Supplemental Information For:
The genetics of splicing in neuroblastoma

Justin Chen, Christopher S. Hackett, Shile Zhang, Young K. Song, Robert J.A. Bell, Annette M. Molinaro, David A. Quigley, Allan Balmain, Jun S. Song, Joseph F. Costello, W. Clay Gustafson, Terry Van Dyke, Pui-Yan Kwok, Javed Khan, and William A. Weiss

SUPPLEMENTAL TABLES:

Supplemental Table 1: CB *cis*-sQTL

Supplemental Table 2: SCG *cis*-sQTL

Supplemental Table 3: *trans*-sQTL

Supplemental Table 4: Genes differentially expressed on Chr10 in 95% CI

Supplemental Table 5: Genes differentially expressed on ChrX in 95% CI

Supplemental Table 6: *Cis*-sQTL Co-localizing with SCNVs

Supplemental Table 7: Genotyping markers used in this study

Supplemental Table 8: Oligonucleotides used in this study

SUPPLEMENTAL FIGURE LEGENDS:

Supplemental Figure S1: *DDX26B* and *RBMX2* Influence Alternative Splicing

A. Quantitative-RT-PCR of HEK293T cells showing reduction of *DDX26B* and *RBMX2* transcript levels after 72 hour siRNA treatment. Expression levels are calculated using the ΔΔCT method, normalized to *GAPDH* expression and relative to the Non-Targeting siRNA control. Error bars indicate the standard deviation from a representative experiment.

B. RT-PCR of *C5* and *FBXW12* in HEK293T cells identify novel alternatively spliced isoforms after 72 hour siRNA treatment of *DDX26B* or *RBMX2*. Bands representing alternative isoforms are highlighted in red (*C5*) and orange (*FBXW12*) were subcloned and sequenced.

C. Visualization of BLAT alignment of the novel *C5* splice isoform. This isoform lacks 6 internal exons that are included in all known UCSC and RefSeq splice variants.

D. Visualization of BLAT alignment of the novel *FBXW12* splice isoform. This isoform skips an exon that is present in all known UCSC and RefSeq splice variants. This isoform also utilizes an alternative splice site.

E. *Ddx26b* and *Rbmx2* expression is not significantly different between the parental mouse strains.

Supplemental Figure S2: *Astn2* possesses a *cis*-sQTL

A. Normalized Exon Expression (NEE) levels for exon 5 of *Astn2* show loss of expression in the homozygotes (129/SvJ)

B. The sQTL for *Astn2* has a LOD score of 65.1 on chromosome 4 where the gene is located, indicating a *cis* effect.

C. RT-PCR of the parental 129/SvJ and FVB/NJ strains indicate that 129/SvJ expresses an alternative isoform.

D. Chromatogram from Sanger sequencing of the bands extracted in C indicating loss of Exon 5.

E. Exon 5 of *Astn2* resides in a SCNV that is lost in 129/SvJ. *Astn2* exons (orange) are overlayed on a plot depicting copy number at this locus in 129/SvJ (top) and FVB/NJ (bottom). The blue line indicates exon skipping of exon 5 that corresponds with copy number loss specifically in 129/SvJ. *Astn2* resides on the (-) strand.

Supplemental Figure S3: SCNVs Identified by WGS

A) SCNVs in 129/SvJ

B) SCNVs in FVB/NJ

Supplemental Figure S4: *Fubp1* RT-PCR From Parental Mouse Strains

RT-PCR of *Fubp1* in parental 129/SvJ and FVB/NJ strains did not indicate complete exon skipping of exon 5.

Supplemental Figure S5: *FUBP1* Expression Correlates with Survival in Two Neuroblastoma Datasets

Kaplan-Meier survival analysis indicates that patients with lower levels of total *FUBP1* expression (blue) have a reduced overall survival compared to patients with higher levels of total *FUBP1* (red) in a dataset containing patients spanning all stages (A, $p=1.48x10^{-4}$, $FDR=2.70x10^{-3}$) and a dataset comprised solely of MYCN non-amplified patients (B, $p=7.48x10^{-8}$, $FDR=6.51x10^{-6}$).

Supplemental Figure S6: Total Set of Identified Unique Splicing Motifs

Enrichment analysis of 19-mer sequences surrounding SNPs between FVB/NJ and 129/SvJ using MEME identified 22 unique motifs (E-Value < 0.05)

Supplemental Figure S7: Motifs with Matches to an RNA Binding database

Identified splicing motifs were compared to a database of known RNA binding motifs using TOMTOM (FDR < 0.05). Known protein binding relationships are indicated below the indicated motifs. Those in bold are known to have roles as splicing factors.

Supplemental Figure S8: Recurrent Somatic Mutations Occur in Intronic Splicing Motifs in GBM

Analysis of 42 high-risk GBM samples reveals recurrent somatic mutations in *LOC100505811* (A), *TRAM2-AS1* (B), and *NPIPA1* (C) that occur in separate splicing motifs. 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 physical location of the splicing motif mutation is indicated in black. 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. Boxplots indicate the normalized exon expression of the 5' exon and 3' exon flanking the intron with recurrent splicing mutations. Significance was assessed by Student's t-test. Normalized exon expression data are based on 25 overlapping samples with RNA-seq.

A. Three GBM samples (7.1%) had a recurrent C>T mutation at position 117,618,623 on chromosome 5 in a splicing motif within the intron of the uncharacterized lncRNA *LOC100505811.* The difference in normalized exon expression of either flanking exon was not statistically significant when comparing tumors with the mutation to those with the wild-type allele.

B. Three samples had a recurrent T>G mutation at chr6:52,445,058 in a splicing motif within an intron of the antisense RNA *TRAM2-AS1.* This mutation is associated with a decrease in normalized exon expression of the exon 3' to the intron ($p < 0.05$).

C. Five samples (11.9%) possessed intronic somatic splicing motif mutations within the Nuclear Pore Complex Interacting Protein Family, Member A1 gene, *NPIPA1.* One of these was a G>C mutation at chr16:15031765 and the remaining four were recurrent G>A mutations at chr16:15040359. A separate germline sample was also G/A heterozygous at the recurrent chr16:15040359 position. The chr16:1504359 mutation was associated with a decrease in normalized exon expression of the exon 3' to the intron ($p < 0.05$).