loss-of-function, but may act in a dominant-negative fashion akin to FUBP1975. Despite the relationship between MYC and FUBP1 and the critical importance of MYC proteins in neuroblastoma tumorigenesis (29, 34), extensive genomic sequence analysis in neuroblastoma tumors has yet to show recurrent mutations in FUBP1 or MYC. In the absence of recurrent mutation, our analysis reveals splicing differences in FUBP1, not detectable by conventional expression analysis, which associate with risk.

Although lacking in whole-exome sequencing analyses, intronic mutations in key regulatory motifs are potentially just as disruptive as mutations in coding regions. Our analysis identified novel splicing motifs that reside along the length of the intron. Previous splicing analyses have focused on exonic splicing motifs, using methods such as analyzing evolutionary conservation or systematic evolution of ligands by exponential enrichment (SELEX) on known and purified splicing factors. Until now, the most definitive analysis of ISEs used a splicing reporter construct to test the effects of random decamers inserted into the intron (47). Our unbiased analysis is the first of which we are aware to identify recurrent intronic splicing motifs in vivo.

We identified and established relevance for these splicing motifs by analyzing WGS from two different human tumor types, identifying novel candidate genes with functional implications in the process. The TCGA WGS cohort provided us with additional insights into the effects of these mutations, as a portion of the samples were accompanied by RNA-seq data. Despite this, not all samples had both transcriptomic and genomic data available, limiting our analysis of normalized exon expression for the NPIPA1 splicing motif mutation to just three tumors. Nevertheless, four out of five in vitro splicing reporters containing these introns with splicing motif mutations confirm functional significance, showing that even single-nucleotide changes hundreds of bases away from the nearest intron–exon boundary could affect mRNA splicing. These splicing motifs may function to bind accessory splicing factors to promote exon skipping, enhance exon inclusion, or even drive truncated transcripts through intron-retention or alternative polyadenylation. They may also alter splicing by changing the transcript’s secondary structure; without highly accurate RNA folding predictions, this analysis remains challenging and uncertain. It is also possible that these splicing motifs reside in noncoding RNA, which could in turn regulate expression of splicing factors.

The recurrent somatic mutations we identified in neuroblastoma and GBM might be easily dismissed because of their intronic nature, but our analysis extends understanding of the cancer genome by identifying functional ramifications for these intronic sites. The somatic mutations we identify within splicing motifs represent just one avenue whereby cancers may benefit from the deregulation of these gene products. Our analysis also broadens the applicability of these data beyond the limited individuals who possess these specific mutations by identifying novel candidate tumor suppressors and oncogenes in neuroblastoma.

**Methods**

**Generation of a Heterogeneous Cohort of Backcrossed Mice**

Mice were obtained from The Jackson Laboratory and were housed and treated following the guidelines of the University of California, San Francisco (UCSF; San Francisco, CA) Institutional Animal Care and Use Committee (IACUC). FVB/NJ TH-MYCN transgenic mice were bred to wild-type 129/SvJ mice. Transgenic animals from the resulting F1 generation were identical genetically, possessing an N1 mutation to just three tumors. Nevertheless, four out of five in vitro splicing reporters containing these introns with splicing motif mutations confirm functional significance, showing that even single-nucleotide changes hundreds of bases away from the nearest intron–exon boundary could affect mRNA splicing. These splicing motifs may function to bind accessory splicing factors to promote exon skipping, enhance exon inclusion, or even drive truncated transcripts through intron-retention or alternative polyadenylation. They may also alter splicing by changing the transcript’s secondary structure; without highly accurate RNA folding predictions, this analysis remains challenging and uncertain. It is also possible that these splicing motifs reside in noncoding RNA, which could in turn regulate expression of splicing factors.

The recurrent somatic mutations we identified in neuroblastoma and GBM might be easily dismissed because of their intronic nature, but our analysis extends understanding of the cancer genome by identifying functional ramifications for these intronic sites. The somatic mutations we identify within splicing motifs represent just one avenue whereby cancers may benefit from the deregulation of these gene products. Our analysis also broadens the applicability of these data beyond the limited individuals who possess these specific mutations by identifying novel candidate tumor suppressors and oncogenes in neuroblastoma.

**Methods**

**Generation of a Heterogeneous Cohort of Backcrossed Mice**

Mice were obtained from The Jackson Laboratory and were housed and treated following the guidelines of the University of California, San Francisco (UCSF; San Francisco, CA) Institutional Animal Care and Use Committee (IACUC). FVB/NJ TH-MYCN transgenic mice were bred to wild-type 129/SvJ mice. Transgenic animals from the resulting F1 generation were identical genetically, possessing an allele for each gene from both parental strains. F1 mice were then backcrossed to 129/SvJ mice to generate the N1 generation used in this analysis. SCG and cerebellum were surgically isolated and snap-frozen in liquid nitrogen (n = 102).

**Genotyping**

DNA was isolated from spleen tissue using a proteinase K lysis followed by phenol chloroform extraction. Microsatellite marker genotyping was carried out by the Marshfield Clinic (Marshfield, WI) and the
Center for Inherited Disease Research (Baltimore, MD). SNP genotyping was performed using template-directed primer extension with fluorescence polarization detection (FP-TDI, Aycloprime II; PerkinElmer) and SNPSstream 48-plex (Beckman Coulter). Markers and map positions are shown in Supplementary Table S7. The marker set had an average spacing of 8 Mb genome-wide (excluding the high density of markers on chromosome 10).

RNA Isolation

RNA was isolated from SCG using the RNeasy kit (Qiagen). All other RNA isolations used TRIzol (Invitrogen/Life Technologies) for phase separation before purification with the RNeasy kit.

Expression Arrays

One microgram of RNA was used as a starting template for RiboMinus rRNA subtraction (Invitrogen/Life Technologies) followed by the ST labeling protocol (Affymetrix). Labeled samples were hybridized to Affymetrix Mouse Exon 1.0 arrays. Arrays were normalized using RMA in the XPS Bioconductor package within R at both the exon level and the transcript level using core probes. Differential expression between tissue types at the transcript level was examined using significance analysis of microarrays (SAM; ref. 48).

sQTL Analysis

Exon expression was normalized to gene expression by calculating normalized exon expression, a ratio of exon expression to gene expression (both values determined using the XPS Bioconductor package). sQTL were calculated using normalized exon expression as a quantitative trait in sQTL software as previously described (49).

Splicing Motifs

Spliced exons were identified as possessing a cis-sQTL in both tissues, and the direction of splicing in each strain was noted. 19-mer sequences surrounding polymorphic nucleotides and their positions relative to the spliced exon were acquired in the direction of transcription from consensus FASTA sequences generated from strain-specific WGS. Sequences were trimmed to exclude exonic bases, and regions where either strain returned a poly(N) motif indicative of a lack of sequencing coverage were discarded. Extracted sequences were binned on the basis of the direction of the associated splicing and their origin (5′ intron or 3′ intron).

MEME (52) was used to identify recurrent motifs with a width of at least 10 bases in these bins using a background guanine-cytosine content of 42%. Motifs with an E value of <0.05 and derived from at least 20 sequences were reported as significant. Redundant motifs were identified as those that shared a Pearson correlation coefficient > 0.60, and the representative motif was chosen with the lowest E value. TOMTOM (52) was used to compare sets of identified motifs at an FDR < 0.05 using Pearson correlation coefficients and requiring a minimum overlap of five bases. Cumulative matching of splicing motifs was determined by generating a multi-FASTA file consisting of 1,048,576 possible decamer permutations. Seven hundred forty-three, or 0.07%, of these sequences matched the splicing motifs when compared with this file using MAST (52) at a sequence P value threshold of 0.0001. Introns containing splicing motif matches were calculated by extracting unique RefSeq introns from the University of California, Santa Cruz (UCSC) hg19 build (n = 195,956) and mm9 assemblies of FVB/NJ and 129/SvJ (n = 184,637) created in this study. Splicing motif matches were determined using FIMO (52) at a P value of 0.0001. The number of unique intron matches found in FVB/NJ (174,842) and 129/SvJ (174,746) was averaged when representing the percentage of murine introns containing at least a single splicing motif.

Analysis of Splicing Motif Mutations in Primary Neuroblastoma and GBM

Primary neuroblastoma and patient-matched normal DNA WGS data for 40 patients from the St. Jude Children’s Hospital—Washington University Pediatric Cancer Genome Project were downloaded from the European Genome-Phenome Archive (Dataset ID: EGAD00001000135). Primary GBM and patient-matched normal DNA WGS data for 42 patients were downloaded from the TCGA project’s Cancer Genomics Hub (www.cghub.ucsc.edu; dbGAP Study Accession number: phs000178; refs. 39, 40).

Single-nucleotide variants (SNV) were detected with MuTect, a Bayesian framework for the detection of somatic mutations (53). Somatic and germline SNVs were filtered according to MuTect defaults [germline variants were kept at a LOD(N) threshold of 2.3]. SNVs for TCGA GBM samples aligned to hg18 were converted to hg19 coordinates using the UCSC liftover tool (54). For technical reasons, one GBM sample had SNV calls only from chromosomes 1–7, and another GBM sample had SNV calls only from chromosomes 1–12. The overall somatic mutation rates were calculated by dividing the total number of somatic mutations by the total number of “cappable bases” identified by MuTect. The GBM samples with missing SNV calls were omitted from the calculation of somatic mutation rates.

Intronic sequences were extracted from the hg19 reference genome encompassing 9 bp upstream and downstream of the identified SNVs in the 5′–3′ direction of transcription. Reference sequences in addition to sequences containing the alternative allele were analyzed for splicing motif occurrences using MAST at a sequence P value threshold of 0.0001, recording the best motif match to either the reference or alternative sequence as determined by the lowest sequence P value.

GBM RNA-Seq Analysis

Open-access level 3 RNA-Seq data for 25 samples that overlapped with the 42 WGS GBM samples were downloaded through the TCGA Data Portal (55). In the absence of transcript expression levels, normalized exon expression was calculated by taking the ratio of individual exon expression (RPKM) to the maximally expressed exon within the gene. Significance was determined by the Student t test.

Neuroblastoma RNA-Seq and Survival Analysis

RNA-Seq was performed on samples collected for the TARGET program (currently unpublished) and represent an interim analysis of these data. The cohort consisted of all high-risk and stage IVS tumors (n = 20), and included stage III tumors (n = 2) that had similar tumor aggressiveness to stage IV.

Total RNA was extracted from the fresh-frozen tumor samples by the TRizol/RNeasy kit protocol described previously (56). Whole transcriptome libraries for Illumina HiSeq 2000 were prepared
according to the TruSeq RNA protocol, in which poly-A mRNA is purified from total RNA at the initial step. Two indexed samples were pooled and sequenced on an Illumina HiSeq 2000 with 100-bp paired end. Quality of RNA and library was assessed by an Agilent BioAnalyzer.

Hundred-base pair paired end reads were first aligned to the reference human genome (hg19) using spliced read mapper Tophat 2.0.8 (57). We then used Cufflinks 2.1.1 (58) for gene abundance estimation. Abundances in fragments per kilobase of exon per million fragments mapped (FPKM) were calculated for each annotated RefSeq gene. The Kaplan–Meier analysis was performed using GraphPad Prism. High and low distal:proximal isoform ratios were determined by a median split. Significance was determined by the log-rank test.

**Splicing Reporters**

The region of interest from NIPPIA1 was PCR-amplified using AccuPrime Taq High Fidelity (Invitrogen/Life Technologies) from genomic DNA obtained from HEK293T cells using gNIPPIA1, gTTRAM2-AS1, gMOXD2P, gCCB7 Cloning-F and Cloning-R primers (Integrated DNA Technologies; Supplementary Table S8) and TOPO-cloned into the PCR2.1-TOPO vector (Invitrogen/LifeTechnologies). The pGINT fluorescent splicing reporter was obtained from Addgene (Plasmid 24217) and used as a template for PCR amplification of the GFP exons using the GFPEx1-F/R and GFPEx2-F/R primers (Integrated DNA Technologies and Phusion High-Fidelity PCR Master Mix (Thermo Scientific) according to the manufacturer’s instructions. GFP fragments containing specific introns were generated via PCR using the GFPEx1-F, GFPEx2-R, and the appropriate Intron-F/R primer pair (Integrated DNA Technologies). The GFP fragment containing the second MOXD2P intron (containing the chr7:141944631 mutation) was generated through PCR amplification of the GFP exons using the GFPEx1-F/R and GFPEx2-F/R primers (Integrated DNA Technologies and Phusion High-Fidelity PCR Master Mix. The GFP inserts were cloned into the NotI and XhoI restriction sites in the pGINT plasmid, replacing the modified adenovirus intron, using the Quik Ligation Kit (New England Biolabs) according to the manufacturer’s instructions. Mutant alleles were obtained using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene/Agilent Technologies) and the appropriate SDM-F/R primers (Integrated DNA Technologies). Plasmids were verified using Sanger sequencing (Quintara Biosciences).

One microgram of each splicing reporter was transfected into SHEP cells with 2 μL of Lipofectamine 2000 (Invitrogen/Life Technologies) in individual wells of 24-well plates according to the manufacturer’s instructions. RNA was harvested after 16 hours using TRIzol (Invitrogen/Life Technologies) and the RNeasy Mini Kit (Qiagen). RNA (500 ng) was used as input for cDNA synthesis. Gene expression on five-times diluted cDNA was measured through quantitative RT-PCR (qRT-PCR) using 5’ nuclease assays with the GINT and NeoR qPCR primers/probes (Integrated DNA Technologies; Supplementary Table S8) and the KAPA PROBE FAST qPCR Kit (Kapa Biosciences). Reactions were run in triplicate on an ABI 7900HT Fast Real-Time PCR system (Applied Biosystems/Life Technologies) and analyzed using SDS 2.3 software. Splicing efficiency was calculated using the ΔΔCT method, comparing GFP expression with the expression of the Neomycin resistance gene included on the pGINT plasmid to control for plasmid copy number.

**Tissue Culture**

All cell lines were obtained from the UCSF Cell Culture Facility but not reauthenticated by the authors. SHEP cells were grown in RPMI 1640 supplemented with 10% FBS. SK-N-AS and HEK293T cells were grown in DMEM supplemented with 10% FBS, nonessential amino acids, and sodium pyruvate. All cell lines were maintained at 37°C with 5% CO2 and tested for Mycoplasma contamination by PCR.

**Lentiviral Transduction**

Human FUBP1 cDNA was obtained from the IMAGE consortium (Accession: BC017247) and PCR-cloned into the pENTR-D/TOPO gateway vector (Invitrogen/Life Technologies) using the FUBP1 Cloning-F and FUBP1 Cloning-R primers (Integrated DNA Technologies; Supplementary Table S8) to add the directional TOPO motif and exclude a stop codon. This entry clone was used as a substrate to create an FUBP1 entry vector using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene/Agilent Technologies) and Site-Directed Mutagenesis Primers FUBP1975 SDM-F/FUBP1975 SDM-R (Integrated DNA Technologies; Supplementary Table S8). Plasmid sequences were verified using Sanger sequencing (Quintara Biosciences). Lentiviral constructs were made by LR recombination (Invitrogen/Life Technologies) with the pLenti 6.3 destination vector (Invitrogen/Life Technologies). Virus was packaged in HEK293T cells using the ViraPower lentiviral packaging plasmid mix and Lipofectamine 2000 (Invitrogen/Life Technologies) over the course of 72 hours. Viral supernatant was harvested and filtered through a 0.45 micron syringe filter and used to directly transduce neuroblastoma cells for 24 hours. Stably transduced cells were selected with 10 μg/mL blasticidin (Invitrogen/Life Technologies).

**Neuroblastoma Expression Profiling and Survival Analysis**

Survival data were obtained from two datasets on the R2: microarray analysis and visualization platform (59). The Versteeg dataset (GEO Accession ID: GSE3705; ref. 31) spanned all neuroblastoma stages and included 88 samples profiled on Affymetrix U133P2 expression arrays. The Seeger dataset (GEO Accession ID: GSE3446; ref. 32) included 102 primary untreated neuroblastoma tumors without MYCN gene amplification profiled on Affymetrix HG-U133A and HG-U133B platforms. Expression data for both datasets were calculated using the MAS 5.0 algorithm. FUBP1, DYNLRB1, AIMPI1, and NIPPIA1 expression was assayed using the following probes: 214893_s_at, 217918_at, 202542_s_at, and 214870_s_at. High and low expression groups were separated by scanning along the cohort with a minimum size of 8. Significance was assessed by the log-rank test. The FDR was determined by the associated Q-value.

**Western Blotting**

Cell lysate was harvested using Cell Lysis Buffer (Cell Signaling Technology) supplemented with protease inhibitor (Roche) and 1% SDS. Protein was quantitated with the BCA assay kit (Pierce Biotechnology). Equal amounts of total protein were loaded and run on 4% to 12% SDS-polyacrylamide gels (Invitrogen/Life Technologies) and transferred to polyvinylidene difluoride (PVDF) membranes using the iBlot (Invitrogen/Life Technologies). After blocking (1 hour at room temperature, 5% nonfat milk in TBS-T), membranes were incubated overnight (4°C, 5% BSA in TBS-T) with V5-specific antibody (1:5,000; Invitrogen/Life Technologies), GAPDH-specific antibody (1:10,000; Millipore), or c-MYC-specific antibody (XR-P, 1:1,000; Cell Signaling Technology). Antibodies were detected with horseradish peroxidase (HRP)-linked mouse or rabbit (Calbiochem/Millipore) secondary antibodies followed by enhanced chemiluminescence (Amersham/GE).

**cDNA Synthesis and RT-PCR**

cDNA synthesis was performed using SuperScript VILO MasterMix (Invitrogen/Life Technologies) according to the manufacturer's instructions. One microgram of total RNA was used as starting material in the reverse transcriptase reaction. RT-PCR to determine retention of Fubp1 exon 5 and validate the alternative Astra2 isoform was performed using GoTag Green MasterMix (Promega) in 25-μL reactions according to the manufacturer’s instructions using 1 μL of cDNA and FUBP1-1F/FUBP1-7R or Astra2-F/R primers, respectively (Integrated DNA Technologies;
Supplementary Table S8). qPCR products were run on a 1.5% agarose gel and Sanger sequenced (Quintara Biosciences). Sequence analysis and chromatogram generation was performed with Geneious.

qRT-PCR was performed using 1 μl of a 1:10 dilution of cDNA using the KAPA SYBR FAST qPCR Kit (Kapa Biosystems) according to the manufacturer’s instructions and the appropriate primer sets (Integrated DNA Technologies; Supplementary Table S8). Reactions were run in triplicate on an ABI 7900HT Fast-Real-Time PCR system (Applied Biosystems/Life Technologies) and analyzed using SDS 2.3 software.

**Knockdown of Putative Trans-sQTL Splicing-Related Genes**

siRNA targeting human DDX26 and RBMX2 were purchased as ON-TARGETPlus SMARTpools (Dharmacon/Thermo Scientific Bio). ON-TARGETPlus Non-Targeting siRNA #1 (D-001810-01-05) was used as a negative control. siRNA (140 pmol) was transfected into HEK293T cells using Lipofectamine 2000 for a period of 72 hours, after which RNA was harvested using Trizol.

cDNA was synthesized and RT-PCR was performed using primers C5 RT-F/RT-R and FBXW12 RT-F/RT-R (Integrated DNA Technologies; Supplementary Table S8). The targeted amplicons for C5 and FBXW12 spanned exons 4-21 and exons 4-10, respectively. Products were run on a 1.5% agarose gel, and bands indicating the molecular weight of the full-length isoform were excised. The remaining gel products were extracted using the Gel-Extraction Kit (Qiagen) and reamplified using the same primers. The final products were analyzed on a 2% agarose gel. Differentially expressed products were gel-extracted and TOPO cloned into PCR2.1-TOPO vectors (Invitrogen).

**References**


