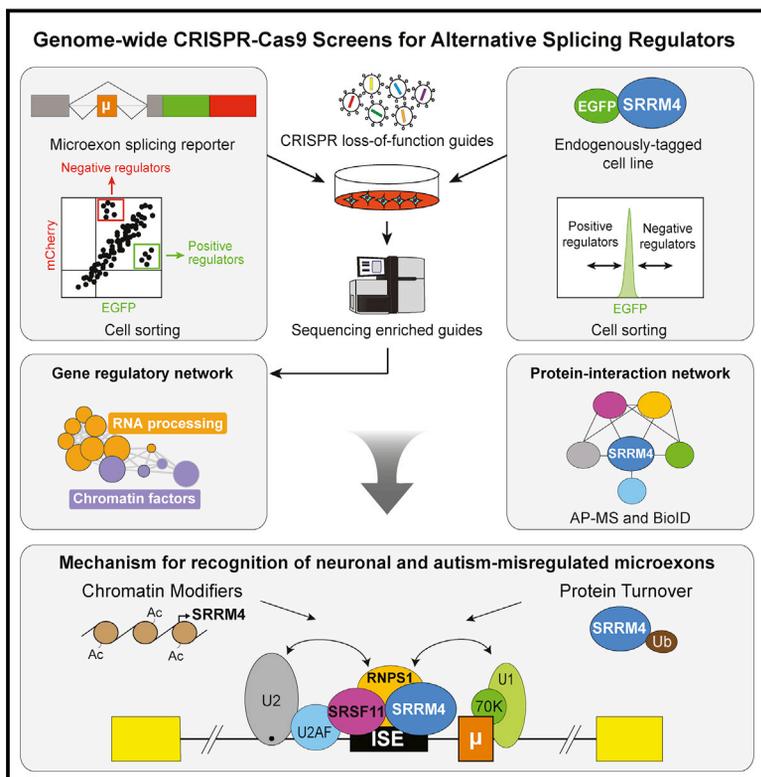


# Molecular Cell

## Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons

### Graphical Abstract



### Authors

Thomas Gonatopoulos-Pournatzis, Mingkun Wu, Ulrich Braunschweig, ..., Jason Moffat, Anne-Claude Gingras, Benjamin J. Blencowe

### Correspondence

t.gonatopoulos@gmail.com (T.G.-P.), b.blencowe@utoronto.ca (B.J.B.)

### In Brief

Gonatopoulos-Pournatzis et al. report a CRISPR-based strategy for the genome-wide identification of genes that control alternative splicing. Application of this technology to neuronal microexons frequently disrupted in autism reveals a mechanism for how very short exons, not amenable to previously described exon-definition interactions, are recognized and spliced in the brain.

### Highlights

- Genome-wide CRISPR-Cas9 screens for detection of alternative splicing regulators
- Identification of ~200 regulators of neuronal microexons often disrupted in autism
- Diverse regulators include chromatin, protein turnover and RNA processing factors
- A mechanism for the definition of neuronal microexons



# Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons

Thomas Gonatopoulos-Pournatzis,<sup>1,6,\*</sup> Mingkun Wu,<sup>1,2,6</sup> Ulrich Braunschweig,<sup>1,6</sup> Jonathan Roth,<sup>1,2,3</sup> Hong Han,<sup>1</sup> Andrew J. Best,<sup>1</sup> Bushra Raj,<sup>1,5</sup> Michael Aregger,<sup>1</sup> Dave O'Hanlon,<sup>1</sup> Jonathan D. Ellis,<sup>1</sup> John A. Calarco,<sup>2,4</sup> Jason Moffat,<sup>1,2</sup> Anne-Claude Gingras,<sup>2,3</sup> and Benjamin J. Blencowe<sup>1,2,7,\*</sup>

<sup>1</sup>Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada

<sup>2</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

<sup>3</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada

<sup>4</sup>Department of Cell and Systems Biology, University of Toronto, Toronto, ON M5S 3G5, Canada

<sup>5</sup>Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead Contact

\*Correspondence: [t.gonatopoulos@gmail.com](mailto:t.gonatopoulos@gmail.com) (T.G.-P.), [b.blencowe@utoronto.ca](mailto:b.blencowe@utoronto.ca) (B.J.B.)

<https://doi.org/10.1016/j.molcel.2018.10.008>

## SUMMARY

Alternative splicing is crucial for diverse cellular, developmental, and pathological processes. However, the full networks of factors that control individual splicing events are not known. Here, we describe a CRISPR-based strategy for the genome-wide elucidation of pathways that control splicing and apply it to microexons with important functions in nervous system development and that are commonly misregulated in autism. Approximately 200 genes associated with functionally diverse regulatory layers and enriched in genetic links to autism control neuronal microexons. Remarkably, the widely expressed RNA binding proteins Srsf11 and Rnps1 directly, preferentially, and frequently co-activate these microexons. These factors form critical interactions with the neuronal splicing regulator Srrm4 and a bi-partite intronic splicing enhancer element to promote spliceosome formation. Our study thus presents a versatile system for the identification of entire splicing regulatory pathways and further reveals a common mechanism for the definition of neuronal microexons that is disrupted in autism.

## INTRODUCTION

Transcriptome profiling across diverse cell and tissue types, developmental stages, and conditions has revealed “networks” of coordinated alternative splicing (AS) events that establish fundamental properties of biological systems (Braunschweig et al., 2013; Jangi and Sharp, 2014). These networks are enriched for evolutionarily conserved AS events in functionally coherent sets of genes, the perturbation of which often affects specific functions and phenotypes (Baralle and Giudice, 2017;

Scotti and Swanson, 2016). For example, networks of conserved and coordinated AS events are prevalent in the mammalian nervous system, where they play important roles in processes such as neurogenesis, axon guidance, synapse formation, and neurotransmission (Raj and Blencowe, 2015; Vuong et al., 2016).

Programs of AS required for nervous system development and function are regulated by RNA binding proteins (RBPs) that are differentially expressed between neural and other tissues, such as members of the NOVA, RBFOX, CELF, and PTBP families of proteins, as well as SRRM4 (also known as the neuronal-specific Ser-Arg [SR] repeat-related protein of 100 kDa [nSR100]; Raj and Blencowe, 2015; Vuong et al., 2016). These factors bind short, linear sequences in pre-mRNA referred to as exonic or intronic splicing enhancers and silencers to facilitate or repress the formation of spliceosomes at adjacent splice sites, respectively. Remarkably, disruption of these proteins and their individual splicing targets has been linked to defects in nervous system development and neurological disorders (Buckanovich et al., 1996; Gehman et al., 2011; Quesnel-Vallières et al., 2015).

SRRM4 activates a highly conserved program of activity-dependent neuronal microexons (i.e., 3- to 27-nt exons). This microexon regulatory network is disrupted in the brains of approximately one-third of analyzed autistic subjects (Irimia et al., 2014), and individual microexons within the network have been linked to important neurodevelopmental and brain functions (Parras et al., 2018; reviewed in Ustianenko et al., 2017 and in M. Quesnel-Vallières, R.J. Weatheritt, S.P. Cordes, and B.J.B., unpublished data). Moreover, mice haploinsufficient for Srrm4 recapitulate microexon misregulation and display hallmark autistic-like features, including altered social behaviors and synaptic transmission (Quesnel-Vallières et al., 2016). Disruption of the SRRM4-regulated neuronal microexon program has thus emerged as a convergent mechanism underlying autism spectrum disorder (ASD). However, how SRRM4-dependent microexons are recognized despite their short length, and how this mechanism is disrupted in autism, is poorly understood.



The discovery of the SRRM4-dependent microexon program highlights a fundamental question relating to AS regulatory networks, namely, which pathways and repertoires of factors control biologically important exons and introns? Such knowledge is necessary in order to fully understand the mechanisms and functions of these networks, as well as to identify possible targets for their therapeutic modulation (Scotti and Swanson, 2016). Previous studies employing global-scale small interfering RNA (siRNA) knockdowns of genes have provided insight into the sets of genes that control splicing events linked to apoptosis, proliferation, cell fate, and cancer (Han et al., 2017; Moore et al., 2010; Tejedor et al., 2015; Venables et al., 2008). However, the extent and functional range of the regulatory pathways impacting different classes of alternative exons, in particular those that are tissue, developmentally, and disease or disorder regulated, has not been determined. CRISPR-Cas technology offers a powerful and widely applicable genetic-based approach for the sensitive, genome-wide interrogation of gene function (Shalem et al., 2015; Wright et al., 2016) and is therefore well suited for addressing the fundamental questions concerning microexons and splicing regulatory networks described above.

In this study, using genome-wide CRISPR-Cas9 inactivation of protein-coding genes in cells expressing dual-fluorescent splicing reporters, we have systematically identified genes that regulate Srrm4-dependent neuronal microexons. An additional, complementary genome-wide CRISPR-Cas9 screen identified genes that impact endogenous Srrm4 expression. These screens reveal that microexon splicing is controlled by ~200 genes enriched in genetic links to neurological disorders and acting at multiple regulatory levels. An in-depth analysis of screen hits representing common regulators revealed that the SR-related proteins, Srsf11 and Rnps1, preferentially regulate neuronal microexons. These factors form mutually stabilizing interactions with Srrm4 and a specialized bi-partite intronic enhancer element that are important for the formation of early splicing complexes. Our results thus introduce a highly effective strategy for the comprehensive definition of splicing regulatory pathways and further shed light on how very short neuronal exons are recognized and spliced.

## RESULTS

### Generation of Neural Cell Lines Expressing Microexon Splicing Reporters

To systematically identify factors that impact neuronal microexon splicing, we constructed bichromatic splicing reporters (Kuroyanagi et al., 2006; Norris et al., 2014; Orenge et al., 2006) containing microexons, together with their upstream and downstream native intronic sequences, from the SH3 and multiple ankyrin repeat domains 2 (Shank2) and myocyte-specific enhancer factor 2d (Mef2d) genes (Figures 1A, S1A, and S1B). These microexons are of interest because they are highly conserved in mammals, activity dependent, misregulated in individuals with ASD, located in gene families that are genetically linked to ASD, and are regulated by SRRM4 and RBFOX1, which also have been associated with ASD (Ebert and Greenberg, 2013; Irimia et al., 2014; de la Torre-Ubieta et al., 2016; Quesnel-Vallières et al., 2016; Singh et al., 2014). Additional informa-

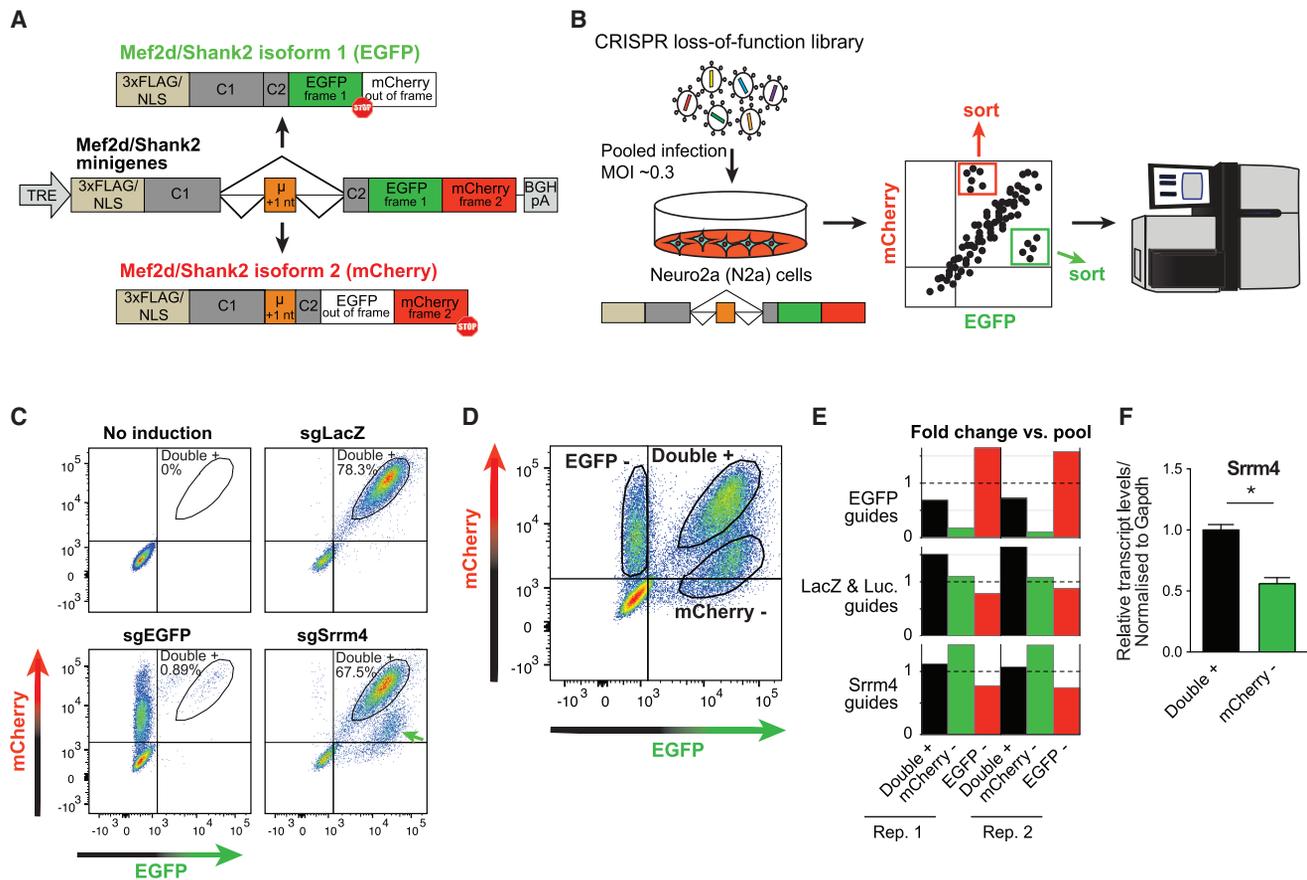
tion on these microexons and their host genes can be found in [Supplemental Information](#).

A single nucleotide was added to each microexon such that their inclusion and skipping results in expression of alternate downstream reading frames that express mCherry and EGFP proteins, respectively (Figure 1A). A nuclear localization signal (NLS) was added to each construct to concentrate fluorescent signals in the nucleus. Mouse neuroblastoma (N2A) Flp-In lines were created (STAR Methods) to enable the stable expression of each reporter from the Rosa26 locus under doxycycline (dox)-inducible control. These features were introduced to minimize cell-to-cell variability in signal and eliminate background fluorescence from prolonged transgene expression.

Confirming the responsiveness of the splicing reporters, short hairpin RNA (shRNA) knockdown of Srrm4 results in microexon skipping, which is associated with increased EGFP and reduced mCherry expression, whereas ectopic expression of an Srrm4 transgene has the opposite effect (Figures S1C and S1D). To test the responsiveness of the reporters to CRISPR-Cas9 targeting, each cell line was transduced with lentiviral vectors expressing Cas9 with four independent, single guide (sg)RNAs targeting Srrm4 or control guides targeting LacZ, Luciferase, or EGFP. As expected, expression of the Srrm4-targeting sgRNAs results in cell populations with increased EGFP and reduced mCherry expression, compared to expression of non-targeting controls, whereas expression of sgRNAs targeting EGFP eliminates EGFP expression (Figures 1C and S1E–S1G). To evaluate responsiveness in a screen format, we next transduced a test library of pooled lentiviral vectors expressing the guides described above (Figure 1D). Following fluorescence-activated cell sorting (FACS), genomic DNA was isolated and integrated sgRNA sequences were amplified from the sorted cell populations and subjected to high-throughput sequencing. As expected, Srrm4-targeting guides were enriched in the mCherry-minus population, EGFP targeting guides in the EGFP-minus population, and LacZ and Luciferase guides in the double EGFP- and mCherry-plus population (Figure 1E). Further confirming successful targeting, sorted mCherry-minus cells have decreased levels of Srrm4 transcripts (Figure 1F) and reduced levels of splicing of the reporter transcripts as well as endogenous microexons compared to EGFP-plus/mCherry-plus cells (Figure S1H). Collectively, these results demonstrate that the microexon splicing reporter cell lines are sensitive to CRISPR-Cas9 targeting of endogenous splicing regulators and thus are amenable to genome-wide CRISPR-based screening.

### Systematic Identification of Microexon Regulators Using Genome-wide CRISPR-Cas9 Screens

The microexon splicing reporter cell lines were transduced with a sgRNA lentiviral library targeting 19,674 protein-coding genes, with four guides per gene (Doench et al., 2016). Cells within ~2% or ~30% of the highest or lowest EGFP:mCherry ratios were collected, and sgRNA enrichment was analyzed by high-throughput sequencing, in each case using data from three independent replicate experiments (Figure 1B). Collectively, the screens captured 233 high-confidence hits that impact microexon splicing (false discovery rate [FDR] < 0.1; Figures 2A and S2A; Table S1). Importantly, among the top



**Figure 1. Generation of Neural Cell Lines (N2A) Expressing Dual Fluorescent Microexon Splicing Reporters**

(A) Schematic diagram of the bichromatic microexon reporters (from *Mef2d* or *Shank2* genes) whose alternative splicing results in the expression of either EGFP or mCherry. The reporters include the microexons (engineered with a +1 reading frameshift such that inclusion or skipping results in mCherry or EGFP expression, respectively), flanking introns, and constitutive C1 and part of C2 exons.

(B) Overview of genome-wide CRISPR loss-of-function screens for the identification of novel microexon regulators.

(C) FACS analysis of N2A cells expressing the *Shank2* microexon reporter (24 hr induction) and transduced with lentiviral expression cassettes for Cas9 and sgRNAs targeting the indicated genes. The population expressing EGFP and mCherry (double +) is indicated. Green arrow, emerging population after sgSrrm4 treatment resulting in microexon skipping.

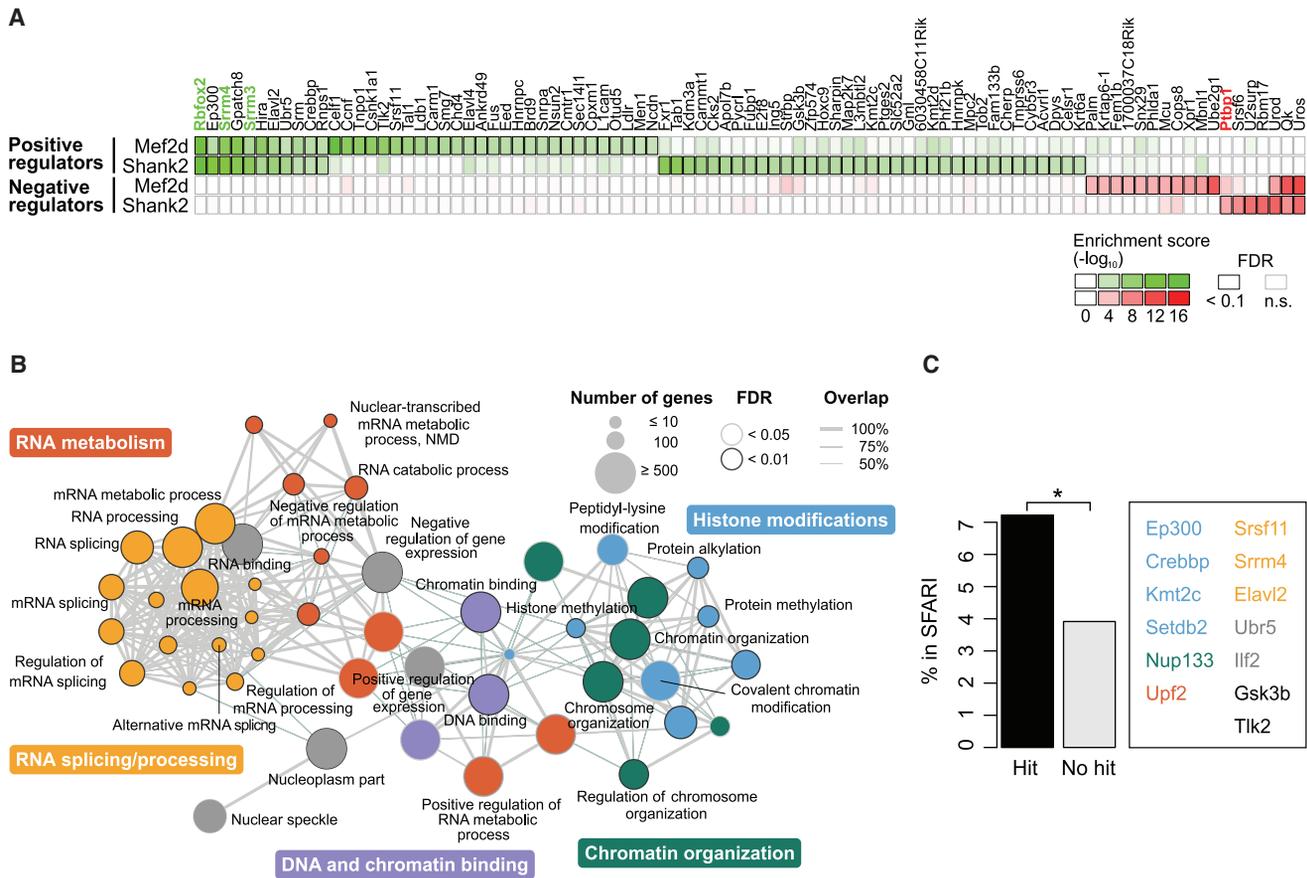
(D) FACS analysis as in Figure 1C transduced with a library consisting of sgRNAs targeting EGFP, Srrm4, and non-targeting controls (LacZ and Luciferase). Three distinct populations were sorted based on EGFP and mCherry expression as indicated.

(E) Enrichment of sequenced sgRNAs (targeting EGFP, non-targeting controls, or Srrm4) in cell populations sorted as depicted in Figure 1D and compared to unsorted cells. Results from two independent replicate experiments are shown.

(F) Real-time qRT-PCR quantification of Srrm4 transcript levels in the indicated sorted populations. Transcript levels are normalized to Gapdh. Error bars indicate SD. \* $p < 0.05$ ; two-tailed unpaired t test.

hits for exon inclusion-promoting genes in both reporter screens are Srrm4 and Rbfox2, which is the only Rbfox family member with appreciable expression in N2A cells. Srrm3, a previously uncharacterized paralog of Srrm4 (sharing ~30% identity) was also identified among the top-scoring regulators. Gene ontology analysis of the screen hits reveals enrichment for specific terms related to pre-mRNA processing and chromatin organization (Figure 2B; Table S1), with other screen hits comprising genes with diverse additional regulatory functions associated with cell signaling, protein turnover, and mRNA metabolism (see below). Remarkably, the screen hits further display significant enrichment for genes that are genetically associated with ASD (Figure 2C; odds ratio: 1.91;  $p < 0.05$ ; Fisher's exact test).

To validate the screen results, we performed transduction with vectors expressing Cas9 along with two independent sgRNAs targeting 39 of the genes identified by the genome-wide CRISPR screens and spanning a broad range of different guide enrichment scores (Figure S2B; Table S2; STAR Methods). We additionally tested a *Mef2d* microexon reporter engineered such that inclusion and skipping result in an opposite switch in the EGFP:mCherry ratio as detected with the reporter used in the primary screen (Figure S2C). This control serves to distinguish genes that may impact the relative expression levels of EGFP or mCherry independently of affecting microexon splicing levels. Disruption of the majority (97%) of the independently targeted genes recapitulates results from the genome-wide screen (Figures S2B and S2D),



**Figure 2. Identification of Microexon Regulators by Genome-wide CRISPR Loss-of-Function Screens**

(A) Relative effects of CRISPR screen hits, as detected by sgRNA enrichment in sorted cell populations. Genes identified as positive (green) or negative (red) regulators of microexon splicing in N2A cells expressing the Shank2 or Mef2d microexon reporters with FDR < 0.1 and with a fold enrichment higher than 1.4 are shown.

(B) Analysis of Gene Ontology (GO) enrichment among genes identified as microexon regulators by the genome-wide CRISPR loss-of-function screens.

(C) Percentage of autism spectrum disorder (ASD)-related genes identified as hits in the CRISPR screens compared to the percentage of ASD-related genes represented by targeting sgRNAs and that are expressed in N2A. ASD-related genes are indicated. Colors correspond to GO categories as depicted in Figure 2B (black indicates genes with unrelated GO categories). \* $p = 0.033$ ; Fisher's exact test.

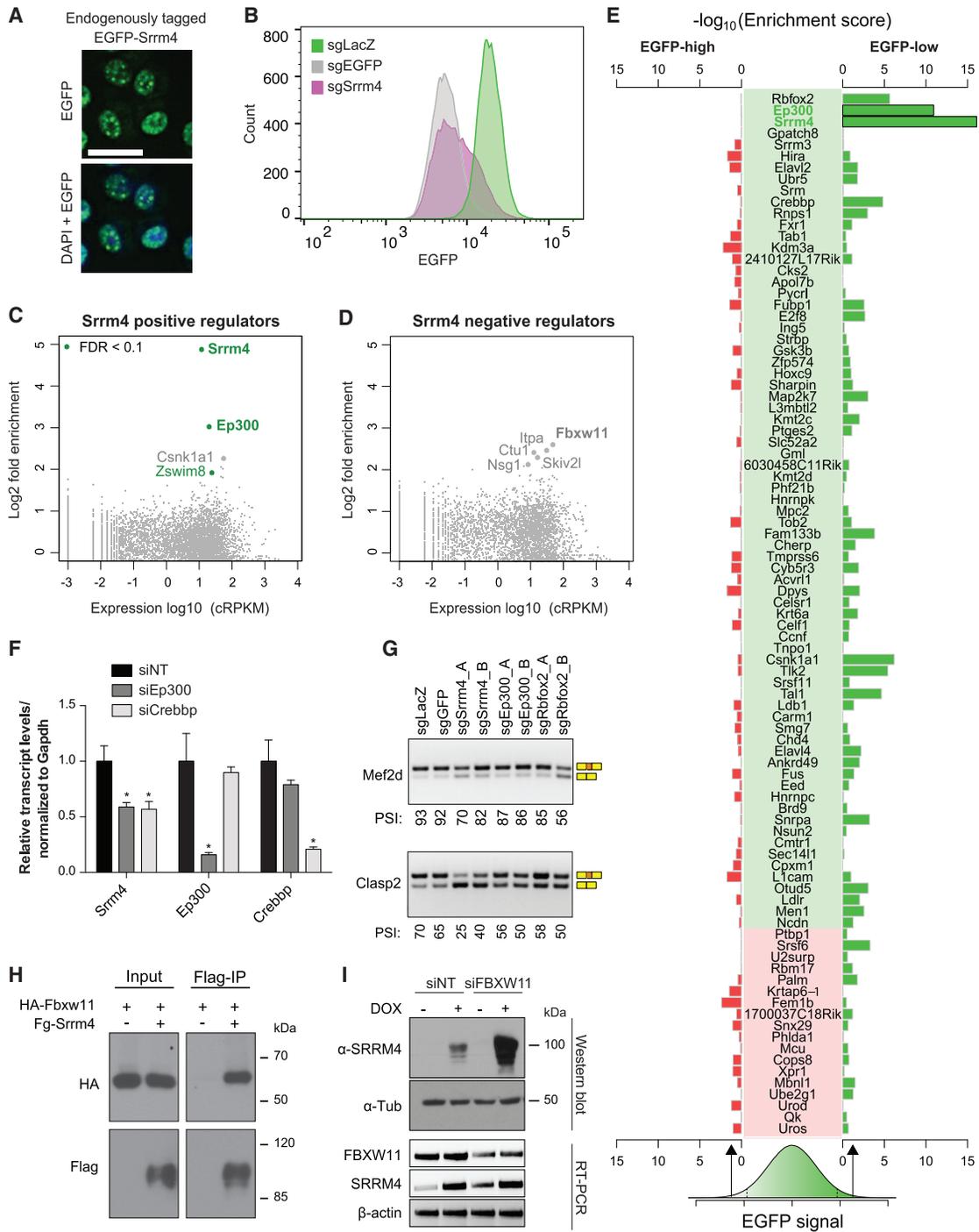
and 80% of the tested genes result in the expected reciprocal shift in EGFP:mCherry ratio using the control Mef2d splicing reporter (Figures S2B and S2E). The remaining genes often reflect false-positive hits that, in some cases, arise due to background fluorescence from non-splicing related mechanisms. For example, ablation of *Uros* and *Urod*, which encode components of the porphyrin biosynthesis pathway, results in red fluorescence independently of microexon splicing (Figure S2F; STAR Methods). By tracking the loss of sgRNA sequences in the unsorted cell populations over time, which chiefly arises due to targeting of fitness-related genes, we estimate a false negative rate of 0.4% (Figures S2G–S2J; STAR Methods).

The results so far demonstrate that our CRISPR screening strategy is effective for the genome-wide identification of genes that control AS events of interest and, more specifically, that it has identified known and new regulators of microexons that are functionally diverse and enriched in genetic links to autism.

### Systematic Identification of Srrm4 Expression Regulators

Genes identified as microexon regulators by our screen may act through direct or indirect mechanisms. To distinguish these possibilities, we next performed a genome-wide, CRISPR screen to detect genes that affect the expression of *Srrm4*. *Srrm4* was endogenously tagged with EGFP in N2A cells by CRISPR-Cas9-mediated gene editing (Figures 3A and S3A). Interestingly, similar to untagged endogenous protein, the EGFP-Srrm4 fusion protein displays a speckle-like nuclear localization pattern that only partially overlaps that of the canonical SR protein SRSF2/SC35 (Figures S3B and S3C). As expected, expression of siRNA or sgRNAs targeting *Srrm4* or EGFP results in reduced EGFP-Srrm4 expression (Figures 3B, S3A, and S3D).

The N2A EGFP-Srrm4 cell line was transduced with a genome-wide loss-of-function library, and cells with the 2% highest or lowest EGFP expression were collected and analyzed, as described above. Confirming the effectiveness of the screen,



**Figure 3. Systematic Identification of Srrm4 Expression Regulators**

(A) Fluorescent micrographs of N2A cells with endogenously EGFP-tagged Srrm4. Scale bar represents 27  $\mu$ m.

(B) FACS analysis of the EGFP-Srrm4 N2A cell line transduced with lentiviral expression cassettes for Cas9 and sgRNAs targeting Srrm4, EGFP, or LacZ (non-targeting).

(C and D) Scatterplots representing the average log<sub>2</sub> fold enrichment of guides targeting each gene in the sorted population with changed EGFP-Srrm4 expression (y axis) and the corresponding expression of each gene in N2A cells (x axis), as detected by RNA-seq analysis. Significant hits (FDR < 0.1) are in green. (C) Reduced EGFP expression. (D) Increased EGFP expression.

(legend continued on next page)

Srrm4-targeting sgRNAs are the most highly enriched in the EGFP-minus population (Figure 3C). Notably, only one of the positive regulators of microexon splicing identified in the bichromatic reporter screens, Ep300, is also identified as a positive regulator of EGFP-Srrm4 expression, although several others, including the Ep300 paralog Crebbp, Rbfox2, Csnk1a1, and Tlk2, display modest effects on Srrm4 expression and are below the significance threshold applied in our analysis (Figures 3C and 3E). Conversely, the E3 ubiquitin ligase Fbxw11 was detected as the strongest negative regulator of Srrm4 expression (Figure 3D).

Previously, we reported that Srrm4 expression is transcriptionally repressed by Rest in non-neural cells (Raj et al., 2011). However, positive regulators of Srrm4 expression and orthologous mechanisms controlling Srrm4 levels have not been previously described. EP300 and its paralog CREBBP are lysine acetyltransferases and chromatin remodelers that predominantly regulate gene expression via acetylation of histones. Depletion of Ep300 and Crebbp results in a significant reduction in Srrm4 transcript levels (Figures 3F and S3F; FDR < 0.05), suggesting that these factors promote Srrm4 gene expression in neural cells. To investigate this, we analyzed Ep300 chromatin immunoprecipitation sequencing (ChIP-seq) data (Visel et al., 2009) and identified significant peaks in the promoter region of Srrm4 (STAR Methods) in brain, but not limb, tissue (Figure S3G). Consistent with Ep300 and Crebbp promoting Srrm4 expression via histone acetylation at the Srrm4 promoter, treatment of neurons with the deacetylase inhibitor trichostatin A results in a two-fold increase in Srrm4 expression (Figure S3H). Moreover, as expected, depletion of Ep300 and Crebbp results in reduced expression of Srrm4 protein levels (Figure S3E) and in skipping of endogenous microexons (Figure 3G). Collectively, these data provide evidence that Ep300 and Crebbp stimulate microexon splicing by promoting Srrm4 expression, at least in part through increasing histone acetylation.

Fbxw11 acts as part of a Skp-Cullin-F-box (SCF) complex that regulates ubiquitin-dependent proteasomal degradation (Suzuki et al., 1999). Previous affinity purification coupled to mass spectrometry (AP-MS) experiments provided evidence that FBXW11 interacts with SRRM4 (Raj et al., 2014), which we confirm in the present study by co-immunoprecipitation (co-IP) western blot analysis (Figure 3H). This interaction suggests that Fbxw11 regulates Srrm4 protein levels by controlling its proteolysis. Indeed, knockdown of FBXW11 in a dox-inducible 293T cell line expressing exogenous SRRM4 or, in N2A cells, results in pronounced increases in SRRM4/Srrm4 protein levels (Figures 3I, S3I, and S3J). Furthermore, Fbxw11 overexpression results in reduced

levels of Srrm4, and this activity is inhibited by the proteasome inhibitor MG132 (Figure S3K). These data provide evidence that Fbxw11 reduces Srrm4 expression via ubiquitin-dependent proteolysis, thereby revealing an additional layer of control of neuronal AS.

To summarize, through a genome-wide screen for regulators of Srrm4 protein levels, only a small percentage of genes have a significant impact and overlap those that control microexon inclusion levels in the genome-wide biochromatic splicing reporter screens (Figure 3E). This suggests that the large majority of the detected microexon regulators function through mechanisms other than those that impact Srrm4 protein levels.

### Secondary SPAR-Seq Screening of Regulators of Endogenous Microexons

To investigate whether high-scoring hits from the bichromatic reporter screens more generally regulate microexons, we next performed secondary screens using our previously described method for simultaneously linking *trans*-acting factors to multiple AS events, “systematic parallel analysis of endogenous RNA regulation coupled to barcode sequencing” (SPAR-seq) (Han et al., 2017). siRNA knockdowns of 15 genes identified in the CRISPR-based screen and involved in RNA processing were tested for effects on the splicing of 20 microexons (including orthologs of 10 ASD-misregulated microexons) and, for comparison purposes, 12 longer neural cassette exons (Figures 4A and S4A). Consistent with the results from the bichromatic reporter screens, SPAR-seq reveals that knockdown of Srrm4 and Srrm3 have the most pronounced and highly correlated effects on microexon splicing, with 16 of the 20 tested microexons being affected (Figures 4A and S4A; Table S3). Interestingly, changes in microexon splicing upon knockdown of Rnps1, an SR-related splicing activator and auxiliary component of the exon-junction complex (EJC) (Le Hir et al., 2016; Lykke-Andersen et al., 2001; Mayeda et al., 1999), correlate strongly with those detected upon knockdown of Srrm4 and Srrm3 (Figures 4A and S4A). Furthermore, knockdown of Srsf11 (also known as SRp54), an SR family protein that interacts with polypyrimidine tract binding components (Page-McCaw et al., 1999; Zhang and Wu, 1996) and Rnps1 (Sakashita et al., 2004), impacts a similar set of microexons as Srrm4/Srrm3 and Rnps1 (Figures 4A, 4B, and S4A). A distinct, although partially overlapping, subset of microexons displays correlated changes in splicing upon knockdown of Rbfox2, Gpatch8, and Hnrnpk (Figure 4A). As mentioned above, Rbfox family proteins are known to control the splicing of some microexons (Li et al., 2015; see below). In

(E) Enrichment scores for EGFP-Srrm4 expression in relation to CRISPR screen hits identified in the bichromatic reporter screens (Figures 2A and S2A). Scores in the populations with the 2% lowest (EGFP-low) or highest (EGFP-high) EGFP expressions are depicted as for screen hits in Figure 2A. Significant hits (FDR < 0.1) are highlighted in green.

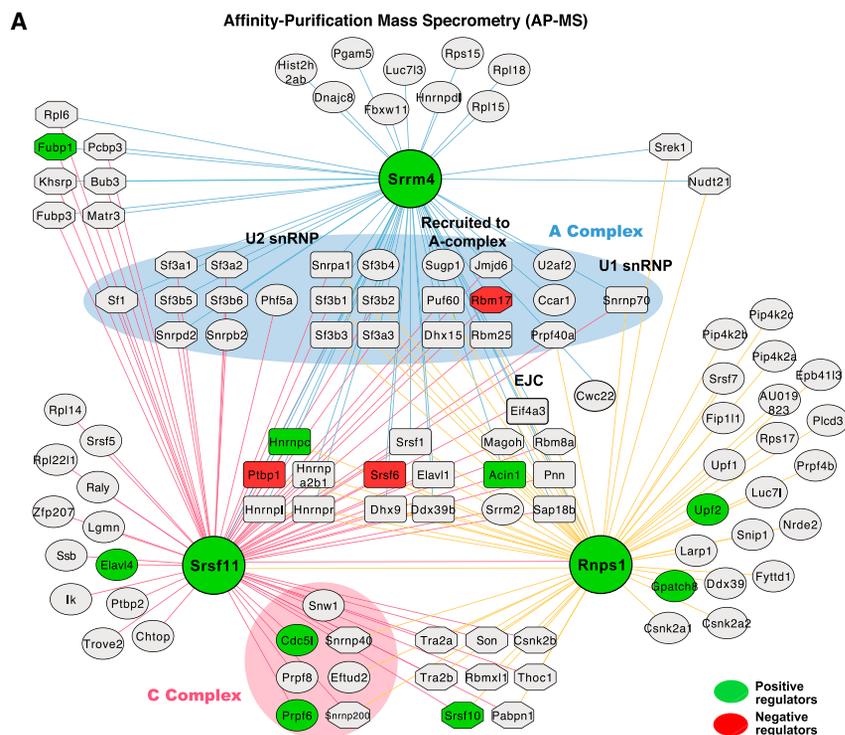
(F) Real-time qRT-PCR analysis quantifying Srrm4, Ep300, and Crebbp transcript levels after siRNA (si) knockdown of Ep300, Crebbp, or transfection of a non-targeting (siNT) control. Transcript levels are normalized to Gapdh. Error bars indicate SD. \*p < 0.05; two-tailed unpaired t test.

(G) RT-PCR assays monitoring endogenous microexon splicing in N2A cells transduced with lentiviral expression cassettes for Cas9 and two independent sgRNAs targeting Srrm4, Rbfox2, Ep300, or non-targeting controls, as indicated.

(H) Western blot analysis of inputs and Flag immunoprecipitates from N2A Flp-In cells expressing Flag-Srrm4 and transfected with constructs expressing HA-Fbxw11, using Flag and HA antibodies. Flag immunoprecipitation was performed in lysates treated with benzonase, RNase A, and RNaseT1.

(I) Upper panels: western blot analysis of 293T cells expressing Flag-SRRM4 under dox induction and transfected with siRNAs targeting FBXW11 or a non-targeting siRNA (siNT) control pool. Blots were probed with antibodies specific for SRRM4 and  $\alpha$ -tubulin, as indicated. Lower panels: RT-PCR analysis of RNA samples corresponding to those shown in the upper panels, monitoring the expression levels of FBXW11 and SRRM4.



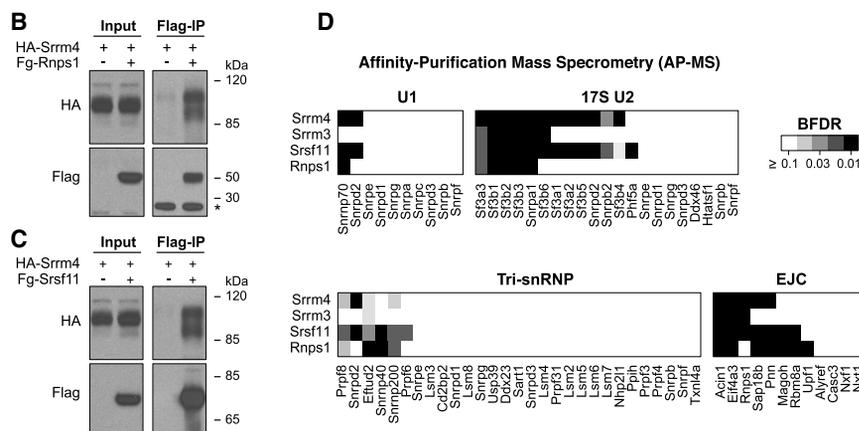


**Figure 5. Analysis of the Srrm4 Interactome**

(A) Protein-protein interaction networks involving Srrm4 (blue edges), Srsf11 (red edges), and Rnps1 (yellow edges) as detected by AP-MS. Interactors with Bayesian false discovery rate (BFDR) < 0.01 are depicted. Interactors identified as positive or negative regulators of microexon splicing in the CRISPR bichromatic reporter screens are visualized in green or red, respectively. Preys shared between Srrm4, Srsf11, and Rnps1 are indicated by round-edged rectangles; preys shared between pairs of the baits are indicated by octagons, and preys unique to Srrm4, Srsf11, or Rnps1 are indicated by ovals.

(B and C) Western blot analysis of inputs (left panel) and Flag immunoprecipitates (right panel) from N2A cells expressing HA-Srrm4, Flag-Rnps1 (B), Flag-Srsf11 (C), or both proteins using anti-Flag and anti-HA antibodies. Flag immunoprecipitation was performed in lysates treated with benzonase, RNase A, and RNaseT1. \* indicates Flag antibody light chain.

(D) BFDR values indicating relative degrees of enrichment of interactors (from AP-MS analysis of Srrm4, Srrm3, Srsf11, or Rnps1) among all proteins annotated in SplicosomeDB (see STAR Methods) and sorted according to membership in splicing-associated complexes (U1 snRNP, U2 snRNP, tri-snRNP, or EJC).



These data reveal that genes identified in our genome-wide splicing reporter screens form a network of protein-protein interactions that function early in the formation of spliceosomes and further suggest that Srrm4 activates microexon splicing through interactions involving Srsf11 and Rnps1.

### Srsf11 and Rnps1 Co-regulate a Program of Srrm4-Dependent Microexons

To investigate whether interactions between Srsf11, Rnps1, and Srrm4 are

Srrm4 co-immunoprecipitates with both Flag-Rnps1 and Flag-Srsf11 proteins expressed in the N2A Flp-In cells in an RNA-independent manner (Figures 5B and 5C).

In light of the similar effects of depletion of Srrm4, Srsf11, and Rnps1 on microexon splicing (Figure 4) and detection of interactions between these factors, we further investigated the “Srrm4 interactome” by performing reciprocal AP-MS and BiO-MS experiments using tagged Srsf11 and Rnps1 proteins expressed in N2A cells. Rnps1 and Srsf11 indeed share multiple interaction and proximal partners with Srrm4, particularly members of the spliceosomal U2 snRNP and the EJC (Figures 5A, 5D, and S5B; Table S4). For comparison and specificity control purposes, we compared the BiO-MS profiles of Srrm3, Rbfox2, Ptpb1, and Ptpb2. As expected, the interaction profile of Srrm3 is similar to that of Srrm4, whereas Rbfox2, Ptpb1, and Ptpb2 have distinct interaction profiles (Figure S5B).

more generally important for the activation of neuronal microexons, we knocked down each of these proteins in N2A cells and performed RNA-seq. This analysis assessed a total set of 531 detected microexons, of which 193 (36%) are neural specific, as well as effects on longer alternative exons (Figures S6A and S6B). Similar to knockdown of Srrm4, which affects 51% of neural microexons, knockdown of Srsf11 or Rnps1 results in widespread skipping of neural microexons, impacting 32% and 47%, respectively. In contrast, knockdown of these factors affected significantly smaller numbers (~5% in each knockdown) of non-neural microexons (Figures 6A, S6C, and S6D; Table S5;  $p < 0.001$ ; Fisher’s exact test). Moreover, consistent with the results in Figure 4, the subsets of microexons affected by each knockdown show extensive and highly significant degrees of overlap (Figure 6B;  $p < 0.001$ ; Fisher’s exact test), as well as extensive overlap with orthologous microexons



misregulated in autistic brains (Figure S6F). In contrast, longer (>27 nt) affected exons show substantially lower although significant degrees of overlap ( $p < 0.001$ ; Fisher's exact test; Figure S6E). Moreover, the correlation coefficients of knockdown-induced AS changes are higher for microexons than longer cassette alternative exons (Figure S6G). Importantly, the effects of *Srrm4*, *Srsf11*, and *Rnps1* knockdown are not due to off-target effects, because ectopic expression of these factors rescues splicing levels after siRNA treatment (Figures S1B, S6H, and S6I).

*Ptbp1*, which primarily negatively regulates neuronal exons (Boutz et al., 2007; Li et al., 2015; Raj et al., 2014; Figure 2A), has opposite effects as knockdown of *Srrm4*, *Rnps1*, and *Srsf11* (Figure 6A), and knockdown of *Rbfox2* results in a significantly lower relative proportion of splicing changes impacting neuronal microexons compared to non-neuronal microexons (Figures 6A and S6J;  $p < 0.03$ ; one-sided binomial test). Together with the proteomics data, the extensive overlap between the subsets of neural microexons regulated by *Srrm4*, *Srsf11*, and *Rnps1* demonstrates that these factors function together to promote the splicing of a large program of ASD-associated microexons.

To further investigate how *Srsf11* and *Rnps1* function with *Srrm4* to promote microexon splicing, we next performed individual nucleotide crosslinking and immunoprecipitation coupled to sequencing (iCLIP-seq) analysis of these factors in N2A cells. Because *Rnps1* functions as a peripheral component of EJC in addition to its role in splicing (Lykke-Andersen et al., 2001; Mayeda et al., 1999), we focused our analysis on iCLIP-seq reads derived from pre-mRNA (i.e., intronic reads or reads spanning exon-intron junctions) so as to detect possible targets of *Rnps1*-mediated splicing regulation (Figure S6K), although it should be noted that *Rnps1* may also function in splicing in association with the EJC (Wang et al., 2018). Strikingly, the resulting binding maps for *Srsf11*, *Rnps1*, and *Srrm4* reveal that they form strong, overlapping occupancy peaks positioned proximally upstream of exon 3' splice sites (Figure 6C; Raj et al., 2014). Moreover, consistent with the observation that *Srsf11* and *Rnps1* preferentially regulate *Srrm4*-dependent neural microexons (Figures 6A, S6C, and S6D), the binding of these factors is enriched on pre-mRNA proximal to *Srrm4*-regulated microexons compared to longer cassette exons (Figures S6L and S6M). The results suggest that *Srsf11* and *Rnps1* function to preferentially facilitate *Srrm4*-dependent functional interactions on pre-mRNA that activate microexon splicing.

Considering that *Srsf11* and *Rnps1* are widely expressed, it is striking that they preferentially promote splicing of neural-regulated microexons. However, they are not sufficient to promote microexon splicing in a non-neural context (Figure S6N; data not shown). In contrast, expression of *Srrm4* in non-neural cells activates the splicing of otherwise silent microexons (Figure S6N; Irimia et al., 2014). Accordingly, we hypothesized that expression of *Srsf11* and/or *Rnps1* facilitate *Srrm4*-dependent splicing activity. To test this, we knocked down *Srsf11*, *Rnps1*, and/or *Srrm4* in an N2A cell line engineered to ectopically express *Srrm4* under dox-inducible control. Following knockdowns and induction of *Srrm4* expression, the splicing of *Srrm4*-dependent microexons was assayed by RT-PCR. Notably, in the absence of

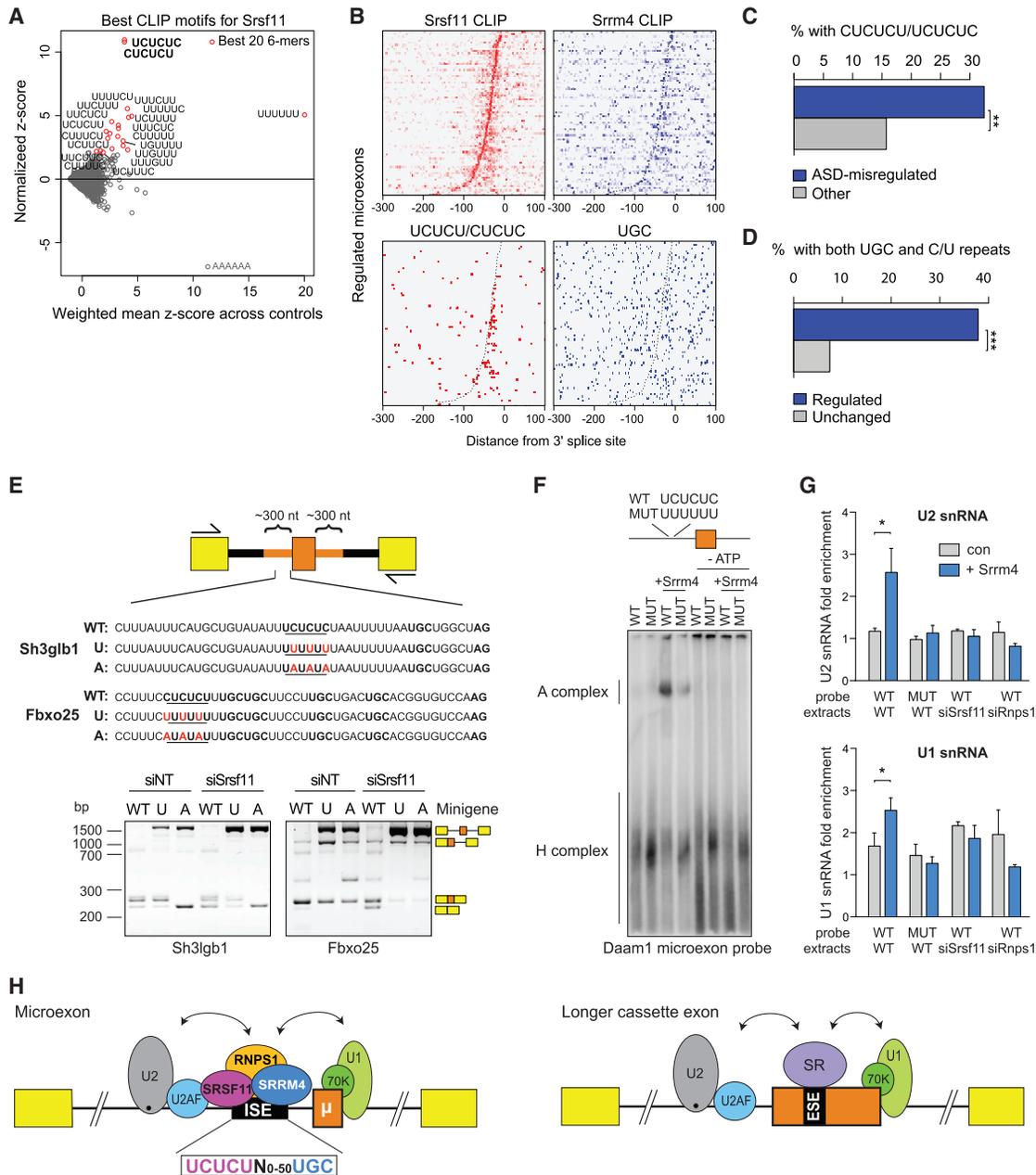
*Srsf11* or *Rnps1*, microexon splicing is markedly reduced upon expression of *Srrm4* (Figures 6D and 6E). Similarly, knockdown of *SRSF11* in 293T cells reduced the levels of *Srrm4*-dependent microexon splicing (Figure S6N). To investigate whether these effects are the consequence of altered physical interactions and not due to indirect effects, we asked whether *SRSF11* and *RNPS1* are required for *SRRM4*-dependent splicing of a microexon (exon 16 from the *Daam1* gene) *in vitro*. Consistent with a requirement for *SRSF11* in promoting microexon splicing, 293T cell extracts depleted of *SRSF11* display reduced levels of exon 16 inclusion in the presence of recombinant *SRRM4* (Figure 6F), whereas addition of recombinant *Srsf11* stimulates *SRRM4*-dependent activity and rescues reduced splicing of *Daam1* exon 16 in the *SRSF11*-depleted extract (Figure 6F). Similarly, recombinant *Rnps1* promotes *SRRM4*-dependent splicing of *Daam1* exon 16 (Figure 6G). These effects are specific because addition of equivalent levels of a control protein (BSA) does not appreciably alter *Daam1* exon 16 splicing levels.

Next, to address whether *Srsf11* and *Rnps1* facilitate recruitment of *SRRM4* to RNA, or else whether these proteins might support each other's interaction with transcripts, gel shift assays were performed using RNA probes containing the *Daam1* or *Mef2d* microexons and 100 nt upstream intronic sequence. Purified recombinant *SRRM4* protein binds RNA, albeit weakly. However, when *SRRM4* is incubated with *Srsf11* or *Rnps1*, each pair of proteins binds more stably to RNA, forming super-shifted complexes (Figures 6H, 6I, and S6P), suggesting that physical interactions involving *Srsf11* and/or *Rnps1* are critical for *SRRM4*-dependent microexon splicing. To confirm this, we performed splicing complex formation assays. Recombinant *SRRM4* promotes A complex formation in an ATP-dependent manner (Figure S6Q). Importantly, addition of recombinant *Rnps1* or *Srsf11* strongly enhances *SRRM4*-dependent A complex formation (Figures 6J, S6R, and S6S). Collectively, these data reveal that binding of *SRRM4*, *Rnps1*, and *Srsf11* upstream of neuronal microexons promotes early steps in spliceosome assembly.

### **SRRM4-Dependent Neuronal Microexons Are Spliced via a Unique Exon-Definition Mechanism**

The data presented above suggest that microexons could be recognized by a set of interactions in which *Srrm4*, *Rnps1*, and *Srsf11* form a specialized intronic splicing enhancer (ISE) complex that "substitutes" for the role of SR proteins bound to exonic enhancers, a key feature of classical exon definition interactions (Berget, 1995). In this model, the *Srrm4*/*Rnps1*/*Srsf11* ISE complex forms U2 snRNP stabilizing and cross-microexon bridging interactions with U1 snRNP. To test this model, we determined the effects of 5' splice site inactivating mutations (Figure S7A) on three microexons using reporter constructs containing native upstream and downstream intronic and flanking constitutive exonic sequences. Remarkably, all three of these mutations prevented splicing of the upstream introns (Figure S7A).

We next investigated the mechanism by which *Srsf11*, *Rnps1*, and *Srrm4* promote exon-definition interactions and microexon splicing. Previous studies identified UGC-containing motifs as *Srrm4* binding sites that are critical for the splicing of microexons



**Figure 7. Binding of SRSF11 to C/U Intronic Enhancer Motifs Promotes Srm4 Recruitment and Microexon Splicing**

(A) Enrichment of hexamers at Srsf11 iCLIP-seq binding peaks compared to background enrichment (see STAR Methods).

(B) Occurrences of U/C repeat (bottom left) and UGC motifs (bottom right) coinciding with Srsf11 iCLIP (top left) and Srm4 iCLIP (top right) binding peaks and relative to the 3' splice sites of microexons regulated by Srsf11. Dotted line indicates highest Srsf11 signal.

(C) Percentage of ASD-misregulated or non-misregulated microexons containing a U/C repeat motif. Odds ratio = 2.6; \*\**p* < 0.01 Fisher's exact test.

(D) Percentage of Srm4/Srsf11/Rnps1 co-regulated or non-regulated microexons with co-occurrence of U/C repeats and UGC motifs. Odds ratio 9.6; \*\*\**p* < 0.001 Fisher's exact test.

(E) RT-PCR assays monitoring splicing of "wild-type" (WT) and mutant minigene reporters transfected into N2A cells treated with siRNAs targeting Srsf11 or a non-targeting (NT), control siRNA pool. U/C repeat elements were mutated to U/U (U) or U/A (A) repeats.

(F) Splicing complex assembly assays using WT and U/C mutated Daam1 microexon probe in the presence and absence of recombinant Srm4 (250 ng) and/or ATP as indicated.

(G) Real-time qRT-PCR analysis for the quantification of U1 and U2 snRNA after streptavidin capture of biotinylated WT or mutant (MUT) Daam1 probes incubated with 293T cell extracts, WT, or depleted of Rnps1 or Srsf11. Error bars indicate SE. \*, significant difference; *p* < 0.05; one-tailed unpaired *t* test.

(legend continued on next page)

and longer neuronal exons (Irimia et al., 2014; Nakano et al., 2012; Raj et al., 2014). However, these motifs are prevalent and lack sufficient information to specify microexon recognition. Analysis of the Srsf11 iCLIP-seq binding peaks reveals enrichment of U/C repeat motifs (Figure 7A; Table S6), consistent with previous evidence that this protein associates with polypyrimidine sequences (Page-McCaw et al., 1999). Indeed, overall, Srsf11 binding peaks correlate strongly with the presence of a U/C repeat motif upstream of 3' splice sites (Figures 7B and S7B), and this motif is preferentially enriched upstream of microexons compared to longer cassette exons regulated by Srsf11 (Figure S7C). This motif is also significantly enriched upstream of microexons misregulated in ASD subjects (Figure 7C;  $p < 0.01$ ; Fisher's exact test). Importantly, we also detect a significant co-occurrence of U/C repeat and UGC motifs adjacent to neuronal microexons compared to non-regulated microexons or longer cassette alternative exons, as well as adjacent to microexons co-regulated by Srrm4, Rnps1, and Srsf11 (Figures 7D and S7H;  $p < 0.001$ ; Fisher's exact test). In contrast, Rnps1 binding peaks do not display clear enrichment of specific sequence motifs (data not shown). Moreover, illustrating specificity of these findings, Rbfox2-regulated microexons do not display a significant enrichment for the co-occurrence of U/C repeat and UGC motifs (Figure S7H).

The combination of U/C repeat and UGC motifs thus appears to represent a composite, bi-partite interaction site for an ISE complex containing Srsf11 and Srrm4 that is required for microexon recognition and splicing. From analyzing the full network of Srrm4, Rnps1, and Srsf11 co-regulated microexons, the distance between the UGC and U/C repeat motifs generally is between 0 and 50 nt, with the U/C repeat typically located upstream of the UGC motif, which is within 2–20 nt of the 3' splice site (Figures S7D and S7F). Confirming the importance of the predicted bi-partite ISE, mutation of the U/C repeat motifs (to either U/U or U/A repeat sequences; Figures 7E and S7I) upstream of three analyzed microexons results in pronounced increases in the levels of unspliced pre-mRNA, including an accumulation of the downstream introns (Figures 7E and S7I). Moreover, depletion of Srsf11 results in a further increase in the levels of unspliced pre-mRNA, especially in the context of the U/C repeat motif mutations (Figures 7E and S7I). Finally, we tested the requirement of the U/C repeat motif for splicing complex formation using “wild-type” (WT) or mutant (MUT) microexon probes. We observed a dramatic reduction in A complex formation on the mutant probes (Figures 7F and S7J). Further corroborating these results, the Srrm4-promoted association of U2 and U1 snRNA with splicing signals flanking microexons is lost in the context of the U/C repeat motif mutation (MUT) as well as in the absence of Rnps1 or Srsf11 (Figure 7G). These data provide strong evidence that Srsf11, through its binding to C/U repeat motifs and interaction with Srrm4 bound at adjacent UGC motifs, as well as with Rnps1, contributes to a unique set of exon-definition interac-

tions that are critical for neuronal microexon recognition and splicing (Figure 7H).

## DISCUSSION

In this study, we describe a CRISPR-Cas9-based screening strategy for defining the full repertoires of genes that impact the regulation of AS. An important advantage of this strategy is that CRISPR-based screens, compared to RNAi screens, have increased sensitivity, reduced off-target effects, and are relatively facile to perform on a genome-wide scale (Evers et al., 2016; Hart et al., 2015). Demonstrating efficacy, we identify ~230 genes representing diverse regulatory layers that control neuronal microexons, many of which have not been previously linked to splicing or microexon regulation. Our results thus introduce a new functional genomics platform with which to comprehensively elucidate pathways and mechanisms that regulate functionally important AS events.

The results from our screens and follow-up experiments further provide a broadly applicable model for the tissue-specific recognition of very short exons. In this model, the bi-partite ISE bound by Srsf11, Rnps1, and Srrm4 obviates the requirement for an exonic splicing enhancer (ESE) and thus enables the recognition of exons that are too short to harbor these elements or else may be subject to protein coding constraints that preclude the positioning of ESEs within exons. This model contrasts with a previous proposal for sequential, intron-definition-type interactions in the recognition of microexons (Sterner and Berget, 1993) and is also different from a more recent model invoking exon-enhancer-dependent interactions in the context of the extensively studied neuronal (N1) microexon of the Src gene (Wongpalee et al., 2016). This microexon is repressed in non-neural cells by Ptbp1-dependent interactions (Sharma et al., 2005), yet its mechanism of activation in neural cells has largely remained unclear (Wongpalee et al., 2016). We observe that knockdown of Srrm4, Srsf11, and Rnps1 all result in increased skipping of the N1 microexon (Figure S4B). Moreover, these interactions broadly activate neuronal microexons, affecting ~65% of those detected in our data. In contrast, Rbfox regulates a distinct and smaller subset (17%) of microexons that is enriched for its cognate binding site (GCAUG), but not the bi-partite ISE, and it also lacks a significant preference for regulating neuronal versus non-neuronal microexons. Collectively, our findings illustrate how ubiquitously expressed RNA-binding proteins, namely Srsf11 and Rnps1, can promote highly specific neuronal splicing patterns. The unique positioning of the specialized ISE to which these proteins bind, in combination with their interaction with Srrm4, all critically contribute to the activation of neuronal microexons.

Although our CRISPR-based strategy is highly effective in capturing regulators of AS, a limitation is that potential factors corresponding to essential or fitness genes may not be detected due to cell loss. In this regard, differences in the detection of

(H) Mechanistic model for the AS of neuronal microexons (top) compared to longer cassette exons (bottom). The longer distance between the 3' splice site and branch point (Figure S7G) allows the accommodation of critical intronic splicing enhancers (ISE) that compensate for the absence of exonic splicing enhancers (ESE) in microexons ( $\mu$ ) due to their short size. A bi-partite ISE comprising a U/C repeat followed by UGC motif (Figures S7D–S7F) is cooperatively recognized by Srsf11 and Srrm4 to form critical protein-protein interactions required for microexon recognition and subsequent formation of spliceosomes.

fitness genes may contribute to the overall low (less than 5%) degree of overlap between genes defined as regulators of Fas and Bcl-X alternative splicing in genome-wide, siRNA-based screens (Moore et al., 2010; Tejedor et al., 2015) and genes defined as microexon regulators in the present study. However, our approach does capture numerous fitness genes, and although such genes are underrepresented among the significant hits, we estimate a false-negative detection rate of 0.4% when considering all 19,674 analyzed genes (Figures S2I and S2J; STAR Methods). As such, the majority of the differences between factors identified in the CRISPR- and previous RNAi-based screens may be due to pre-mRNA-sequence- and cell-context-dependent requirements for splicing. This is also apparent when considering the substantial differences between the factors required for the regulation of the Shank2 and Mef2d reporters analyzed in the present study.

An intriguing observation in the present study is that genes identified as microexon regulators by our CRISPR screen, including Srsf11 and Rnps1, are enriched for genetic links to neurodevelopmental disorders and autism (C Yuen et al., 2017; Nguyen et al., 2013). Given the widespread and specific roles for Srsf11 and Rnps1 in promoting Srrm4-dependent splicing of neuronal microexons, it is interesting to consider that disruption of any of these components, whether through their misregulation or genetic alteration, may collectively be responsible for a substantial fraction of ASD cases. Identification of important interactions between these critical splicing factors thus expands the possibilities for pharmacomodulation of the Srrm4-dependent microexon network disrupted in ASD (Irimia et al., 2014; Quesnel-Vallières et al., 2016). Moreover, the CRISPR-Cas9-based screening strategy described in the present study opens the door to comprehensively elucidating pathways underlying the regulation of other AS events with critical roles in development, disease, and disorders.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Cell Culture
  - Generation of a Neuro-2a (N2A) Cell Line Containing an FRT Recombination Site Using CRISPR-Cas9 (N2A Flp-In)
  - Generation of Stable Flp-In-N2A Cell Lines
  - Endogenous N-terminal EGFP-tagging of Srrm4
  - siRNA Transfections
  - RNA Extraction and RT-PCR Assays
  - Quantitative RT-PCR
  - Assessment of CRISPR/Cas9 Editing Efficiency by Endonuclease Cleavage (Surveyor) Assay
  - Splicing of Minigene Constructs
  - Fluorescence Microscopy
  - Co-immunoprecipitation Experiments
  - Immunoblotting

- FLAG Affinity Purification Coupled with Mass Spectrometry (AP-MS) Sample Preparation
- BioID Sample Preparation for MS
- Mass Spectrometry Data Acquisition
- Neuronal differentiation Trichostatin A treatment
- Expression and Purification of Recombinant Srsf11 and Rnps1
- *In Vitro* Transcription
- *In Vitro* Splicing Assays
- Electrophoretic Mobility Shift Assays (EMSA)
- Splicing Complexes Assembly Assays
- Affinity Purification of Splicing Complexes
- Individual-nucleotide Resolution Cross-linking and Immunoprecipitation (iCLIP-Seq)
- RNA-Sequencing
- SPAR-Seq
- Engineering Bichromatic Microexon Splicing Reporters
- Cloning of sgRNAs into LentiCRISPRv2 Vectors for Validation
- Virus Production and MOI Determination
- Test Library Screen
- Genome-wide CRISPR-Cas9 Screens
- Validation of Hits Identified by the Genome-wide CRISPR-Cas9 Screens Using Flow Cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - CRISPR Screen Analysis
  - Analysis of Flow Cytometry Validations
  - SPAR-Seq Analysis
  - RNA-Seq Analysis
  - Annotation of Exons Misregulated in ASD
  - iCLIP-Seq Analysis
  - ChIP-Seq Analysis
  - Mass Spectrometry Data Analysis
  - Analysis of Interactions with Spliceosomal snRNPs
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.molcel.2018.10.008>.

## ACKNOWLEDGMENTS

The authors thank J. Valcarcel, M. Irimia, S. Bonnal, R.J. Weatheritt, and R. Barutcu for helpful discussions and comments on the manuscript. T. Durbic and G. O'Hanlon in the Donnelly Sequencing Centre are gratefully acknowledged for generating data. We also thank S. Angers, S. Nabeel-Shah, S. Bonnal, M. Mis, K.R. Brown, Y.-Q. Lü, and A. Tong for advice and assistance on experiments and D. White, A. Bang, and J. Warzyszyńska in the University of Toronto Faculty of Medicine and Lunenfeld-Tanenbaum Research Institute Flow Cytometry Facilities for assistance with FACS. T.G.-P. was supported by postdoctoral fellowships from EMBO, CIHR, and OIRM; U.B. was supported by an HFSP postdoctoral fellowship; J.R. was supported by an Ontario Graduate Scholarship; H.H. was supported by a Donnelly Centre Home Fellowship; A.J.B. was supported by a CIHR Postdoctoral Fellowship; M.A. was supported by an SNSF fellowship; and B.R. and J.A.C were supported by NSERC studentships. This research was supported by grants from CIHR (A.-C.G., J.M., and B.J.B.), Canada First Research Excellence Fund (B.J.B.), and the Simons Foundation (B.J.B.). B.J.B. holds the Banbury Chair of Medical Research at the University of Toronto.

## AUTHOR CONTRIBUTIONS

T.G.-P. conceptualized the project with input from B.J.B. and M.W. T.G.-P. and M.W. developed the CRISPR-based screens, with contributions from J.R., M.A., and J.M. U.B. analyzed CRISPR screen, SPAR-seq, RNA-seq, CLIP-seq, and proteomics data. J.R., A.-C.G., T.G.-P., and M.W. generated and analyzed AP-MS and BiolD data. H.H. and A.J.B. performed SPAR-seq experiments. B.R., J.D.E., T.G.-P., and M.W. analyzed Fbxw11 interactions. T.G.-P. and D.O. generated recombinant proteins and performed *in vitro* splicing, gel-shift assays, and splicing complex assembly assays. M.W. and T.G.-P. performed RNA affinity purification assays. J.A.C. performed Srm4 localization experiments. T.G.-P. and M.W. performed cell-based splicing assays and RT-PCR validation experiments. B.J.B. and T.G.-P. supervised the study and wrote the paper, with input from the other authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 20, 2018

Revised: August 27, 2018

Accepted: October 4, 2018

Published: November 1, 2018

## SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Hart et al. (2017); Parikshak et al. (2016).

## REFERENCES

- Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451.
- Berget, S.M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414.
- Boutz, P.L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., Jr., and Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**, 1252–1269.
- Buckanovich, R.J., Yang, Y.Y., and Darnell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16**, 1114–1122.
- C Yuen, R.K., Merico, D., Bookman, M., L Howe, J., Thiruvahindrapuram, B., Patel, R.V., Whitney, J., Deflaux, N., Bingham, J., Wang, Z., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611.
- Calarco, J.A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinis, D., van der Kooy, D., et al. (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* **138**, 898–910.
- Cvitkovic, I., and Jurica, M.S. (2013). Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.* **41**, D132–D141.
- de la Torre-Ubieta, L., Won, H., Stein, J.L., and Geschwind, D.H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191.
- Ebert, D.H., and Greenberg, M.E. (2013). Activity-dependent neuronal signaling and autism spectrum disorder. *Nature* **493**, 327–337.
- Evers, B., Jastrzebski, K., Heijmans, J.P.M., Grennum, W., Beijersbergen, R.L., and Bernards, R. (2016). CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633.
- Gehman, L.T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., Ares, M., Jr., Mody, I., and Black, D.L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711.
- Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., et al. (2017). Multilayered control of alternative splicing regulatory networks by transcription factors. *Mol. Cell* **65**, 539–553.e7.
- Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526.
- Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* **7**, 2719–2727.
- Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523.
- Jangi, M., and Sharp, P.A. (2014). Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498.
- Kuroyanagi, H., Kobayashi, T., Mitani, S., and Hagiwara, M. (2006). Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms *in vivo*. *Nat. Methods* **3**, 909–915.
- Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* **17**, 41–54.
- Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554.
- Li, Y.-I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* **25**, 1–13.
- Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**, 1836–1839.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Kraimer, A.R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J.* **18**, 4560–4570.
- Moore, M.J., Wang, Q., Kennedy, C.J., and Silver, P.A. (2010). An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142**, 625–636.
- Nakano, Y., Jahan, I., Bonde, G., Sun, X., Hildebrand, M.S., Engelhardt, J.F., Smith, R.J.H., Cornell, R.A., Fritzsche, B., and Bánfi, B. (2012). A mutation in the Srm4 gene causes alternative splicing defects and deafness in the Bronx waltzer mouse. *PLoS Genet.* **8**, e1002966.
- Nguyen, L.S., Kim, H.-G., Rosenfeld, J.A., Shen, Y., Gusella, J.F., Lacassie, Y., Layman, L.C., Shaffer, L.G., and Gécz, J. (2013). Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neurodevelopmental disorders. *Hum. Mol. Genet.* **22**, 1816–1825.
- Norris, A.D., Gao, S., Norris, M.L., Ray, D., Ramani, A.K., Fraser, A.G., Morris, Q., Hughes, T.R., Zhen, M., and Calarco, J.A. (2014). A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol. Cell* **54**, 946–959.
- Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148.
- Page-McCaw, P.S., Amonlirdviman, K., and Sharp, P.A. (1999). PUF60: a novel U2AF65-related splicing activity. *RNA* **5**, 1548–1560.

- Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., Ubieta, L.T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540, 423–427.
- Parras, A., Anta, H., Santos-Galindo, M., Swarup, V., Elorza, A., Nieto-González, J.L., Picó, S., Hernández, I.H., Díaz-Hernández, J.I., Belloc, E., et al. (2018). Autism-like phenotype and risk gene mRNA deadenylation by CPEB4 mis-splicing. *Nature* 560, 441–446.
- Quesnel-Vallièrès, M., Irimia, M., Cordes, S.P., and Blencowe, B.J. (2015). Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* 29, 746–759.
- Quesnel-Vallièrès, M., Dargaï, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J.Y., Wu, M., Sterne-Weiler, T., Nakagawa, S., Woodin, M.A., Blencowe, B.J., and Cordes, S.P. (2016). Misregulation of an activity-dependent splicing network as a common mechanism underlying autism spectrum disorders. *Mol. Cell* 64, 1023–1034.
- Raj, B., and Blencowe, B.J. (2015). Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* 87, 14–27.
- Raj, B., O'Hanlon, D., Vessey, J.P., Pan, Q., Ray, D., Buckley, N.J., Miller, F.D., and Blencowe, B.J. (2011). Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol. Cell* 43, 843–850.
- Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z.-Y., Chen, G.I., Easton, L.E., Ule, J., Gingras, A.-C., et al. (2014). A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* 56, 90–103.
- Sakashita, E., Tatsumi, S., Werner, D., Endo, H., and Mayeda, A. (2004). Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. *Mol. Cell Biol.* 24, 1174–1187.
- Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., Kidd, V.J., Manley, J.L., Lahti, J.M., and Reinberg, D. (2003). ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol. Cell Biol.* 23, 2981–2990.
- Scotti, M.M., and Swanson, M.S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32.
- Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16, 299–311.
- Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell* 19, 485–496.
- Singh, R.K., Xia, Z., Bland, C.S., Kalsotra, A., Scavuzzo, M.A., Curk, T., Ule, J., Li, W., and Cooper, T.A. (2014). Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. *Mol. Cell* 55, 592–603.
- Stern, D.A., and Berget, S.M. (1993). In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol. Cell Biol.* 13, 2677–2687.
- Suzuki, H., Chiba, T., Kobayashi, M., Takeuchi, M., Suzuki, T., Ichiyama, A., Ikenoue, T., Omata, M., Furuichi, K., and Tanaka, K. (1999). IkappaBalpha ubiquitination is catalyzed by an SCF-like complex containing Skp1, cullin-1, and two F-box/WD40-repeat proteins, betaTrCP1 and betaTrCP2. *Biochem. Biophys. Res. Commun.* 256, 127–132.
- Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallièrès, M., Permanyer, J., Sodaï, R., Marquez, Y., et al. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 27, 1759–1768.
- Tejedor, J.R., Papisaiakas, P., and Valcárcel, J. (2015). Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol. Cell* 57, 23–38.
- Ustianenko, D., Weyn-Vanhenryck, S.M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. *Wiley Interdiscip. Rev. RNA* 8, e1418.
- Venables, J.P., Koh, C.-S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E.R., Watier, V., Durand, M., et al. (2008). Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell Biol.* 28, 6033–6043.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858.
- Vuong, C.K., Black, D.L., and Zheng, S. (2016). The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* 17, 265–281.
- Wang, Z., Ballut, L., Barbosa, I., and Le Hir, H. (2018). Exon junction complexes can have distinct functional flavours to regulate specific splicing events. *Sci. Rep.* 8, 9509.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707.
- Wongpalee, S.P., Vashisht, A., Sharma, S., Chui, D., Wohlschlegel, J.A., and Black, D.L. (2016). Large-scale remodeling of a repressed exon ribonucleoprotein to an exon definition complex active for splicing. *eLife* 5, e19743.
- Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164, 29–44.
- Zhang, W.J., and Wu, J.Y. (1996). Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol. Cell Biol.* 16, 5400–5408.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
anti-Srrm4	<a href="#">Calarco et al., 2009</a>	
anti-Srsf2	Sigma-Aldrich	S4045; RRID:AB_477511
anti-Flag M2	Sigma-Aldrich	F1804; RRID:AB_262044
anti-Tubulin	Sigma-Aldrich	T6074; RRID:AB_477582
anti-Srsf11	ThermoFisher Scientific	PA5-37056; RRID:AB_2553885
<b>Bacterial and Virus Strains</b>		
One Shot Stbl3 Chemically Competent	ThermoFisher Scientific	C737303
Subcloning Efficiency DH5 $\alpha$ Competent Cells	ThermoFisher Scientific	18265017
Endura electrocompetent cells	Lucigen	60242-2
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
MG132	Sigma-Aldrich	M7449
Trichostatin A	Sigma-Aldrich	T8552
His-SRRM4	<a href="#">Calarco et al., 2009</a>	
His-GST-Srsf11	This study	
His-GST-Rnps1	This study	
<b>Critical Commercial Assays</b>		
OneStep RT-PCR Kit	QIAGEN	210210
SensiFAST SYBR No-ROX Kit	BIOLINE	BIO-98005
Maxima First Strand cDNA Synthesis Kit for RT-qPCR	ThermoFisher Scientific	K1671
<b>Deposited Data</b>		
RNA-Seq dataset to detect AS	<a href="#">Raj et al. (2014)</a>	GEO: GSE57278
Genome-wide CRISPR Screens	This study	GEO: GSE112599
SPAR-Seq	This study	GEO: GSE120164
SPAR-Seq	<a href="#">Han et al., 2017</a>	GEO: GSE80196
RNA-Seq datasets to detect AS	This study	GEO: GSE112600
CLIP-Seq	This study	GEO: GSE112598
Affinity-Purification Mass-Spectrometry	This study	MSV000082174 and PXD009226
Proximity ligation proteomics (BioID) data	This study	MSV000082169 and PXD009213
<b>Experimental Models: Cell Lines</b>		
Mouse: CGR8	ECACC	07032901
Mouse: N2A	ATCC	CCL-131
Mouse: N2A Flp-In (and derivatives thereof)	This study	
Mouse: N2A EGFP-Srrm4 endogenously tagged		
Human: 293T Cell Line		
<b>Oligonucleotides</b>		
See <a href="#">Table S7</a>		
<b>Software and Algorithms</b>		
MAGECK 0.5.6	<a href="#">Li et al., 2014</a>	<a href="https://sourceforge.net/p/mageck/wiki/Home/">https://sourceforge.net/p/mageck/wiki/Home/</a>
vast-tools 1.0	<a href="#">Tapial et al., 2017</a>	<a href="https://github.com/vastgroup/vast-tools">https://github.com/vastgroup/vast-tools</a>
FlowJo	TreeStar Inc.	
Prism 6	Graph.Pad Prism	
MS data storage and analysis: ProHits-LIMS		<a href="http://prohitsms.com/Prohits_download/list.php">http://prohitsms.com/Prohits_download/list.php</a>

## CONTACT FOR REAGENT AND RESOURCE SHARING

As Lead Contact, Benjamin J. Blencowe is responsible for all reagent and resource requests. Please contact Benjamin J. Blencowe at [b.blencowe@utoronto.ca](mailto:b.blencowe@utoronto.ca) with requests and inquiries.

## METHOD DETAILS

### Cell Culture

Mouse neuroblastoma (N2A) and 293T cells were grown in DMEM (high glucose; Sigma-Aldrich) supplemented with 10% FBS, sodium pyruvate, non-essential amino acids, and penicillin/streptomycin. Mouse embryonic stem cells (mESC) were grown in gelatin coated plates in GMEM supplemented with 100  $\mu$ M  $\beta$ -mercaptoethanol, 0.1 mM nonessential amino acids, 2 mM sodium pyruvate, 2.0 mM L-glutamine, 5,000 units/mL penicillin/streptomycin, 1000 units/mL recombinant mouse LIF (all Life Technologies) and 15% ES fetal calf serum (ATCC). Cells were maintained at sub-confluent conditions. Embryonic stem cell (ESC)-derived neurons were generated and cultured as described below. All cell lines were maintained at 37°C with 5% CO<sub>2</sub>. Cells were regularly monitored for absence of mycoplasma infection.

### Generation of a Neuro-2a (N2A) Cell Line Containing an FRT Recombination Site Using CRISPR-Cas9 (N2A Flp-In)

Integration of the flippase recognition target sequence (FRT) at the Rosa26 locus was accomplished using CRISPR-Cas9 editing. 500 ng of pX330 vector targeting the Rosa26 locus (cgccctctctctagaagac) and 500ng of a linearized homology directed repair (HDR) template containing an SV40 promoter upstream of an FRT site along with a Blastocidin resistance cassette and a rtTA3 cassette driven by a constitutive EF1- $\alpha$  promoter with homology arms for the Rosa26 locus (a kind gift from Monika Mis and Stéphane Angers, University of Toronto) was transfected into N2A cells in a 6 well format, as described previously. Following transfection and recovery, cells were passaged to a 10 cm dish and selected using 6  $\mu$ g/mL of Blastocidin S. To isolate clonal populations, N2A cells were harvested and passaged at limiting dilution (6 cells/mL) to eight, 96-well plates and allowed to grow in 6  $\mu$ g/mL Blastocidin S until visible colonies were formed. Colonies were first screened by extracting genomic DNA and using PCR reactions spanning the genomic and inserted regions, followed by Sanger sequencing to confirm insertion at the correct site. Of the clones showing successful integration, a single clone was selected for use based on similar growth characteristics and morphology to the parental line (N2A Flp-In).

### Generation of Stable Flp-In-N2A Cell Lines

Doxycycline-inducible stable Flp-in N2A cells were generated by transfecting 500 ng of pcDNA5/FRT/TRE-based plasmid with 2  $\mu$ g of plasmid encoding pOG44 recombinase in N2A Flp-In cells using Lipofectamine 2000 (Life Technologies) according to the manufacturer's recommendations. The pcDNA5/FRT based plasmid (Thermo Fisher Scientific, V6010-20) was modified to include a rtTA3-compatible Tet Response Element (TRE). Cell lines with stably integrated constructs were selected and maintained with 10  $\mu$ g/mL Blastocidin S and 200  $\mu$ g/mL Hygromycin B. Transgene expression was induced by addition of 2  $\mu$ g/mL Doxycycline.

### Endogenous N-terminal EGFP-tagging of Srrm4

Srrm4 was endogenously EGFP-tagged at its N-terminus in N2A cells using CRISPR-directed targeted genomic insertion. N2A cells were co-transfected with pX459 plasmid expressing Cas9 and an sgRNA targeting Srrm4 near the translation initiation site, as well as with a repair template, using Lipofectamine 2000 (Invitrogen). The repair template containing the EGFP sequence, flanked by 800-900 bp homology arms corresponding to the up- and down-stream sequences of the Srrm4 translation initiation site, was cloned into a pBluescript vector (with the hPGK driven Blastocidin resistance gene inserted). The homology arms were flanked with DNA sequences corresponding to the sgRNA target sequence so that the vector would be linearized in cells. The PAM sequence of the sgRNA sequence upstream of the ATG was mutated in the repair template so as not to be targeted by Cas9. For generating the knock-in lines the following sgRNA target sequence was used: 5'-GTTTCACGCGGACAGCGCCC(CGG)-3'.

Cells were selected for 3 days with Puromycin (2.5  $\mu$ g/mL) and Blastocidin S (10  $\mu$ g/mL) for pX459 and pBluescript transfection, respectively, and cultured for seven more days before single cell sorting for EGFP+ cells with a BD Influx FACS sorter (BD Biosciences). The insertion sequence in each sorted clone was confirmed at the genomic level using Sanger sequencing, and clones were tested for EGFP-Srrm4 expression using fluorescent microscopy, western blotting, and 3' RACE PCR.

### siRNA Transfections

N2A and 293T cell lines were transfected with 10 nM of siGENOME siRNA pools (Dharmacon) using RNAiMax (Life Technologies), as recommended by the manufacturer. A non-targeting siRNA pool (D-001206-13) was used as control. N2A and 293T cells were harvested 48 hours or 72 hours post transfection. For rescue experiments, knock-down was performed for 48 hours and ectopic expression of the 3xFlag-tagged transgene was induced for 24 hours (24 hours post siRNA transfection) prior to extraction of RNA or proteins for RT-PCR or immunoblot analysis, respectively.

### RNA Extraction and RT-PCR Assays

Total RNA was extracted from cells using the QIAGEN RNeasy Mini Kit as recommended by the manufacturer. To assess inclusion of alternative microexons, forward and reverse primers were designed to anneal to the constitutively included exons upstream and downstream of the alternative exon, respectively. RT-PCR assays were performed using the OneStep RT-PCR kit (QIAGEN) according to the manufacturer's recommendations. Reaction products were separated on 3%–4% agarose gels.

Percent Spliced In (PSI) values were calculated using ImageJ software. First the exon-included and exon-excluded band intensities were corrected by subtracting background. Then, intensity of the exon-included band was divided by the sum of the exon-included and exon-excluded bands. The result was multiplied by 100% to obtain the PSI value, which was rounded to the nearest whole integer.

### Quantitative RT-PCR

For quantitative RT-PCR (qRT-PCR), first-strand cDNAs were generated from 0.25–3  $\mu\text{g}$  of total RNA using the Maxima H Minus First Strand cDNA synthesis Kit (Thermo Scientific), as per the manufacturer's recommendations, and diluted to 2 ng/ $\mu\text{L}$ . qPCR reactions were performed in a volume of 10  $\mu\text{L}$  using 1  $\mu\text{L}$  of diluted cDNA, 500 nM primers and 5  $\mu\text{L}$  SensiFAST SYBR No-ROX Kit (BIOLINE), using a CFX96 Real-Time PCR Detection System (BIO-RAD) following the manufacturer's recommendations.

### Assessment of CRISPR/Cas9 Editing Efficiency by Endonuclease Cleavage (Surveyor) Assay

ON-target genomic editing efficiency was estimated using the Surveyor assay. N2A cells were transduced with multiple independent Cas9 and sgRNA-expressing viruses targeting *Srrm4* or control genes, respectively (gRNA sequences listed in [Table S7](#)). Cells were selected in Puromycin (2.5  $\mu\text{g}/\text{mL}$ ) for 72 hours and 8 days post-selection genomic DNA was extracted using the PureLink Genomic DNA Kit (Thermo Fisher Scientific), as per the manufacturer's recommendations. After amplification of the targeted loci by PCR ([Table S7](#)), PCR products were denatured and re-annealed to form heteroduplexes. The re-annealed PCR products were incubated with T7 endonuclease (NEB) for 20 minutes at 37°C, and the cleavage efficiency was determined by agarose gel electrophoresis. Estimated editing efficiency was calculated with the formula: percentage efficiency = 100%  $\times$  (1 - (1 - cleaved fraction)<sup>1/2</sup>).

### Splicing of Minigene Constructs

For minigene experiments we used the Exontrap Cloning Vector pET01. The general strategy was to introduce the genomic region including the alternative microexon and 300 intronic nucleotides of both the upstream and downstream flanking introns between the 5' and 3' exons of the base vector exons (from the insulin gene). Native microexon and flanking intronic sequences were PCR amplified from mouse genomic DNA and cloned into the vector using 5' Apal and 3' NotI restriction sites.

To monitor minigene splicing levels, minigenes were transfected using Lipofectamine 2000 (Thermo Fisher Scientific), as per the manufacturer's recommendations. 24 hours post-transfection, RNA was extracted and microexon splicing was monitored using primers annealing to the 5' and 3' exons of the base vector, as described above.

### Fluorescence Microscopy

N2A cells transduced with virus expressing shRNA targeting *Srrm4* or a non-targeting control ([Raj et al., 2011](#)) were plated onto coverslips in 6 well plates at 2–4  $\times$  10<sup>5</sup> cells/well. 24 hours after seeding, cells were fixed in 4% paraformaldehyde/PBS. Fixed cells were stored at 4°C overnight in PBS prior to permeabilization with PBS containing 0.1% Triton X-100 for 2 minutes at room temperature (RT). Cells were blocked in blocking buffer (PBS with 10mg/mL BSA, 1% normal goat serum and 0.2% Tween 20) for 60 minutes at 37°C, rinsed with PBS containing 0.2% Tween 20 and incubated with primary antibodies diluted in blocking buffer for 60 minutes at RT (antibody dilutions: rabbit anti-*Srrm4* at 1:2,000; mouse anti-Srsf2-Sigma S4045 at 1:2,000). Coverslips were washed three times with PBS containing 0.2% Tween 20 and incubated with secondary antibodies diluted in blocking buffer (anti-mouse IgG Alexa Fluor 488 at 1:1,500; anti-rabbit IgG Alexa Fluor 594 at 1:1,500) at 37°C for 60 minutes. Cells were then washed six times with PBS containing 0.2% Tween 20 prior to mounting with Permafluor (Thermo Fisher Scientific). For endogenous EGFP-*Srrm4* visualization, N2A cells were fixed, permeabilized and DNA was labeled with Hoechst 33258 (Sigma Aldrich) to indicate nuclei. Imaging was performed on a Leica DMI6000 B inverted microscope. Image capture was performed using Volocity 6 software (PerkinElmer).

### Co-immunoprecipitation Experiments

HA tagged constructs were transiently transfected into N2A Flp-In cells grown in 10cm plates using Lipofectamine 2000. After 24 hours, cells were treated with 2  $\mu\text{g}/\text{mL}$  Doxycycline and, 24 hours later, were harvested in cold phosphate buffered saline (PBS) and pellets were flash-frozen in liquid nitrogen. Frozen pellets were resuspended in 600  $\mu\text{L}$  of lysis buffer (100 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, 1% NP-40, and protease inhibitors). Lysates were subject to sonication (30 1 second pulses with 1 second in between at 30% power). For nuclease digestion, 10  $\mu\text{g}$  RNase A, 25 Units RNase T1 and 75 Units of benzonase were added and lysates were incubated at 37°C with shaking for 10 minutes. Lysates were cleared in a microcentrifuge by spinning at 15,000 g for 10 minutes at 4°C. Anti-flag immunoprecipitation was performed using magnetic Dynabeads protein G (Thermo Fisher Scientific) complexed with anti-Flag M2 antibody (Sigma-Aldrich). Antibody was incubated with lysates for 1 hours at 4°C followed by incubation with washed Dynabeads protein G for 3 hours at 4°C with rotation. Following incubation, complexes were washed 5 times with lysis buffer. Elution was performed in 1x Laemmli buffer at 95°C for 5 minutes.

### Immunoblotting

Cell lysates and co-immunoprecipitation samples were mixed with Laemmli buffer and heated at 95°C for 5 minutes, separated on variable percentage SDS-PAGE gels, and transferred to PVDF membranes. Blots were incubated overnight with the following primary antibodies at the specified dilutions in 5% milk:

- Mouse anti-Flag M2 (Sigma-Aldrich) at 1:3,000
- Mouse anti-Tubulin (Sigma-Aldrich) at 1:10,000
- Rabbit anti-Srsf11 (ThermoFisher Scientific) at 1:1,000
- Rabbit anti-Srrm4 (Calarco et al., 2009) at 1:5,000

### FLAG Affinity Purification Coupled with Mass Spectrometry (AP-MS) Sample Preparation

For FLAG AP-MS, cell pellets from two 150 mm plates, induced for 24 hours with 2 µg/mL Doxycycline, were lysed in ice cold TAP lysis buffer containing 50 mM HEPES-NaOH (pH 8.0), 100 mM NaCl, 2 mM EDTA, and 10% glycerol with freshly added 0.1% NP-40, 1 mM DTT, 1 mM PMSF, and Protease inhibitor cocktail (Sigma-Aldrich P8340, 1:500) at a 1:6 pellet weight to volume ratio (for 0.1 gr add 0.6 mL of lysis buffer). Resuspended pellets were frozen and thawed by placing the pellets on dry ice for 5 minutes and thawed in a 37°C water bath until only a small ice pellet remained. Samples were quickly moved back on to ice and sonicated with three 10 s bursts with 2 s rest at an amplitude of 35%. To solubilize chromatin and reduce the detection of interactions mediated by RNA or DNA, 1 µL of Benzonase nuclease (Sigma-Aldrich, E8263, 250U) was added to each tube and incubated for 30 minutes with rotation at 4°C. Lysates were cleared by centrifugation at 20,000 *rcf*. for 20 minutes at 4°C and the lysates transferred to tubes containing 25 µL of 50% magnetic anti-FLAG M2 beads slurry (Sigma-Aldrich, M8823) pre-washed in lysis buffer. FLAG immunoprecipitation was performed for 3 hours at 4°C with rotation.

After incubation, beads were pelleted by centrifugation (1,000 rpm for 1 minute) and magnetized to aspirate the unbound lysate. The beads were then demagnetized and washed with 1 mL of lysis buffer and the total volume (with beads) transferred to a new tube. Beads were washed once more with 1 mL of lysis buffer followed by one wash with 50 mM ammonium bicarbonate (ABC) at pH 8. All wash steps were performed on ice using cold lysis buffer and ABC. After the final wash, any residual ABC was aspirated from the beads, and 1 µg of trypsin (Sigma-Aldrich, T6567), in 10 µL of ABC was added to each tube. The samples were incubated at 37°C overnight with rotation. After the initial incubation, the beads were magnetized, and the supernatant transferred to a new tube. Another 250 ng of trypsin was added to each supernatant in 5 µL of ABC (total volume of 15 µL) and further digested with rotation for another 4 hours at 37°C. Samples were treated with formic acid to a final concentration of 2.5% and dried in a centrifugal evaporator.

### BioID Sample Preparation for MS

For BioID samples, cell pellets from one 150 mm plate, induced for 24 hours with 2 µg/mL Doxycycline and 50 µM biotin, were lysed in ice cold RIPA buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA and 0.1% SDS with freshly added 0.5% sodium deoxycholate and protease inhibitor cocktail (Sigma-Aldrich P8340, 1:500) at a 1:10 pellet weight to volume ratio (i.e., 0.1 g in 1.0 mL lysis buffer).

The lysates were sonicated, treated with Benzonase, and centrifuged as described above for FLAG AP-MS. After centrifugation, cleared lysates were added to 60 µL of streptavidin-Sepharose bead slurry (GE Healthcare, Cat 17-5113-01), pre-washed three times with 1 mL of lysis buffer by pelleting the beads with gentle centrifugation (6000 rpm, 30 seconds), and aspiration of the supernatant before the next wash. Biotinylated proteins were captured on beads for 3 hours at 4°C with rotation.

After affinity purification, streptavidin beads were next pelleted (6,000 rpm, 30 seconds), and the supernatant removed. For the Srrm4, Srrm3, Ptbp1, and Ptbp2 samples, the streptavidin beads were resuspended in RIPA lysis buffer and transferred to a new microcentrifuge tube. Beads were washed a second time with RIPA buffer, centrifuged (6,000 rpm, 30 s) and resuspended in RIPA lysis buffer. Subsequent washes were performed in a similar manner, twice with TAP buffer (see above), and twice with 50 mM ammonium bicarbonate (ABC, pH 8). For Rnps1, Rbfox2, and Srsf11 samples, a more stringent buffer was used following streptavidin bead incubation, consisting of 2% SDS and 50 mM Tris pH 7.5 in water, prior to transfer of the beads to a new tube. Two additional washes were performed, twice with RIPA lysis buffer and three times with ABC buffer, to remove detergent. Control samples consisting of BirA<sup>\*</sup>-FLAG GFP, and BirA<sup>\*</sup>-FLAG alone, were prepared for both lysis conditions to control for non-specific biotinylation. For all samples, after the final wash residual ABC was pipetted off the beads and replaced with 60 µL of ABC containing 1 µg of trypsin (Sigma-Aldrich, T6567). Samples were incubated with trypsin overnight with rotation at 37°C.

Beads were next pelleted by centrifugation (6,000 rpm, 30 seconds) and the supernatant removed to a new tube. To remove remaining peptides, beads were rinsed with an additional 60 µL of ABC with the lysate added to the supernatant (total of ~120 µL). 0.5 µg of trypsin was added to each supernatant and incubated for another 4 hours at 37°C with rotation. Finally, samples were treated with formic acid to a final concentration of 2.5% and dried in a centrifugal evaporator.

### Mass Spectrometry Data Acquisition

AP-MS and BioID samples were analyzed by mass spectrometry in at least two biological replicates. Digested peptides were dissolved in 5% formic acid in a volume in which 6 µL contained purified material from half of a 150 mm plate for FLAG AP-MS,

or a quarter of a 150 mm plate for BioID. For each sample, 5  $\mu$ L was directly loaded at 800 nL/minute onto a 15 cm 100  $\mu$ m ID emitter tip packed in-house with 3.5  $\mu$ m Reprosil C18 (Dr. Maisch GmbH, Germany). The peptides were eluted from the column at 400 nL/min over a 90 minutes gradient generated by a 425 NanoLC (Eksigent, Redwood, CA) and analyzed on a TripleTOF<sup>TM</sup> 6600 instrument (AB SCIEX, Concord, Ontario, Canada). The gradient started at 2% acetonitrile with 0.1% formic acid and increased to 35% acetonitrile over 90 minutes followed by 15 minutes at 80% acetonitrile, and then 15 min at 2% acetonitrile for a total of 120 minutes. To minimize carryover between each sample, the analytical column was flushed for 1 hour at 1500 nL/min with an alternating sawtooth gradient from 35% acetonitrile to 80% acetonitrile, holding each gradient concentration for 5 min. Analytical column and instrument performance were verified after each sample by analyzing a 30 fmol bovine serum albumin (BSA) tryptic peptide digest with a 60 fmol  $\alpha$ -casein tryptic digest with a short 30 minutes gradient. MS mass calibration was performed on BSA reference ions between each sample. Acquisition was in Data Dependent mode and consisted of one 250 ms MS1 TOF survey scan from 400–1250 Da followed by twenty 100 ms MS2 candidate ion scans from 100–2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 4+ which exceeded a threshold of 200 counts per second were selected for fragmentation, and former precursors were excluded for 10 s after 1 occurrence.

### Neuronal differentiation Trichostatin A treatment

Neuronal cultures grown in plates coated with poly(D)-Lysine and laminin were maintained for 10 days *in vitro* and then treated with 100 nM Trichostatin A (Sigma-Aldrich) for 6 hours. RNA extraction and qRT-PCR was performed as described above.

### Expression and Purification of Recombinant Srsf11 and Rnps1

Recombinant 6xHIS-SRRM4 was produced as described previously (Calarco et al., 2009; Figure S6J). Recombinant baculovirus expressing either 6xHIS-GST-Srsf11 or 6xHIS-GST-Rnps1 were generated using the Flashbac system and protocols (Mirusbio). Sf9 (*Spodoptera frugiperda*) cells were transfected with Flashbac DNA and transfer vectors (pOET1) to produce recombinant virus (P0). The P0 recombinant virus was used to infect Sf9 cells producing a high titer recombinant virus stock (P1). The P1 stock was titered and used to infect Hi5 (*Trichoplusia ni*) cells for protein production. Approximately  $2 \times 10^6$ /mL logarithmic, suspension Hi5 cells were infected with recombinant virus at an MOI of 10. Infected cells were incubated at 27°C for 48 hours. Cells were then harvested by centrifugation, washed with cold PBS and subsequent pellets snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until protein purification. Sf9 cells were grown and maintained in Sf900 III SFM media (ThermoFisher Scientific) at 27°C. Hi5 cells were adapted to suspension culture and maintained in suspension at 27°C in serum-free media (Express Five media, ThermoFisher Scientific).

In order to purify recombinant proteins, cell pellets were lysed in Tris Lysis buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 10% glycerol, 0.04% v/v  $\beta$ -mercaptoethanol, 1.5 mM PMSF, protease and phosphatase inhibitors) using 10 mL/g pellet, sonicated with  $3 \times 15$  seconds bursts followed by a 10-minute incubation on ice. Samples were centrifuged at 20,000 g for 20 minutes at 4°C and the supernatant was collected. This was passed through a column containing Ni-NTA agarose beads (ThermoFisher Scientific) three times at 4°C. Columns were washed with 20 bed volumes of Tris Lysis buffer and eluted with Tris Lysis buffer containing 250 mM Imidazole, and collected in  $\sim 500$   $\mu$ L fractions. Recombinant protein containing fractions were pooled and bound to a glutathione Sepharose 4B beads (GE Healthcare) for 2 hours at 4°C. Beads were washed and eluted in Tris Lysis 0.1% NP-40 buffer containing 10 mM reduced glutathione (Sigma). Purified recombinant proteins were analyzed by Coomassie-stained SDS-polyacrylamide gel electrophoresis (Figure S6J).

### In Vitro Transcription

*In vitro* transcribed RNAs were synthesized from double-stranded linear DNA templates amplified from minigene constructs (Exon-trap Cloning Vector pET01) by incorporating a T7 promoter sequence upstream of the 5' end of the forward primer. Full sequence added was: 5'-GAAAT(TAATACGACTCACTATAG)GGAGA, with the minimum T7 promoter in parentheses. The upstream sequence was added to aid T7 binding and the downstream nucleotides to aid transcription efficiency. The full primer sequences, including gene-specific sequence used for probe amplification were:

Probe Daam1 (assembly gel) –Fwd GCTGCTGTGGTTTCTGAATTG  
 Probe Daam1 (assembly gel) –Rev TAAGGGTTATATAGTAGCTTGC  
 Probe Sh3glb1 (assembly gel) –Fwd CTGACTAAGTTGTAGGAAATG  
 Probe Sh3glb1 (assembly gel) –Rev CTGATCCAGGCGTTTAAGGTTTAC

The amplified DNA fragments were purified on a 1% agarose gel for use in the *in vitro* transcription reaction. *In vitro* transcription was performed using the MEGAshortscript T7 Transcription Kit (Life technologies) as per the manufacturer's recommendations. For a 20  $\mu$ L reaction: 2  $\mu$ L of 10x reaction buffer, 2  $\mu$ L of 100 mM DTT, 2  $\mu$ L of Ribolock RNase inhibitor, 400 ng of template DNA, 2  $\mu$ L of T7 polymerase enzyme and 0.5 mM nucleotides were mixed with the appropriate amount of nuclease-free water and incubated for 4 hours at 37°C. For radiolabelling reactions, the nucleotide consisted of 1  $\mu$ L of a UTP-reduced nucleotide mix (10 mM AGC, 0.5 mM U) and 5  $\mu$ L 32P-UTP (3000 curies /mmol; 10  $\mu$ Ci/ $\mu$ L). Following transcription, the reaction mix was treated with 2.5  $\mu$ L of TURBO DNase, phenol/chloroform extracted, and resuspended in 40  $\mu$ L of nuclease-free water.

### **In Vitro Splicing Assays**

Preparation of whole cell splicing extracts and purification of recombinant SRRM4 has been previously described in detail (Calarco et al., 2009). *In vitro* splicing assays were performed in a volume of 20  $\mu$ L contained 1.5 mM ATP, 5 mM creatine phosphate, 5 mM DTT, 3 mM MgCl<sub>2</sub>, 2.6% PVA, 30 units of RiboLock RNase inhibitor (Thermo Scientific), 40 ng of splicing substrate and 12  $\mu$ L of splicing extract, with or without the addition of recombinant proteins. Reactions were incubated at 30°C for one hour. RNA was extracted using TRI Reagent (Sigma) and then resuspended in 10  $\mu$ L of DEPC-treated water. Spliced products were amplified by RT-PCR assays using 2  $\mu$ L of the recovered RNA and primers specific for the constitutive exons. RT-PCR products were resolved on a 3% agarose gel.

### **Electrophoretic Mobility Shift Assays (EMSA)**

20  $\mu$ L binding reactions used for gel shift assays contained 12 mM HEPES-KOH pH 7.9, 60 mM KCl, 0.12 mM EDTA, 6% glycerol, 0.02% NP-40, 1 mM DTT, and 100 ng/ $\mu$ L of tRNA. Recombinant GST-tagged Srsf11 and His-tagged SRRM4 were added at indicated concentrations and incubated for 10 minutes at 30°C. 1  $\mu$ L of radiolabeled RNA was then added (approximate final concentration of 0.1  $\mu$ M) and each sample was incubated for 15 minutes at 30°C to allow complex formation. 4  $\mu$ L of gel loading dye (50% glycerol, 62.5 mM EDTA pH 8.0, and bromophenol blue) was added and the samples were separated for 4-5 hours on a 4% native polyacrylamide gel in 0.5x TBE buffer. Gels were transferred to Whatman filter paper, dried and imaged using a Typhoon scanner (GE Healthcare).

### **Splicing Complexes Assembly Assays**

Splicing reactions were performed as described above. The reactions were incubated at 30°C for 20 minutes and then 1  $\mu$ L Heparin (50 mg/mL) was added to the reactions which were incubated at room temperature for 10 more minutes. 3  $\mu$ L of 50% Glycerol was added to the reaction mixture which was loaded on a native acrylamide/agarose composite (4% acrylamide / 0.05% bis-acrylamide 80:1 / 0.5% agarose) gel (20 x 20 cm). The reactions were resolved for 4 hours at 200 V in 0.5x TBE buffer. Gels were transferred to Whatman filter paper, dried and imaged using a Typhoon scanner (GE Healthcare).

### **Affinity Purification of Splicing Complexes**

Affinity selection of splicing complexes were performed with 20  $\mu$ L *in vitro* splicing reaction using 40 ng biotinylated RNA probe (118 nt) containing Daam1 microexon and part of surrounding introns. After 30 minutes of assembly at 30°C, reaction were diluted on ice with 280  $\mu$ L buffer E (20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 200  $\mu$ M EDTA, 1 mM DTT, 10% Glycerol with protease inhibitor cocktail), and affinity purified using Dynabead MyOne Streptavidin C1 beads (Invitrogen). Washed Streptavidin C1 beads were incubated with diluted splicing reaction for 2 hours at 4°C with rotation. Beads were washed 4 times (5 minutes/each with rotation at 4°C) with wash buffer (Buffer E + 0.1% Triton X-100). RNA components associated with beads were purified using TRI reagent (Sigma) following manufacturer's suggestions, and subsequently reverse-transcribed into cDNA using Maxima H-minus first strand cDNA synthesis kit (Thermo Fisher Scientific) with random hexamers. U1 or U2 snRNA and RNA probe qPCR were performed using SensiFAST SYBR No-ROX Kit (BIOLINE) as described above. Relative binding of U1 or U2 snRNA was calculated as fold enrichment comparing to no-probe controls pulled down using streptavidin beads and normalized to probe purification efficiency.

### **Individual-nucleotide Resolution Cross-linking and Immunoprecipitation (iCLIP-Seq)**

iCLIP was performed as described previously (Han et al., 2017). Srsf11, Rnps1 or GFP (as negative control) were immunoprecipitated from N2A Flp-In cells induced for 24 hours with 2  $\mu$ g/mL doxycycline to express Flag-tagged Srsf11, Rnps1 or GFP, respectively. The iCLIP of Flag-tagged Srrm4 was performed in mESC-derived neurons. First, we generated CGR8 mESC single cell clones expressing Flag-tagged Srrm4 using the PiggyBac transposase system. Then mESC cells were differentiated into cortical glutamatergic neurons and plated onto 15 cm plates coated with poly-D-Lysine and laminin (~20 million cells per plate). After 8 days Flag-Srrm4 or Flag-EGFP expression was induced with 2  $\mu$ g/ml doxycycline for 48 hours.

Cells were crosslinked (0.15 J/cm<sup>2</sup>) at 254 nm with a Stratalinker 1800. Three independent replicates were used for generating iCLIP samples. Lysates generated from the crosslinked cells were treated with Turbo DNase (Ambion) and diluted RNase I (Ambion) for 5 minutes at 37°C to digest the genomic DNA and trim the RNA to short fragments of an optimal size range. RNA-protein complexes were immunoprecipitated using 100  $\mu$ L of protein G Dynabeads (Life Technologies) and 10  $\mu$ g of anti-Flag (Sigma) antibody. Following stringent high salt washes, the immunoprecipitated RNA was 3' end dephosphorylated and a pre-adenylated adaptor was ligated at the 3' end for adaptor ligation. After additional washes, the immunoprecipitated RNA was 5' end-labeled using radioactive <sup>32</sup>P isotopes. The immunoprecipitated complexes were separated with SDS-PAGE and transferred to a nitrocellulose membrane (Protran). RNA was recovered by digesting proteins using proteinase K and subsequently reverse transcribed into cDNA. The reverse transcription primers include barcode sequences to enable multiplexing and a BamHI restriction site. The cDNA was size selected (low: 70 to 85 nt, middle: 85 to 110 nt, and high: 110 to 180 nt), circularized to add the adaptor to the 5' end, digested at the internal BamHI site, and then PCR amplified using AccuPrime SuperMix I (Life Technologies). The final PCR libraries were purified by agarose gel electrophoresis using gel extraction columns (QIAGEN), eluted DNA was mixed at a ratio of 1:5:5 from the low, middle, and high fractions and submitted for sequencing.

For Flag-Srsf11 the barcoded primers used were: Rt9clip, Rt13clip and Rt16clip. For Flag-Rnps1 the barcoded primers used were: Rt2clip, Rt9clip and Rt13clip. For Flag-Srrm4 the barcoded primers used were: Rt1clip and Rt9clip. For Flag-GFP the barcoded primers used were: Rt1clip, Rt10clip and Rt14clip (as control for the Flag-Srsf11 iCLIP) and Rt1clip, Rt6clip and Rt16clip (as control for the Flag-Rnps1 iCLIP).

Rt1clip: /5Phos/NNAACNNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt2clip: /5Phos/NNACAANNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt6clip: /5Phos/NNCCGGNNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt9clip: /5Phos/NGCCANNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt10clip: /5Phos/NNGACNNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt13clip: /5Phos/NNTCGNNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt14clip: /5Phos/NNTGCCNNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC;  
 Rt16clip: /5Phos/NNTTAANNNAGATCGGAAGAGCGTCGTGgataCTGAACCGC.

### RNA-Sequencing

RNA was extracted from N2A cells treated for 48 hours with siRNA siGENOME SMARTpools (Dharmacon) using RNeasy Mini Kit (QIAGEN) as described above. Extracted total RNA was submitted to the Donnelly Sequencing Centre for further processing. Total RNA was quantified using fluorescent chemistry contained in the Qubit RNA BR Assay Kit (Cat# Q10211, Thermo Fisher Scientific) and diluted to 1 ng/ $\mu$ L. 1  $\mu$ L was used to assess RNA Integrity Number (RIN) using the Agilent Bioanalyzer RNA 6000 Pico kit (Cat # 5067-1513, Agilent). Lowest RIN was 9.5, median RIN was 9.8.

In the subsequent library preparation, 1000 ng RNA per sample was used. To obtain larger insert sizes (540 nt on average after size selection), Illumina's TruSeq Stranded mRNA sample preparation protocol was modified as follows: 1) fragmentation time and temperature was reduced from 8 minutes at 98°C to 2 minutes at 80°C. Average size distribution prior to library preparation was 232 nt versus 740 nt, respectively after the final library was generated. The samples were loaded onto a 2% agarose gel (Cat # 1613101, BioRad) and run for 1 hour at 100 V. The libraries were size selected between 400 and 600 bp. Excised gel fragments were purified using the QIAGEN MinElute Gel Extraction kit (Cat # 28604, QIAGEN).

The rest of the protocol was adapted as described in Illumina's TruSeq Stranded mRNA sample preparation guide (Part# 15031047 Rev. E, Illumina). 1  $\mu$ L top stock of each purified final library was run on an Agilent Bioanalyzer dsDNA High Sensitivity chip (Cat # 5067-4626, Agilent). The libraries were quantified using Kapa Universal qPCR Master Mix (Cat # KK4923, Roche) and Qubit dsDNA HS Assay Kit (Cat # Q32854, ThermoFisher Scientific) and were pooled at equimolar ratios after size-adjustment. The final pool was run on an Agilent Bioanalyzer dsDNA High Sensitivity chip and quantified using the NEBNext Library Quant Kit for Illumina (Cat # E7630L, NEB). The quantified pool was hybridized at a final concentration of 2.1 pM and sequenced paired-end (PE) on the Illumina NextSeq500 platform using high-output v2 300c chemistry. 525 M PF clusters and 482 M PF and > Q30 clusters were obtained for 12 samples (avg. 87 M PE reads/sample).

### SPAR-Seq

Systematic Parallel Analysis of endogenous RNA regulation coupled to barcode Sequencing (SPAR-Seq) was performed as previously described (Han et al., 2017). In brief, following siRNA knockdowns, a multiplex RT-PCR assay was applied to simultaneously amplify and assess 25 alternative splicing (AS) events in a single reaction. These included 13 microexons and 12 longer neural-regulated cassette exons, of which 17 are misregulated in ASD. The multiplex RT-PCR reaction was carried out using the OneStep RT-PCR kit (QIAGEN). To multiplex all samples for sequencing, unique, dual-index barcodes were added to the amplicons in the second PCR reaction, which was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific). The resulting libraries were pooled and sequenced using an Illumina MiSeq at the Donnelly Sequencing Centre.

### Engineering Bichromatic Microexon Splicing Reporters

For CRISPR screens we generated bichromatic reporters that include conserved microexons from the murine Shank2 (ENSE00002478436/ENSMUSE00001006087) and Mef2d (ENSE00001054660/ENSMUSE00000673645) genes. The general strategy for the construction of the bichromatic reporters was to amplify the alternative microexon, its flanking introns, the upstream constitutive exons and 20-50 nt of the downstream constitutive exon from mouse genomic DNA. The forward primers included attB1 sites required for the cloning of products into pDONR<sup>TM</sup>221 (entry vector) using Gateway recombination cloning technology (Life Technologies<sup>TM</sup>) followed by an SV40 nuclear localization signal (NLS). The reverse primer included 40 nt of sequence annealing to EGFP sequence. In parallel, an EGFP-mCherry expression cassette was amplified from vectors containing this bichromatic cassette (Norris et al., 2014) using a reverse primer that included attB2 sites required for the cloning of products into pDONR<sup>TM</sup>221 (entry vector) using Gateway recombination cloning technology. The two PCR products were joined into a single fragment using Gibson Assembly (NEB) via the overlapping ends and the joint fragment was further amplified using the forward primer of the mini-gene fragment and the reverse primer of the bichromatic cassette. The PCR products contained both attB1 and attB2 sites and were cloned into the pDONR<sup>TM</sup>221 (entry vector) using Gateway recombination cloning technology. After selection of successful colonies,

the Shank2 and Mef2d microexons were mutated via the insertion of an extra nucleotide (TGGACAAAAG and ACTGAGGACCATT TAGATCTG, respectively; inserted nucleotide is underlined) using the Q5 Site-Directed Mutagenesis Kit (NEB) such that microexon inclusion would result in a change in the reading frame of the bichromatic cassette. Edited sequences were subcloned into a customized pcDNA5-based Gateway compatible vector (Life Technologies™) that contained 8x TET response elements (TRE) followed by a miniCMV promoter, Kozak sequence, 3xFlag-tag sequence and attR sites. The pcDNA5 vectors containing the bichromatic microexon reporters were subsequently used for the generation of N2A Flp-In lines.

The RT-PCR amplified products of the microexon reporters were confirmed by Sanger sequencing. The Shank2 reporter, in addition to the expected product (microexon sequence: 5'-TGGACAAAAG-3') produced an additional isoform generated by the use of an alternative 3' splice site located 12 nucleotides upstream (microexon sequence; 5'-GCCTCTCACTAG TGGACAAAAG-3'). Use of either of the two 3' splice sites does not impact the downstream reading frame and thus does not affect fluorescent protein expression upon microexon inclusion.

### Cloning of sgRNAs into LentiCRISPRv2 Vectors for Validation

LentiCRISPRv2 (Addgene plasmid # 52961), was modified to generate GCCA and AAAC overhangs after digestion with BsmBI (NEB) restriction enzyme and phosphatase. 20 nt guide sequences were selected from an independent mouse genome-wide library that was designed to have reduced off-target effects and improved on-target activity (J.M., unpublished data).

The forward and reverse oligos with 4 nucleotide overhangs (ACCG for forward and AAAC for reverse oligo) were annealed and ligated into BsmBI digested and rSAP (NEB) treated lentiCRISPRv2 vector using T4 DNA Ligase (NEB) for 90 minutes at 25°C followed by heat inactivation for 10 minutes at 65°C. The ligation reaction was transformed into Stbl3 competent cells (Thermo Fisher Scientific) by heat shock, as per the manufacturer's recommendations. Single colonies were picked for plasmid isolation. Plasmids were analyzed by Sanger sequencing to confirm successful cloning of guides.

### Virus Production and MOI Determination

For all CRISPR screens we used the mouse Brie CRISPR knockout pooled library (Addgene #73633)(Doench et al., 2016). For library virus production, 8 million HEK293T cells were seeded per 15 cm plate and the cells were transfected 24 hours later with a mix of 8 µg lentiCRISPRv2 vector containing the BRIE library, 4.8 µg packaging vector psPAX2, 3.2 µg envelope vector pMD2.G, 48 µL X-treme Gene transfection reagent (Roche) and 700 µL Opti-MEM medium (Life Technologies). 24 hours post-transfection the medium was changed to serum-free, high-BSA growth medium (DMEM, 1.1 g/100 mL BSA, 1% Penicillin/Streptomycin). For the validation experiments, the same protocol was followed but instead of 8 million cells, 450,000 cells were plated per single well of a 6-well plate. The cells were transfected 24 hours later with a mix of 1 µg lentiCRISPRv2 vector, 0.6 µg packaging vector psPAX2, 0.4 µg envelope vector pMD2.G, 6 µL X-treme Gene transfection reagent (Roche) and 100 µL Opti-MEM medium (Life Technologies). The virus-containing medium was harvested 48 hours after transfection, centrifuged at 1,500 rpm for 5 minutes, aliquoted and frozen at -80°C.

In order to determine the volume of the BRIE library virus required to obtain a multiplicity of infection (MOI) of 0.3, 4 million N2A cells were infected with a titration of the lentiviral gRNA library along with polybrene (8 µg/mL). After 24 hours, growth medium was replaced with medium containing Puromycin (2.5 µg/ml) and cells were incubated for an additional 72 hours. The MOI of the titrated virus was determined 96 hours post-infection by comparing percent survival of infected and selected cells to an infected but non-selected control.

### Test Library Screen

To determine whether the N2A Flp-In cells expressing microexon reporters were sensitive for pooled CRISPR screens an initial test screen was performed. Eight lentiCRISPRv2 vectors expressing guides targeting LacZ, Luciferase, EGFP (two independent guides) and Srrm4 (four independent guides) were mixed at an equimolar ratio and this small library was used for the generation of virus as described above. N2A Flp-In cells expressing the Shank2 microexon reporter were transduced at an MOI of 0.3 in the presence of 8 µg/mL polybrene. 24 hours after infection, the medium was replaced with fresh medium containing Puromycin (2.5 µg/mL) and cells were incubated for an additional 72 hours before the cells were passaged again. 10 days post infection, the cells were harvested and resuspended in sorting buffer (Hanks Balanced Salt Solution, 25 mM HEPES pH 7.0, 2 mM EDTA, 1% Fetal Bovine Serum) at a concentration of 5 million cells per mL. The cells were passed through a nylon mesh with a pore size of 40 µm to eliminate large aggregates and sorted based on the relative EGFP:mCherry expression as indicated in Figure 1D using the Flow Cytometry core facility at the University of Toronto and a BD Influx jet-in-air cell sorter. The experiment was performed twice independently. gDNA and RNA was extracted from the unsorted cells as well as the sorted populations using the PureLink Genomic DNA Kit (Thermo Fisher Scientific) and the total RNA purification kit (Norgen), respectively. The purified RNA was used for RT-PCR and qRT-PCR experiments as described in the relevant sections. The gDNA was used for the generation of libraries for high-throughput sequencing and the libraries were sequenced as described below.

### Genome-wide CRISPR-Cas9 Screens

80 million N2A Flp-In cells expressing bichromatic splicing reporters (either for Shank2 or Mef2d microexons) were infected with the lentiviral Brie library (78,637 gRNAs) at an MOI ~0.3, in the presence of 8 µg/mL polybrene, such that every sgRNA was represented in approximately 300 cells. 24 hours after infection, the medium was replaced with fresh medium containing Puromycin (2.5 µg/mL) and

cells were incubated for an additional 72 hours. After selection cells were split such that each sgRNA would be represented by an average of 300 cells in the population (i.e., 24 million cells) and passaged every three days.

For cell sorting, 24 million cells were seeded and the next day 1  $\mu\text{g}/\text{mL}$  doxycycline was added to induce bichromatic reporter expression. 24 hours later the cells were harvested and resuspended in sorting buffer (Hanks Balanced Salt Solution, 25 mM HEPES pH 7.0, 1 mM EDTA, 1% Albumin) at a concentration of 5 million cells per mL. The cells were passed through a nylon mesh with a pore size of 40  $\mu\text{m}$  to eliminate large aggregates. Filtered cells were sorted based on the relative EGFP:mCherry expression using Flow Cytometry core facilities at the University of Toronto and Lunenfeld-Tanenbaum Research Institute (LTRI) using either a BD Influx jet-in-air cell sorter or a Beckman Coulter MoFlo Astrios Cell Sorter, respectively. For each replicate, 300,000–500,000 cells with the highest or lowest 2% EGFP:mCherry ratio were collected. In addition, 24 million cells were also collected prior to sorting as a reference point for comparing enrichment in the sorted populations. For the 30% sorting scheme, the same procedure was followed except that the 30% of cells with the highest or lowest EGFP:mCherry ratio were sorted. Similarly, for the EGFP:Srrm4 expression-CRISPR screen, 24 million cells were also collected prior to sorting as a reference point and the 2% of the cells (300,000–500,000 cells) with the highest or lowest EGFP expression were sorted.

Genomic DNA was extracted from the cell pellets of unsorted samples using the QIAamp Blood Maxi Kit (QIAGEN) while gDNA from sorted populations was purified with the Midi Kit as per the manufacturer's recommendations. Genomic DNA was precipitated using ethanol and sodium chloride, and resuspended in Buffer EB (10 mM Tris-HCl, pH 7.5). gRNA inserts were amplified via PCR using primers harboring Illumina TruSeq adapters with i5 and i7 barcodes, following a two-step PCR approach (Table S7).

In the first step, a total of 50  $\mu\text{g}$  of gDNA was subjected to PCR (25 cycles, temperature of annealing ( $T_a$ ) = 65°C) for enrichment of gRNA cassettes using NEBNext Ultra II Q5 polymerase and staggered primers annealing to the end of the U6 promoter and the beginning of the tracrRNA. For the 2% highest or lowest expressing sorted cell populations 10  $\mu\text{g}$  of gDNA was used. Subsequently, the individual PCR reactions were pooled and 50  $\mu\text{L}$  of the first step PCR product was loaded onto a 2% agarose gel (Cat # 1613101, BioRad) and run for 1.5 hour at 100V. PCR products of 200–230 bp were selected and the excised gel fragments were purified using the QIAGEN MinElute Gel Extraction kit (Cat # 28604, QIAGEN). 1/6th of the purified PCR products were subjected to 2<sup>nd</sup> step PCR (10 cycles,  $T_a$  = 62°C) using NEBNext Ultra II Q5 polymerase and primers harboring Illumina TruSeq adapters with i5 and i7 indices to generate barcoded sequencing libraries ready for Illumina sequencing (Table S7). The PCR amplicons were loaded onto a 2% agarose gel (Cat # 1613101, BioRad) and run for 1 hour at 100V. The libraries were size-selected at 270 bp and 300 bp, and the excised gel fragments were purified using QIAGEN MinElute Gel Extraction kit (Cat # 28604, QIAGEN).

The purified libraries were analyzed by the Donnelly Sequencing Centre for quality control, pooling and sequencing. Size confirmation analysis for each sample was performed on an Agilent Bioanalyzer dsDNA High Sensitivity chip. Amplicons were quantified using Kapa Universal qPCR Master Mix (Cat # KK4923, Roche) and Qubit dsDNA HS Assay Kit (Cat # Q32854, Thermo Fisher Scientific) and pooled at varied ratios by molarity after size-adjustment. To mitigate the effects of index hopping, pooling was performed just prior to sequencing. The final pool was run on an Agilent Bioanalyzer dsDNA High Sensitivity chip and quantified using NEBNext Library Quant Kit for Illumina (Cat # E7630L, NEB). The quantified pool was hybridized at a final concentration of 2.15 pM and sequenced by generating single-end (SE) reads on the NextSeq500 platform using high-output v2 75c chemistry. To circumvent problems caused by low-diversity sequencing on this platform, we used a custom sequencing protocol which included 20 “dark cycles” (i.e., base additions without imaging) in addition to the staggered library design (i.e., primers with staggered regions were used for the generation of sequencing libraries so as to maintain sequence diversity across the flow-cell; see Table S7), followed by 26 imaged cycles (the first 20 bp of which were the guide sequence) followed by two 8-bp index reads.

### Validation of Hits Identified by the Genome-wide CRISPR-Cas9 Screens Using Flow Cytometry

N2A Flp-In bichromatic reporter lines were transduced with lentivirus expressing Cas9 along with sgRNAs and selected with 2.5  $\mu\text{g}/\text{mL}$  Puromycin for 3 days. For validation experiments we interrogated 39 genes with two independent sgRNAs for each gene. The target sequences of all sgRNAs are indicated in Table S7. 7–10 days post selection cells were induced with 2  $\mu\text{g}/\text{mL}$  doxycycline for 24 hr, dissociated with trypsin and fixed in HBSS buffer with a final concentration of 2% formaldehyde and 1% BSA at 4°C for 30 minutes. The fixed cells were washed in HBSS with 0.1% BSA and resuspended in HBSS with 0.1% BSA. The cells were filtered through a 40  $\mu\text{m}$  mesh filter and subsequently subjected to flow cytometry. All samples were analyzed using a BD LSR Fortessa (BD Biosciences) flow cytometer.

Acquired data were analyzed with FlowJo Software (FlowJo, LLC). Single cells with positive EGFP or mCherry signal were gated for two subpopulations: a “red” subpopulation with higher mCherry signal and lower EGFP signal (higher mCherry/EGFP ratio), and a “green” subpopulation with lower mCherry signal and higher EGFP signal (lower mCherry/EGFP ratio). The above-mentioned gates were set in the control samples (reporter lines infected with non-targeting sgRNAs), so that approximately 30% of the population resides in the “red” or “green” subpopulation. The same gates were applied to all other samples, and cell numbers in each gate were counted.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### CRISPR Screen Analysis

sgRNA sequences were extracted from raw reads based on matches to the tracrRNA and U6 promoter, and perfect matches to expected sgRNAs were tallied. Log<sub>2</sub>-fold changes of counts in sorted versus unsorted populations from each time point were

calculated. To score sgRNAs that impact reporter fluorescence, raw counts from sorted and unsorted populations were analyzed with the *count* module of MAGeCK version 0.5.6 (Li et al., 2014) with default settings where 1,000 non-targeting control sgRNAs present in the Brie library were specified with parameter `-control-sgrna`. Genes over-represented at an FDR < 0.1 with any reporter and gating regimen (2% or 30% highest or lowest mCherry/EGFP ratio for bichromatic reporters or 2% lowest/highest EGFP-Srrm4 signal) were considered as hits.

GO term enrichment analysis of hits from the bichromatic reporter screens was performed using g:Profiler with the list of genes targeted by the CRISPR library as a background (Table S1). The network view (Figure 2B) was generated with the Enrichment Map plugin for Cytoscape based on category overlaps. Relative representation of genes with genetic links to autism spectrum disorder (ASD) within the set of hits was evaluated by comparing the hits with mouse orthologs of human ASD genes curated in the Simons Foundation Autism Research Initiative Human Gene Module (<https://gene.sfari.org/database/human-gene/>), restricting to genes scored with evidence codes 1-5 or 'syndromic'. Only genes with an expression of > 1 cRPKM in N2A cells were considered.

To assess the efficiency of CRISPR knockouts and the strategy for sgRNA amplification from genomic DNA, sequencing and scoring, we determined Bayes factors (BF) reflecting the likelihood of genes to affect cell abundance using the BAGEL algorithm. Mouse orthologs of human core-fitness and non-fitness genes (Hart et al., 2015, 2017) were used as positive and negative gene sets, respectively, restricting the analysis to those genes with 1-to-1 orthology. The precision-recall analysis of gold-standard essential genes suggests that our CRISPR screens have an excellent performance in terms of gene inactivation and the drop-out from the cell population of the sgRNAs targeting gold-standard essential genes (Figure S2F) (Hart et al., 2015). Furthermore, the BF distribution of the core fitness and non-fitness genes display very little overlap, further confirming the quality of our screens (Figure S2G). A limitation of CRISPR screening methodologies coupled to phenotypic readouts is that essential genes will likely fail to be detected due to the rapid depletion of guides targeting these genes from the cell population. Although the vast majority of genes in N2A cells are non-essential, with a negative BF score (Figure S2H), the GO categories enriched among our CRISPR screen hits are represented by a higher percentage of essential genes than genes targeted by the CRISPR library (Figure S2D). To further analyze the effect of essentiality on screen hits, genes were binned into ten groups based on their BF. By calculating the percentage of screen hits in each BF bin we observe an increase from low to intermediate BF in parallel with an increasing fraction of genes with GO annotations related to those that impact microexon regulation, followed by an unexpected reduction of hits at high BF (Figure S2E). Assuming an expected hit ratio in these high-BF bins at least not lower than in medium-BF bins, we calculated that our screen failed to detect 82 regulators which corresponds to a false negative rate of ~0.4% and a recall of ~75% (Figure S2E).

### Analysis of Flow Cytometry Validations

For statistical analysis, each sample was compared to control samples treated with sgRNAs targeting LacZ or Luciferase using one-sided binomial tests with the same expected direction as observed in the screen. The less significant p value from the two tests (corresponding to comparisons against the two control samples) was chosen to represent significance of the sample. P values were corrected for multiple testing using the Benjamini-Hochberg procedure. The heatmap in Figure S2F indicates the ratio of cell numbers in the "red" and "green" subpopulations as well as the adjusted p value. For Figure S2H, any hit that was validated by at least one out of the two sgRNAs was considered as positively validated.

Two of the hits identified by the biochromatic CRISPR screen to result in an increased mCherry/EGFP ratio are the Uros and Urod genes, which are key components of the porphyrin biosynthesis pathway. Our validations revealed that genetic depletion of these two genes causes an increased signal in mCherry fluorescence (detected at 610 nm with excitation at 561 nm) channel irrespectively of microexon splicing (Figure S2F). Interestingly, mutations in these genes cause congenital erythropoietic porphyria (CEP), a manifestation of which is red discoloration of the urine due to porphyrin accumulation, suggesting that knockout of these genes may result in increased detection of red fluorescence independently of mCherry due to the accumulations of porphyrins in N2A cells. To test this hypothesis we transduced parental N2A cells (that do not express any fluorescent proteins) with virus expressing sgRNAs targeting Uros and Urod. Indeed, these cells displayed increased signal in mCherry fluorescent channel (Figure S2F). We thus predict that Uros and Urod will be common false-positive hits in CRISPR screens that use mCherry and other red fluorophores as a readout.

### SPAR-Seq Analysis

Raw reads were de-multiplexed based on expected forward and reverse barcode reads and mapped to custom splice junction libraries as previously described (Han et al., 2017). Subsequently, AS changes were quantified by calculating strictly standardized mean difference (SSMD) scores against non-targeting and mock transfection controls, also as described previously (Han et al., 2017). In addition, AS changes of 10 microexons in this publication, of which three overlapped with the experiment described above, was re-analyzed.

### RNA-Seq Analysis

Alternative splicing analysis of RNA-Seq data was performed with vast-tools version 1 (Tapial et al., 2017). From the primary output, events with poor coverage or junction balance were filtered out (vast-tools quality score 3 other than SOK/OK/LOW for cassette exon [CE], microexon [MIC], and alternative 5' or 3' splice site [Alt5/3] events or coverage less than 15 reads for intron retention [IR] events; score 4 other than OK/B1 for CE and MIC events and score 5 of less than 0.05 for IR events). Differential AS was scored using vast-tool's *diff* module requiring  $p(|dPSI| > 0) > 0.05$  and a point estimate of  $|dPSI| > 10$ .

### Annotation of Exons Misregulated in ASD

Information from two sources was combined for ASD annotation of alternative exons: (1) Data from post-mortem human cerebral cortex from individuals with ASD and controls (Parikshak et al., 2016) was downloaded from the github page associated with the publication and overlaid with human VastDB exons including microexons (Tapial et al., 2017), allowing for 6 nt differences in coordinates, followed by looking up of the orthologous mouse exons. Differential AS between ASD and control individuals was scored using a Mann-Whitney U-test followed by Benjamini-Hochberg correction of p values. (2) Differences between ASD and control post-mortem human samples from Brodman areas 9 and 41 (Irimia et al., 2014) were labeled as differential if the average dPSI was  $> 10$  or  $< -10$  in both Brodman areas, and matched to the orthologous mouse AS events. Annotations from both datasets were combined by annotating an event as 'ASD-high' or 'ASD-low' if it was concordantly labeled in both or annotated in only one dataset, ensuring a consistent trend among different brain regions.

### iCLIP-Seq Analysis

Analysis of iCLIP-Seq data was performed as previously described (Han et al., 2017). 51-nt raw reads that consisted of 3 random positions, a 4-nt multiplexing barcode, and another 2 random positions, followed by the cDNA sequence, were initially de-duplicated based on the first 45 nt. Reads were de-multiplexed and the random positions, barcodes, and any 3'-bases matching Illumina adaptors were removed. Finally, reads shorter than 25 nt were filtered out and remaining reads trimmed to 35 nt. These steps were carried out using Trimmomatic. Surviving reads were mapped to the mouse genome/transcriptome (Ensembl annotation of NCBI m37) using tophat with default settings. To prevent false assignments of reads from repetitive regions, any reads with a mapping quality  $< 3$  were removed from further analysis. Plots showing average crosslinking signal of events aligned to exon boundaries were generated as described after first reducing reads to their first position, which is adjacent to the crosslink position. A 21-bp running window average was used for display only, and average signals across replicates are shown. For plots showing intronic and unspliced reads only, reads entirely within an exon or mapping with a splice were excluded and the same procedure was followed.

Motif analysis was carried out as follows: All 6-mers on the same strand as the read, which corresponds to the strand of the crosslinked RNA, and within 101-nt windows centered on each read's first position were extracted and hexamer frequencies from replicates were averaged. To moderate biases arising from the effect of preferential crosslinking to uracil and differential RNA abundance, 6-mer frequencies were then normalized by subtracting mean frequencies across a panel of independent iCLIP experiments conducted in N2A cells in the Blencowe laboratory (Figure 7A), where each experiment received a weight proportional to its mean correlation with all experiments (this study; (Han et al., 2017) and unpublished data). Co-occurrence of the U/C-repeat (UCUCU/CUCUC) and UGC motifs was scored when the former was located between  $-80$  and  $-15$  nt upstream of the 3' splice site, and the latter between  $-20$  nt and the splice site, and the U/C repeat motif was located further upstream than the most exon-proximal UGC, in accordance with their strongest enrichment.

For heatmaps depicting iCLIP and motif occurrences, iCLIP data was smoothed with a 5-nt running mean. The profile within the plot coordinates was then scaled to the maximum for each exon.

### ChIP-Seq Analysis

Mapped reads and peak annotations of p300 ChIP in mouse forebrain, midbrain, and limb were retrieved from (Visel et al., 2009). For plotting, reads were extended to 400 nt in line with the authors' experimental procedures.

### Mass Spectrometry Data Analysis

Mass spectrometry data was stored, searched and analyzed using the ProHits laboratory information management system (LIMS) platform. The WIFF data files were converted to MGF format using WIFF2MGF and subsequently converted to an mzML format using ProteoWizard (3.0.4468) and the AB SCIEX MS Data Converter (V1.3 beta). The mzML files were searched using Mascot (v2.3.02) and Comet (2014.02 rev.2). The results from each search engine were jointly analyzed through the Trans-Proteomic Pipeline (TPP) via the iProphet pipeline. The spectra were searched against a total of 58,206 proteins consisting of the NCBI mouse RefSeq database (v53, Sep 9th, 2015, forward and reverse sequences) supplemented with "common contaminants" from the Max Planck Institute (<http://141.61.102.106:8080/share.cgi?ssid=0f2gfuB>) and the Global Proteome Machine (<https://www.thegpm.org/crap/index.html>) as well as sequences from common fusion proteins and epitope tags. The database parameters were set to search for tryptic cleavages, allowing up to two missed cleavage sites per peptide, MS1 mass tolerance of 40 ppm with charges of 2+ to 4+ and an MS2 mass tolerance of  $\pm 0.15$  amu. Asparagine/glutamine deamidation and methionine oxidation were selected as variable modifications. A minimum iProphet probability of 0.95 was required for protein identification. Proteins detected with a minimal number of two unique peptides were used for protein interaction scoring.

For both AP-MS and BioID analysis, Significance Analysis of INTERactome (SAINTexpress version 3.6.1) was used as a statistical tool to calculate the probability value of each potential interaction from background contaminants. Briefly, our experimental design included specific negative controls (FLAG-GFP, and FLAG for FLAG AP-MS; BirA\*-FLAG-GFP, and BirA\*-FLAG-empty for BioID), each run in several biological replicates (eight and seven total respectively). To increase the stringency in scoring, controls were further "compressed" (i.e., the top x values across y controls are recovered for each prey) prior to running SAINTexpress (compression to four control samples for AP-MS and three control samples for BioID respectively). Each biological replicate analysis of a bait

was analyzed independently against these compressed controls, before averaging of the score values and assessment of the Bayesian False Discovery Rates (BFDR). High-confidence interactions are those with  $\text{BFDR} \leq 1\%$ .

All mass spectrometry data associated with this study, including complete SAINTexpress results, have been deposited at the ProteomeXchange consortium through partner MassIVE (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>). The FLAG dataset has been assigned IDs MSV000082174 and PXD009226 respectively. Likewise, the BioID dataset has been assigned IDs MSV000082169 and PXD009213. Data can be accessed at <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=949fbf1aa5364f28857ab14a13cf9571> (FLAG) and <https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=552ed3d256014deda445199521d88691> (BioID). FLAG and BioID data have also been uploaded to the ProHits-web interaction proteomics repository for query (<https://prohits-web.lunenfeld.ca/>).

### **Analysis of Interactions with Spliceosomal snRNPs**

Membership in spliceosomal sub-complexes of prey proteins identified in AP-MS and BioID experiments was analyzed using mouse orthologs of human protein annotations in the Spliceosome Database (Cvitkovic and Jurica, 2013). In Figures 5D and S5B, only preys with a BFDR < 0.05 for any of the baits are shown. The network map was generated using Cytoscape. Only preys with BFDR < 0.01 are shown.

### **DATA AND SOFTWARE AVAILABILITY**

The GEO accession numbers for sequencing-based data reported in this paper are: CRISPR screens, GSE112599; SPAR-Seq, GSE120164 and GSE80196; RNA-Seq, GSE112600 and GSE57278; CLIP-Seq, GSE112598. The accession numbers for the AP-MS data are: MassIVE: MSV000082174, ProteomeXchange: PXD009226. The accession numbers for BioID data are: MassIVE: MSV000082169, ProteomeXchange: PXD009213.