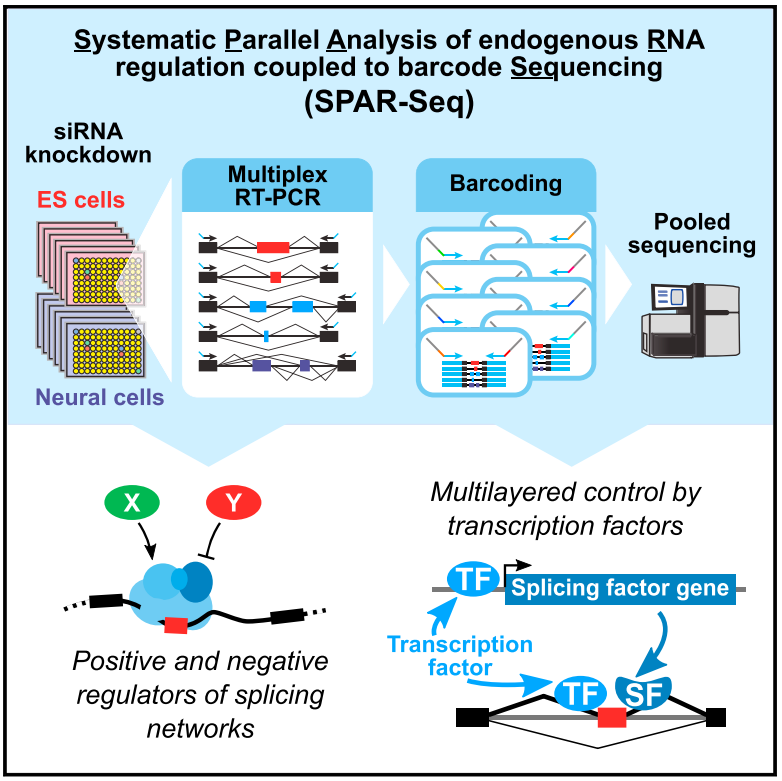
Resource

Molecular Cell

Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors

Graphical Abstract



Authors

Hong Han, Ulrich Braunschweig, Thomas Gonatopoulos-Pournatzis, ..., Jeffrey L. Wrana, Jason Moffat, Benjamin J. Blencowe

Correspondence

b.blencowe@utoronto.ca

In Brief

Han, Braunschweig et al. describe SPAR-seq, a strategy for the systematic elucidation of endogenous RNA regulatory networks. Hundreds of factors controlling cell-fate-associated splicing networks are defined, among which are transcription factors that have dual direct and indirect regulatory roles.

Highlights

- SPAR-seq is a versatile method for elucidating endogenous RNA regulatory networks
- Hundreds of *trans*-acting factors control ES and neural splicing networks
- SPAR-seq defines positive and negative functional and physical interrelationships
- Transcription factors have dual direct and indirect roles in splicing regulation

Han et al., 2017, Molecular Cell *65*, 539–553
February 2, 2017
http://dx.doi.org/10.1016/j.molcel.2017.01.011

CellPress

Molecular Cell
Resource

Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors

Hong Han,[1,2,6] Ulrich Braunschweig,[1,6] Thomas Gonatopoulos-Pournatzis,[1] Robert J. Weatheritt,[1,3] Calley L. Hirsch,[4] Kevin C.H. Ha,[1,2] Ernest Radovani,[1,2] Syed Nabeel-Shah,[1,2] Tim Sterne-Weiler,[1] Juli Wang,[1] Dave O'Hanlon,[1] Qun Pan,[1] Debashish Ray,[1] Hong Zheng,[1] Frederick Vizeacoumar,[4] Alessandro Datti,[4] Lilia Magomedova,[5] Carolyn L. Cummins,[5] Timothy R. Hughes,[1,2] Jack F. Greenblatt,[1,2] Jeffrey L. Wrana,[2,4] Jason Moffat,[1,2] and Benjamin J. Blencowe[1,2,7,*]
[1]Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
[2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada
[3]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
[4]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada
[5]Department of Pharmaceutical Sciences, University of Toronto, Toronto, ON M5S 3M2, Canada
[6]Co-first author
[7]Lead Contact
*Correspondence: b.blencowe@utoronto.ca
http://dx.doi.org/10.1016/j.molcel.2017.01.011

SUMMARY

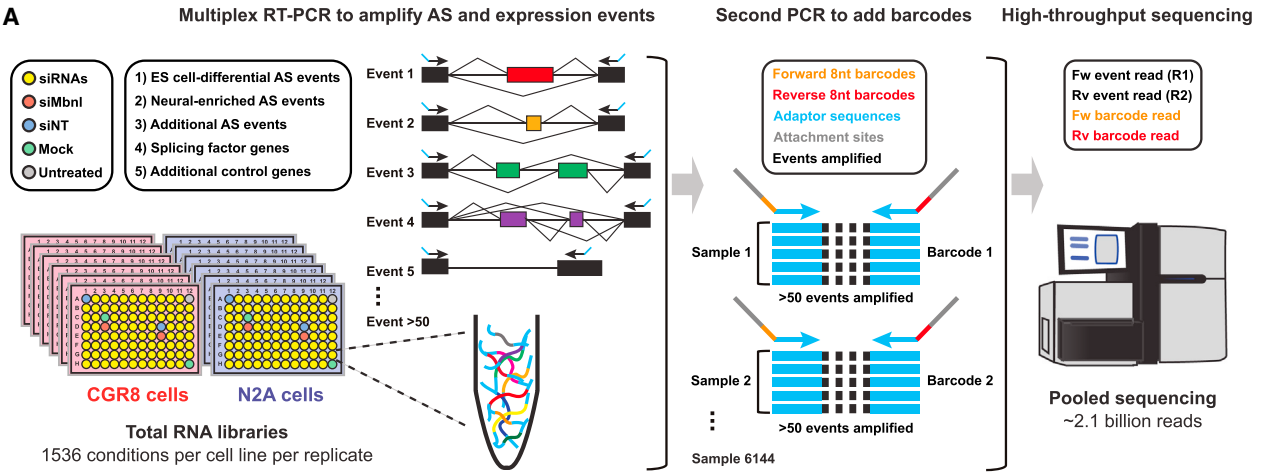
Networks of coordinated alternative splicing (AS) events play critical roles in development and disease. However, a comprehensive knowledge of the factors that regulate these networks is lacking. We describe a high-throughput system for systematically linking *trans*-acting factors to endogenous RNA regulatory events. Using this system, we identify hundreds of factors associated with diverse regulatory layers that positively or negatively control AS events linked to cell fate. Remarkably, more than one-third of the regulators are transcription factors. Further analyses of the zinc finger protein Zfp871 and BTB/POZ domain transcription factor Nacc1, which regulate neural and stem cell AS programs, respectively, reveal roles in controlling the expression of specific splicing regulators. Surprisingly, these proteins also appear to regulate target AS programs via binding RNA. Our results thus uncover a large "missing cache" of splicing regulators among annotated transcription factors, some of which dually regulate AS through direct and indirect mechanisms.

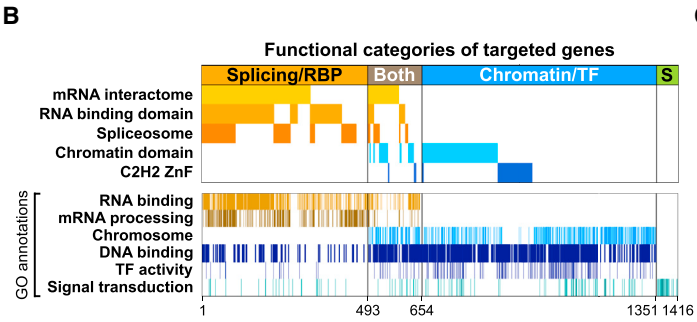
INTRODUCTION

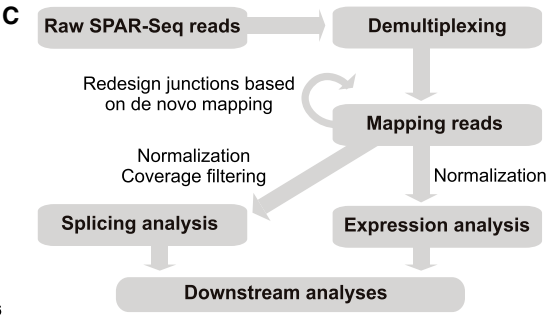
Alternative splicing (AS) is the process by which different combinations of splice sites in precursor mRNA (pre-mRNA) are selected to generate structurally and functionally distinct mRNA and protein variants. It acts widely to expand the functional and regulatory capacity of metazoan genomes (Irimia and Blencowe, 2012; Licatalosi and Darnell, 2010; Nilsen and Graveley, 2010). For example, nearly all transcripts from human multi-exon genes are alternatively spliced, and a substantial fraction of these splice variants are differentially expressed in a cell- and tissue-specific manner (Pan et al., 2008; Wang et al., 2008). AS has critical roles in diverse biological processes, including cell-fate determination, and misregulation of AS is associated with numerous diseases (Daguenet et al., 2015; Jangi and Sharp, 2014; Kalsotra and Cooper, 2011). An important challenge is to understand how networks of AS events are coordinately regulated to impart their biological roles in different cellular contexts.

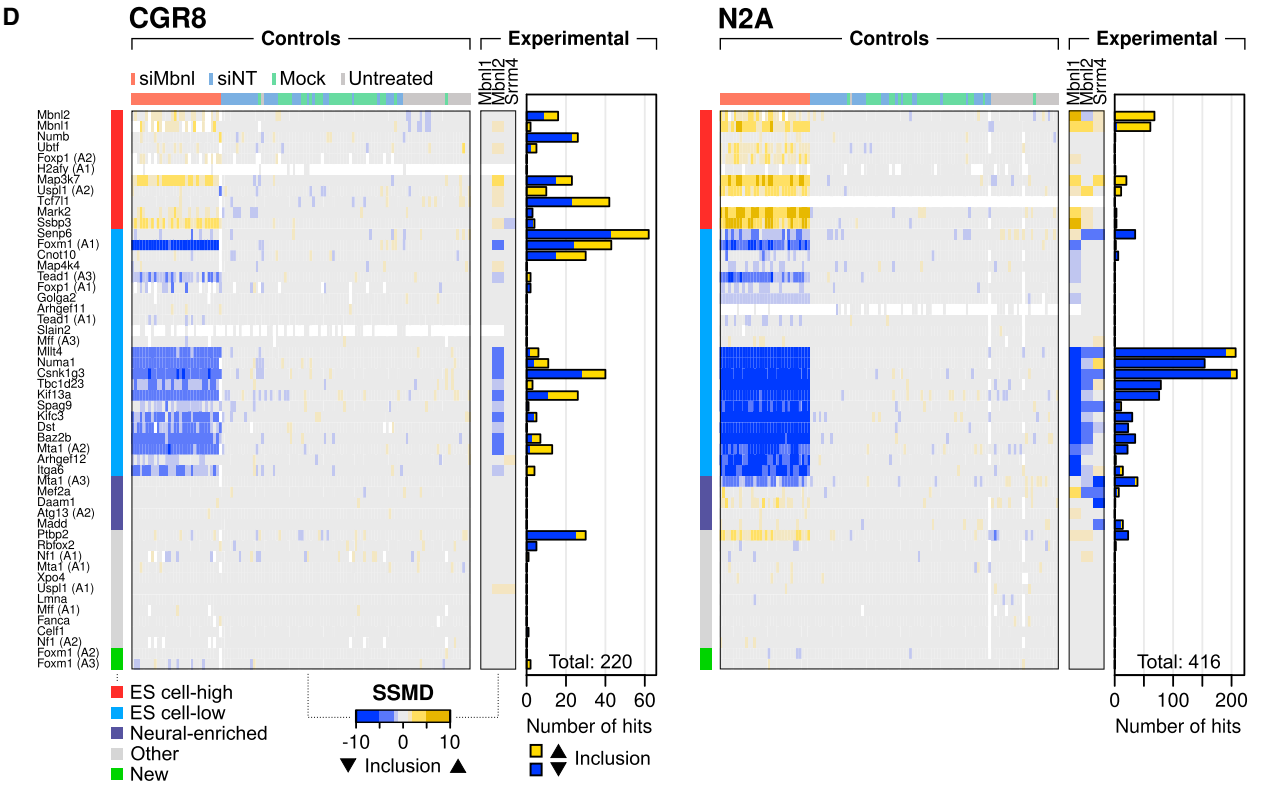
The spatiotemporal specificity of AS is governed by combinations of *cis*-regulatory elements and cognate *trans*-acting factors, which promote or inhibit spliceosome assembly (Chen and Manley, 2009; Fu and Ares, 2014; Wahl et al., 2009). AS is also controlled by coordinated interactions with other regulatory layers, including transcription and chromatin (Braunschweig et al., 2013). Moreover, post-translational and signaling pathways influence AS through different mechanisms, such as by altering the function and/or localization of key splicing regulators (Heyd and Lynch, 2011). However, the full repertoires of splicing regulators and associated mechanisms in different cell types are not known. This question is especially relevant to cell types with relatively complex AS patterns, such as embryonic stem (ES) cells and neural cells.

Considerable progress has been made in the development of strategies for large-scale investigations of AS. Fluorescent- or luciferase-based splicing minigene reporters have been utilized in high-throughput RNA interference (RNAi), small molecule, and cDNA overexpression screens to discover factors that control individual AS events (Moore et al., 2010; Stoilov et al., 2008; Warzecha et al., 2009; Zheng et al., 2013). However, since splicing reporters often do not recapitulate important aspects of AS regulation, such as crosstalk with chromatin and transcription components, high-throughput methodologies that monitor endogenous AS changes are required. Progress in this direction has included the employment of a high-throughput quantitative PCR (qPCR) assay to screen mutant strains of budding yeast for splicing regulators (Albulescu et al., 2012) and an automated RT-PCR platform coupled to capillary gel electrophoresis or sequencing to monitor effects of knockdown of candidate









(legend on next page)

regulators on apoptosis- and proliferation-related AS events (Papasaikas et al., 2015; Tejedor et al., 2015; Venables et al., 2008). These studies have illuminated interesting and important functional relationships within and between core and ancillary splicing regulators, as well as other factors, that control specific AS events.

In this study, we describe "Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing" (SPAR-seq), a multiplexed and quantitative functional genomics screening platform coupled to a sequencing output that is capable of comprehensively linking *trans*-acting factors to dozens of endogenous gene regulation events of interest. We use SPAR-seq to elucidate regulatory networks that control conserved, endogenous AS events linked to ES cell pluripotency, neural differentiation, and somatic cell reprogramming. Our results reveal hundreds of previously unknown splicing regulators associated with different regulatory layers that impact distinct subsets of these AS events in ES and neural cells. Surprisingly, in neural cells, annotated transcription and DNA-binding factors affect AS events at a comparable frequency as defined splicing regulators. Further characterization of Zfp871 and Nacc1 provides evidence that these factors directly and indirectly regulate neural and ES-cell-differential AS networks, respectively. This study thus introduces a versatile technology for elucidating endogenous RNA regulatory networks and highlights its application in revealing *trans*-acting regulators and associated multilayered mechanisms that impact AS events with key roles in cell-fate decisions.

RESULTS

A High-Throughput System for Linking *trans*-Acting Factors to Endogenous AS Events

To systematically discover factors that control endogenous AS networks that impact cell fate, we employed SPAR-seq to interrogate 1,536 small interfering RNA (siRNA) knockdown and control treatments for effects on 52 evolutionarily conserved (i.e., between human and mouse) AS events that are associated with ES cell pluripotency, neural differentiation, and somatic cell reprogramming (Figure 1A; Table S1; STAR Methods). The 52 prioritized AS events were simultaneously amplified by multiplex RT-PCR in 96-well plates, following well-specific siRNA knockdowns. Unique, dual-index barcodes were added to the RT-PCR amplicons to mark each well. Subsequent pooling and high-throughput sequencing analysis of the barcoded amplicons were then used to link well-specific knockdowns to changes in AS (Figure 1A; Figure S1A).

AS events that were analyzed comprised 35 cassette alternative exons, including eight microexons (3–27 nt), 15 alternative exons from eight genes with more complex AS patterns, and two previously unannotated alternative 3′ splice site events in Foxm1 that were identified through de novo splice-junction mapping. The screen was performed in mouse ES (CGR8) and neuroblastoma (N2A) cells to identify positive and negative regulators of the assayed AS events. Genes subject to knockdown included 654 known and putative splicing/RNA-associated factors, encompassing all annotated spliceosomal-associated proteins, and all known and predicted RNA binding proteins (RBPs) (Figure 1B; Table S2). We also interrogated 858 genes with annotations linked to chromatin and transcription that are expressed in CGR8 or N2A cells, 161 of which overlap those linked to splicing and RNA binding. Additionally, we assayed 65 signaling-related and post-translational factors with links to AS regulation (Figure 1B; Table S2).

Total RNA was harvested 48 hr post-transfection of siRNAs and subject to SPAR-seq. An analysis pipeline was designed to extract data, quantify AS levels by calculating percentage of transcripts with the exon spliced in ("PSI") values, and prioritize detected PSI changes for further analysis using the "strictly standardized mean difference" (SSMD; Zhang, 2007) metric (Figure 1C; Figures S1B and S1C; Table S3; STAR Methods). SSMD measures effect size while taking the variance between replicates into account. In total, 316,863 AS measurements were analyzed across two biological replicate screens. The SPAR-seq data were additionally used to monitor mRNA expression levels for all genes assayed for splicing changes, as well as for a representative set of splicing factors (Figure 1C; Table S4; STAR Methods). These data indicated >60% depletion of all monitored genes for which knockdowns were carried out (Figure S1D). (q)RT-PCR experiments using independent samples from CGR8 and N2A cells validated SPAR-seq-detected changes for nearly all analyzed AS and mRNA expression events (Figure S1E). Furthermore, the magnitude of screen-detected PSI changes correlated well with changes measured using independent RT-PCR ($r = 0.77$, $n = 335$) and RNA sequencing (RNA-seq) ($r = 0.89$, $n = 224$) (Figures S1E and S1F).

Figure 1. Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing

(A) Schematic outline of Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing (SPAR-seq) strategy. Total RNA from 1,536 siRNA knockdown and control treatments applied to CGR8 and N2A cells was prepared in biological replicate. A two-step, PCR-based amplification and indexing protocol generated multiplexed barcoded libraries, monitoring >50 endogenous AS and gene expression events.

(B) Functional categories and Gene Ontology (GO) annotations associated with genes analyzed by SPAR-seq. Number of targeted genes is indicated. S, signaling and post-translational factors.

(C) Flowchart of pipeline for analysis of SPAR-seq data.

(D) SSMD scores for representative screen results. Positive (siMbnl) and negative (siNT, mock, and untreated) control treatments, as well as knockdown of splicing factors Mbnl1, Mbnl2, and Srrm4, are shown. Endogenous AS events monitored are grouped into ESC-high, ESC-low, neural-enriched, and others. ESC-high and ESC-low refer to exons that are preferentially included and skipped in ES cells relative to other cell and tissue types, respectively (Han et al., 2013). Two AS events detected by de novo alignment are shown at the bottom. Multiple events within the same gene are denoted as A1 up to A3. Columns (treatments) for controls are clustered in the same way for CGR8 and N2A, and rows (events) are clustered within each group. See Figure S2A for full results. Bar graphs show numbers of knockdowns leading to a significant increase (yellow) or decrease (blue) in the inclusion level of each exon above negative controls. See also Figures S1 and S2, Tables S1, S2, S3, and S4, and STAR Methods.

Negative controls (32 siNTs, non-targeting siRNA; 32 mock controls; 24 untreated samples) uniformly resulted in little to no change in PSI levels, whereas positive controls (32 siMbnl, simultaneous knockdown of Mbnl1 and Mbnl2) resulted in significant changes for many ES-cell-differential AS events toward an ES-like pattern, without substantially affecting neural-enriched AS events, consistent with our previous results (Figure 1D; Figure S2A) (Han et al., 2013). Individual knockdown of Mbnl proteins and additional regulators previously linked to pluripotency and reprogramming, including Rbfox2, Son, Srsf2, Srsf3, and U2af1 (Lu et al., 2013; Lu et al., 2014; Ohta et al., 2013; Venables et al., 2013), also affected ES-cell-differential AS events (Figure 1D; Table S3). Conversely, knockdown of the neuronal-specific splicing regulator nSR100/Srrm4 (Calarco et al., 2009) in N2A cells affected neural-enriched AS events without substantially affecting ES-cell-differential AS events (Figure 1D). Additional analyses further demonstrated the specificity and reproducibility of the screen data, both within and between replicate experiments (Figure S2B; STAR Methods). In total, knockdown of 220 and 416 factors in CGR8 and N2A cells, respectively, resulted in significant changes in one or more endogenous AS events (Figure 1D, bar graphs).

Correlated AS Changes Reveal Multilayered Regulatory Pathways and Complexes

To investigate functional relationships between factors that impact AS, we determined overall correlations between the changes in PSI values of the 52 AS events from all pairwise comparisons of knockdowns in CGR8 and N2A cells. Clustering of the resulting data based on the degree of pairwise correlation similarity revealed groups of factors that positively or negatively regulate specific subsets of AS events in a similar manner (Figures 2A and 2B; Table S5; STAR Methods). Consistent with previous results (Papasaikas et al., 2015; Tejedor et al., 2015), these groups are often significantly enriched in factors that function in distinct biological processes and pathways and, in many cases, are known to interact within complexes. Complementing this approach, we systematically surveyed factors in experimentally defined complexes (from CORUM; Ruepp et al., 2010) for coordinated AS changes by scoring mean pairwise correlations between complex members (Figure 3A; Table S6). Remarkably, of 316 complexes represented by at least three factors in our screen, 38 (12%) in CGR8 and 187 (59%) in N2A cells displayed significant correlations above background (false discovery rate [FDR] < 0.05, Mann-Whitney U test). SPAR-seq data thus links protein complexes to the control of specific subsets of AS events associated with cell fate and, as such, provides insight into physical and functional interactions between *trans*-acting factors.

One of the largest groups of correlated knockdowns in CGR8 cells involves factors associated with the 17S U2 small nuclear ribonucleoprotein particle (snRNP) (Figure 2A, cluster 8), which binds the pre-mRNA branch site during spliceosome formation (Wahl et al., 2009). Remarkably, the degree of correlation reflects physical relationships among these components (Papasaikas et al., 2015): members of the heptameric Sm complex form a central, highly correlated module adjacent to additional (i.e., SF3a/b) U2 snRNP components. In contrast, auxiliary factors that facilitate U2 snRNP recruitment to pre-mRNA, such as U2af1 and U2af2 (Wahl et al., 2009), and the intron binding complex (IBC) components Aqr and Isy1 (De et al., 2015), are less well correlated with core U2 snRNP proteins and form a distinct cluster (Figure 3B). Another distinct sub-cluster comprises Rbm17, Cherp, U2surp, and Dhx15. Rbm17 controls splice site selection by interacting with U2 snRNP and by recognizing the 3′ splice site AG during the second catalytic step of splicing (Corsini et al., 2007; Lallena et al., 2002) and interacts with U2surp and Dhx15 (Hegele et al., 2012). The correlated effects of knockdown of these factors suggest that they may have closely related functions in regulating a subset of AS events linked to cell fate.

Interestingly, U2 snRNP-associated factors are more highly expressed in ES and reprogramming cells than in differentiated cells or tissues (Figure S3A) (Han et al., 2013; Hirsch et al., 2015), and knockdown of these factors shifted a sub-network of ES-cell-differential AS events toward a differentiated cell-like pattern (Figure 2A; Tables S3 and S5). Collectively, these results suggest that increased levels of U2 snRNP-associated components may be important for maintaining AS patterns specific to ES cells, whereas reduced levels are associated with differentiated cells.

Positive and Negative Roles of Different Gene Regulatory Layers in Controlling AS

Other clusters of significantly correlated knockdowns in our screen data include factors that form additional sub-complexes involved in splicing but also complexes and pathways associated with RNA polymerase II (pol II) transcription, chromatin modification, DNA-binding/transcription factor activity, mRNA surveillance, turnover, transport, and translation (Figures 2 and 3A). Many of the factors in these clusters have not been previously linked to splicing control.

One such cluster comprises multiple factors associated with the exon junction complex (EJC), which controls splicing-dependent mRNA export and turnover via the nonsense-mediated mRNA decay (NMD) pathway (Tange et al., 2004). Knockdown of "core" EJC components (i.e., Eif4a3, Magoh, Rbm8a, and Casc3) resulted in more highly correlated effects on AS patterns than knockdown of "non-core" EJC factors (i.e., Ddx39b, Alyref, Srrm1, and Rnps1) (Figure 3C). While the peripheral EJC factors Srrm1 and Rnps1 have reported functions in splicing regulation (Eldridge et al., 1999; Sakashita et al., 2004), the present data are consistent with emerging evidence that core EJC-associated components also regulate AS (Ashton-Beaucage et al., 2010; Michelle et al., 2012; Papasaikas et al., 2015). Moreover, our results link these factors to the control of a subset of AS events associated with cell fate.

Our data also revealed anti-correlated (i.e., antagonistic) effects on AS (Figure 2). For example, knockdown of factors associated with transcription/DNA-binding activity (e.g., Nfyb, Tbx1, and Tfdp1) (Figure 2A, cluster 10) resulted in Foxm1 AS changes that are anti-correlated with the knockdown of a subset of splicing-associated factors (e.g., Snrnp200, Bcas2, Rbfox2, and Mbnl proteins) (Figure 2A, cluster 3) ($p < 0.01$, one-sided binomial tests). Antagonistic (and positive) effects were also observed between Arg/Ser-repeat (RS) domain proteins and other factors (Figure 2A, e.g., clusters 2 and 5). As an example, knockdown of a previously uncharacterized splicing regulator,

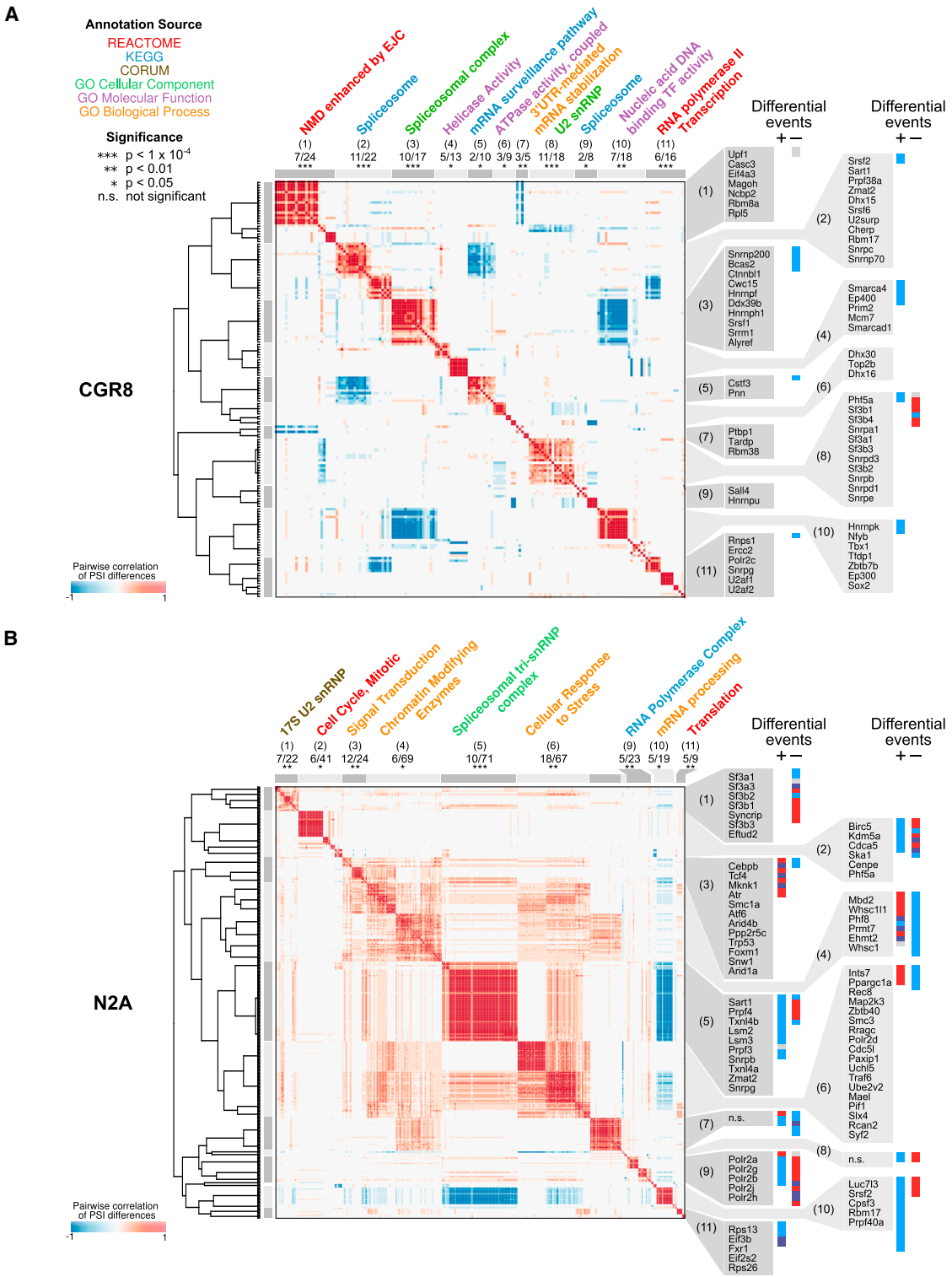


Figure 2. Correlations of AS Changes Reveal Protein Complexes and Pathways

(A and B) Symmetrical heatmaps of pairwise correlation of AS changes in CGR8 (A) and N2A (B) cells. Knockdowns were clustered by affinity propagation and subsequent hierarchical clustering. Sub-clusters are annotated with the most significant terms from GO, CORUM, REACTOME, and KEGG. Numbers of genes annotated with each term and associated enrichment p value are indicated, and individual genes are listed on the right. Colored boxes (right) indicate types of events (see Figure 1D) that are differential for each cluster. Additional details in Table S5 and STAR Methods.

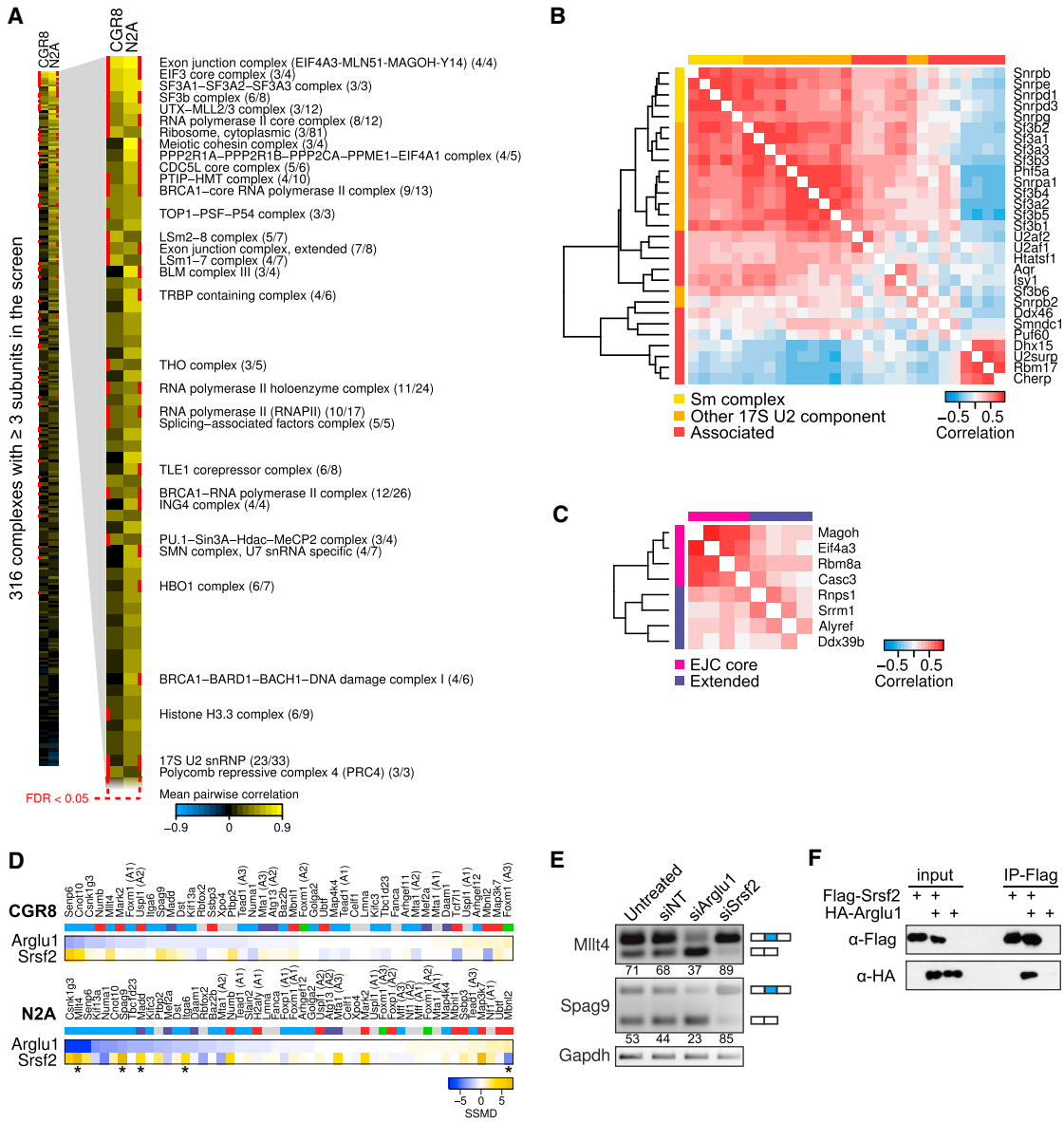


Figure 3. Factors Associated with Different Gene Regulatory Layers Positively or Negatively Control AS

(A) Mean pairwise correlations of AS changes upon knockdown of components of CORUM complexes, where correlations represent averages for at least three components of a complex. Complexes are sorted according to the highest correlation after averaging values between CGR8 and N2A cell lines. FDR reflects significance of pairwise correlation when compared to a background distribution of all pairwise comparisons (see STAR Methods and Table S6).

(B) Hierarchical clustering of components of 17S U2 snRNP analyzed by SPAR-seq, based on similarity between pairwise correlations of SSMD scores (see also Figure S3A).

(C) Same as (B) for EJC components.

(D) SSMD scores upon knockdown of Arglu1 and Srsf2, sorted by SSMD values in the Arglu1 knockdown. Event groups are indicated by colors as in Figure 1D. Asterisks indicate events tested in (E) and Figure S3B.

(E) Representative RT-PCR assays detecting AS changes in Mllt4 and Spag9 transcripts upon knockdown of Arglu1 and Srsf2. Percentage of transcripts with the exon spliced in (PSI) values are indicated below each gel image.

(F) Western blot analysis of immunoprecipitates (IPs) from 293T cells expressing Flag-Srsf2, HA-Arglu1, or both. Co-immunoprecipitation was performed with anti-Flag antibody in the presence of benzonase.

Arginine and Glutamate rich 1 (Arglu1) protein, and of Srsf2, had opposing effects on six AS events monitored in the screen (Figures 3D and 3E; Figure S3B). Arglu1 has previously been associated with estrogen receptor-mediated gene activation (Zhang et al., 2011), whereas the presence of an RS domain in this protein suggested that it also has a function in AS (Boucher et al., 2001). Our data, as well as results from an independent study (L.M. and C.L.C., unpublished data), confirm that Arglu1 is an

AS regulator. Since RS domains function in mediating interactions between splicing components (Lin and Fu, 2007), and because factors that associate physically often have related knockdown AS profiles in our screen, we asked whether Arglu1 and Srsf2 might antagonize each other by interacting. Co-immunoprecipitation in the presence of nuclease treatment followed by western blotting of HA-Arglu1 and FLAG-Srsf2 proteins shows that these proteins can interact (Figure 3F). These results thus highlight Arglu1 as a previously unknown splicing regulator that may function by physically antagonizing Srsf2.

Chromatin and Transcription Factors Function as AS Regulators

In N2A cells, knockdown of chromatin/transcription factors, including proteins with PWWP motif, chromodomain, and various annotated DNA-binding domains, affected AS with a similar frequency as knockdown of splicing/spliceosomal-associated proteins and additional factors with annotated RNA-binding or helicase domains (Figures 4A and 4B; Figure S4A). As examples, knockdown of components of pol II-containing complexes, including core pol II subunits, had pronounced and strongly correlated effects on AS, whereas knockdown of pol II initiation complex factors TFIIB and TFIID had distinct effects (Figure 3A; Figure S4B). AS is thus differentially impacted by perturbing temporarily distinct pol II complexes. Knockdown of structural maintenance of chromosomes (SMC) proteins forming the cohesin complex also markedly affected AS (Figure 3A), consistent with previous evidence of physical and functional links between SMC proteins and splicing factors (McCracken et al., 2005). Furthermore, knockdown of components of complexes regulating histone modification (e.g., the PTIP-HMT, TLE1-corepressor, ING4-containing, and HBO1 complexes), as well as DNA repair (e.g., the BLM complex III and BRCA1-BARD1 complex), also displayed significantly correlated effects on AS. Interestingly, a BRCA1-containing complex has previously been shown to recruit splicing factors to gene promoters upon DNA damage (Savage et al., 2014). Our results support a role for BRCA1 complexes in AS regulation and are further consistent with links between AS and different DNA repair pathways.

An Extensive Role for Zinc Finger Proteins in AS Regulation

Remarkably, the screen data revealed that knockdown of 57% (137/242) of factors with zinc finger (ZnF) domains affected AS. Among this set are 42% (49/116) of analyzed C2H2 ZnF proteins, which displayed striking effects on AS in N2A cells (Figure 4A). C2H2 ZnF proteins represent the largest class of nucleic acid-binding proteins, comprising 718 family members in the human genome and 583 members in the mouse genome (Emerson and Thomas, 2009; Tadepally et al., 2008). However, while the majority of analyzed C2H2 ZnF proteins bind DNA and a subset have been shown to control transcription or silence retrotransposition, the functions of the vast majority of these proteins are not known (Stubbs et al., 2011).

Hierarchical clustering (Figure 4C) and principal component analysis (Figure S4C) showed that knockdown of C2H2 ZnF proteins in N2A cells affected different subsets of AS events in a positive or negative manner. Some of the most pronounced changes are associated with knockdown of Gtf3a/TFIIIA, Yy1, Repin1, and Rest/Nrsf, all of which have previously been reported to bind RNA (Cassiday and Maher, 2002; Jeon and Lee, 2011; Lu et al., 2003; Pelham and Brown, 1980). Moreover, knockdown of Rest, a repressor of neurogenesis genes in non-neural cells, shifted splicing patterns of several AS events in an opposite direction to that of knockdown of Srrm4 (Figure S4C), consistent with our previous finding that Rest transcriptionally represses Srrm4, whereas Srrm4 promotes the splicing of neural exons, including an exon that silences Rest activity (Calarco et al., 2009; Raj et al., 2011).

An interesting candidate AS regulator from our screen is Zfp871, which contains a KRAB domain and 15 C2H2 ZnF domains. Knockdown of Zfp871 affects neural-enriched exons in a manner similar to that of Srrm4 (Figure 4C). Moreover, analysis of RNA-seq data of differentiation of ES cells to glutamatergic neurons (Hubbard et al., 2013) reveals that Zfp871, like Srrm4 (Raj et al., 2014), displays a marked increase in expression during neuronal development (Figure 5A). To further investigate the function of Zfp871 and its relationship with Srrm4, we performed RNA-seq analysis of AS and mRNA expression changes following knockdown of each protein in N2A cells (Figure 5B). Knockdown of Zfp871 affected splicing ($\Delta PSI > 10$) of 479 exons, of which 189 are neural differential, and 198 overlap those regulated by Srrm4 (both $p < 0.001$, Fisher's exact test; Figure 5B). RT-PCR validation confirmed changes for all ten analyzed Zfp871-dependent and -independent neural-differential AS events detected by SPAR-seq and/or RNA-seq (Figure 5C; Figure S5A). Similar to Srrm4 (Irimia et al., 2014), knockdown of Zfp871 predominately reduces the splicing of short neural-differential exons, including 3–27 nt microexons that are also regulated by Srrm4 (Figure 5B) ($p < 0.001$, Fisher's exact test for overlapping effects on $\leq$30 nt exons versus longer exons). Moreover, genes with AS events impacted by Zfp871 knockdown are significantly enriched in functional categories associated with neuronal morphology and function (Figure 5D).

We next investigated the mechanisms by which Zfp871 regulates neural-differential and/or Srrm4-dependent exons. Consistent with an indirect role, knockdown of Zfp871 results in an ~40% reduction in Srrm4 mRNA levels and also reduces the levels of other splicing regulators linked to neural AS, including Celf4, Raver1, Nova1, and Ptbp2 (Figure 5E). Remarkably, using individual-nucleotide resolution crosslinking and immunoprecipitation coupled to sequencing (iCLIP), we observed that Zfp871 preferentially binds intronic sequences proximal to its regulated target exons in N2A cells, in particular sequences adjacent to those exons for which it promotes inclusion (Figure 5F; Figure S5B). These increases in binding occupancy are not due to increased expression of the corresponding genes or retention of bound introns (data not shown). Furthermore, the binding profile of Zfp871 is similar to that of Srrm4, with increased occupancy within an ~50 nt region upstream of the 3′ splice site (Raj et al., 2014). Consistent with this observation, exons controlled by Zfp871 are flanked by sequences enriched in UGC motifs, which represent binding sites for Srrm4 (Figures S5C and S5D). These results suggest that Zfp871 controls neural AS via both direct and indirect mechanisms that impact Srrm4

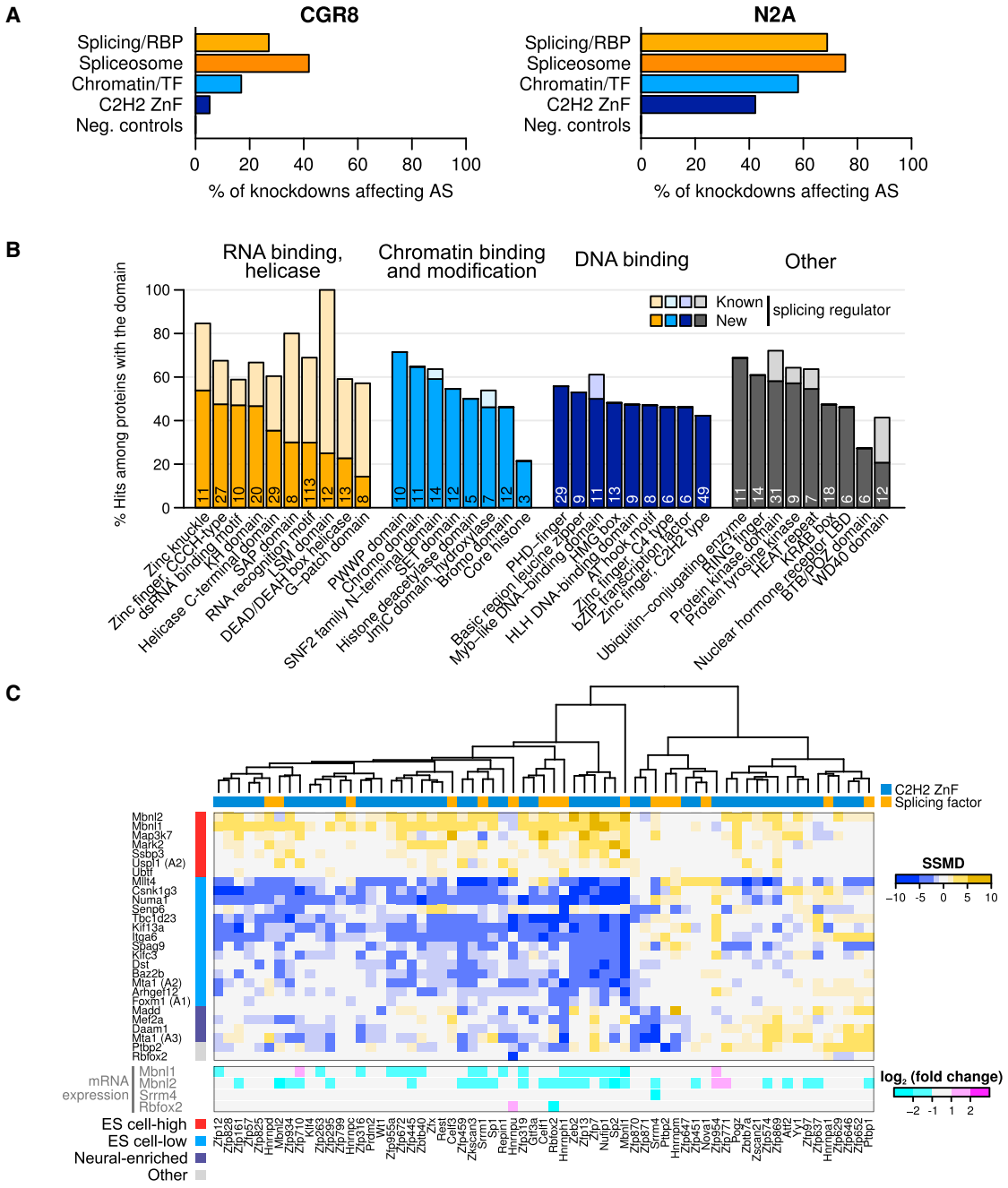


Figure 4. Chromatin and Transcription Factors Frequently Regulate AS in N2A Cells

(A) Percentages of knockdowns affecting at least one AS event with an |SSMD| > 2.25 in CGR8 and 3.00 in N2A, thresholds that robustly discriminate positive and negative controls (see also Figure S4A).

(B) Percentages of knockdowns of factors containing at least one of the indicated domains scored as a hit in N2A cells at the same threshold as in (A). Only domains occurring in at least ten of the factors analyzed in the screen are shown (refer to Table S2 for domain annotations). Absolute numbers of factors identified as regulators are indicated. Bars are sorted by the percentages of newly discovered AS regulators, defined as a factor not previously annotated with a GO category related to splicing (see STAR Methods).

(C) Heatmap and hierarchical clustering of AS changes (represented by SSMD scores) upon knockdown of all C2H2 ZnF factors identified as hits in N2A cells, together with developmentally regulated splicing factors. Only AS events and C2H2 ZnF factors with at least one change > 3 SSMD are shown. Expression changes of representative splicing factors, Mbnl1, Mbnl2, Srrm4, and Rbfox2 are shown. See also Figure S4.

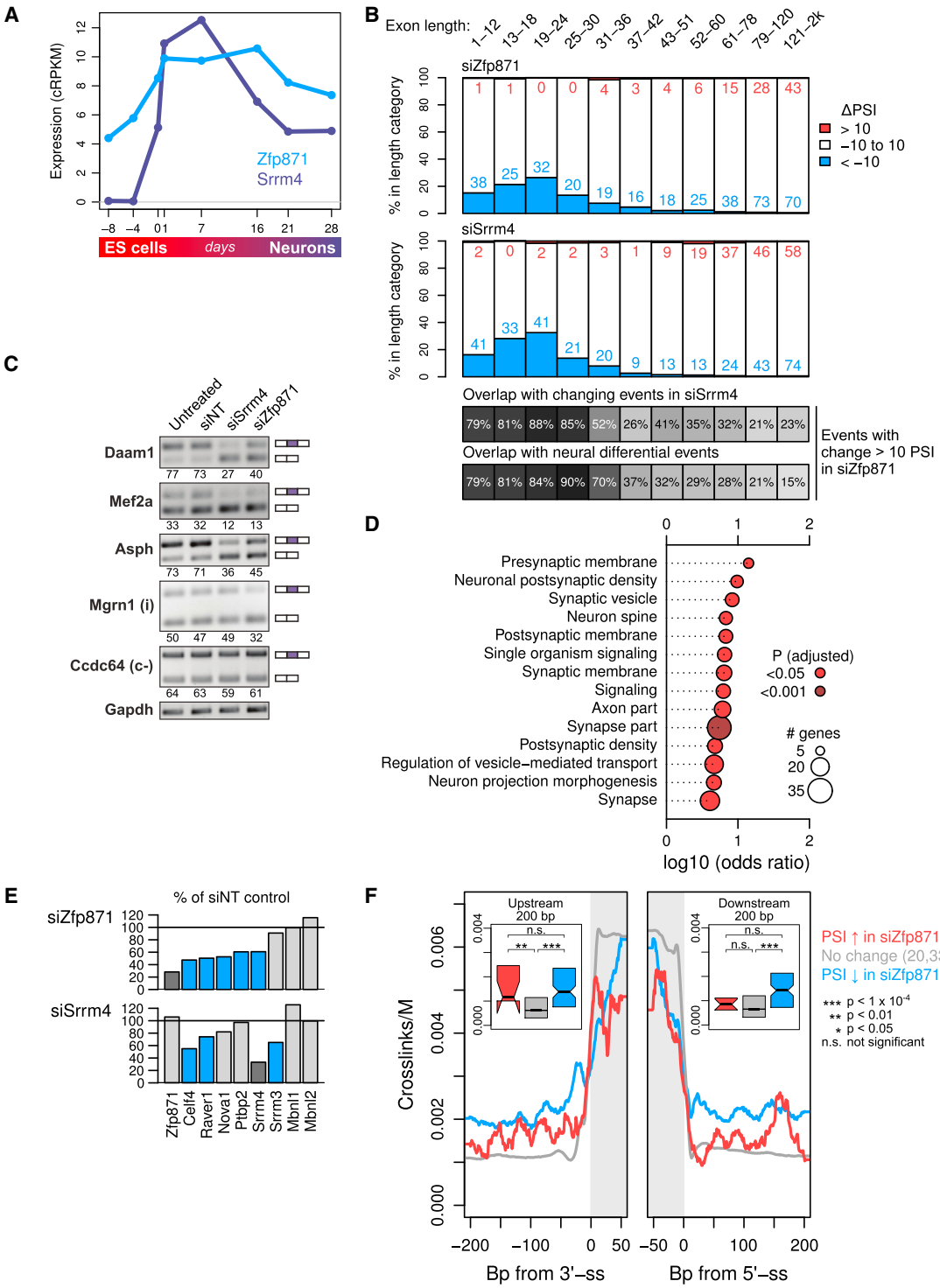


Figure 5. The C2H2 ZnF Factor Zfp871 Regulates Neural-Enriched Exons and Microexons

(A) mRNA expression of Zfp871 and Srrm4 during differentiation of ES cells to glutamatergic neurons.

(B) Alternative exons of different lengths affected by knockdown of Zfp871 or Srrm4, as analyzed by RNA-seq. Bar height indicates percentage of the changing exons in each length category, and absolute numbers of changing exons are also indicated. Expression of genes containing these exons did not change more than expected by chance (data not shown). Grayscale heatmaps indicate overlap with exons in each length group that show AS changes upon knockdown of Srrm4 and with neural-differential exons.

(legend continued on next page)

and its target exons and that it also controls neural exons that are not regulated by Srrm4.

Nacc1 Regulates ES-Cell-Differential AS at Multiple Levels

The results from analyzing Zfp871 raised the intriguing possibility that additional transcription factors/DNA-binding proteins identified in our screen might also have dual direct and indirect AS regulatory activities. To investigate this, we next focused on the POZ/BTB family transcription factor, nucleus accumbens associated 1 (Nacc1). From the screen data, knockdown of Nacc1 has strongly correlated effects with knockdowns of Mbnl and Rbfox2 (Figure 6A; Figure S6A), which, as mentioned earlier, negatively regulate ES-cell-differential AS events and control somatic cell reprogramming (Han et al., 2013; Venables et al., 2013). Moreover, knockdown RNA-seq analysis revealed that Nacc1 has more widespread effects on ES-cell-differential AS, which correlate significantly with knockdown of Mbnl proteins ($\rho = 0.76$, rank correlation; $p < 2.2 \times 10^{-16}$; Figure 6B). The RNA-seq data further revealed that Nacc1 knockdown reduces the expression of Mbnl1 and, to a lesser extent, other splicing regulators, including Rbfox2 (Figure 6C), although there was no significant overlap between genes with AS or expression changes upon knockdown (data not shown). Consistent with these results, chromatin immunoprecipitation sequencing (ChIP-seq) analysis of Nacc1 occupancy in N2A cells shows that it binds proximal to the transcription start site of the *Mbnl1*, but not of the *Mbnl2* or *Rbfox2* genes (Figure 6D; data not shown). Similarly, Nacc1 peaks are generally enriched near transcription start sites of genes with changing expression upon Nacc1 knockdown, supporting a transcription regulatory role (Figure S6B). Collectively, these data provide evidence that Nacc1 indirectly controls ES-cell-differential AS by regulating the expression of Mbnl1 and possibly other splicing regulators.

Surprisingly, Nacc1 also binds RNA sequences surrounding its regulated target AS events. Similar to the results for Zfp871, using iCLIP analysis in N2A cells, Nacc1 frequently crosslinks to exonic RNA, but it also displays enriched occupancy over intronic sequences adjacent to its regulated target exons (Figure 6E; Figures S6C and S6D) ($p < 10^{-5}$, all comparisons, one-sided Mann-Whitney U test). Crosslinking-immunoprecipitation experiments employing transfected wild-type and mutant Nacc1 constructs confirm that Nacc1 binds RNA and further show that this activity is primarily dependent on a linker region located between the POZ/BTB and BEN domains (Figure S6E). These observations suggest that Nacc1 has a direct role in controlling the inclusion of ES-cell-differential exons.

To confirm whether Nacc1 has a direct role in regulating AS, we assayed bacterially expressed recombinant Nacc1 for activity in promoting target exon splicing in reporter transcripts in vitro. Increasing concentrations of Nacc1 stimulated the inclusion of an alternative exon from the Myo9b gene, which, based on the analysis of knockdown RNA-seq and iCLIP data, is regulated by Nacc1 (Figure 6F; Figures S6F and S6G). In contrast, addition of comparable levels of recombinant PTBP1 or BSA did not significantly affect splicing levels. Moreover, recombinant Nacc1 did not promote splicing of a neural-specific exon (Figures 6G). Finally, consistent with its RNA binding activity, the linker region of Nacc1 is capable of promoting exon inclusion in vitro, although with reduced activity compared to the full-length protein (Figure S6H). However, deletion of this domain did not abolish the splicing stimulatory activity of Nacc1, indicating that multiple domains in the protein, at least in vitro, likely forge interactions that promote exon inclusion (Figure S6H). Collectively, these results provide evidence that Nacc1, like Zfp871, has dual indirect and direct roles in the regulation of AS events linked to cell fate.

DISCUSSION

The SPAR-seq system described in this study generates a highly quantitative, sequencing-based readout for dozens of endogenous regulatory events in response to thousands of query conditions. By applying SPAR-seq to discover regulatory networks that control cell-fate-associated AS in mouse ES and neural cells, we have identified extensive positive and negative functional inter-relationships between *trans*-acting factors associated with different gene regulatory layers. An unexpected observation in neural cells is that annotated transcription, chromatin, and DNA-binding domain proteins impact AS at a similar frequency as splicing factors and that a subset of these protein factors dually control cell-fate-associated AS networks through direct and indirect mechanisms.

Transcription and chromatin regulators impact AS by various mechanisms, including the recruitment of splicing components that subsequently influence AS in nascent transcripts, and by effects on pol II elongation rate that control AS by altering the kinetics of exposure of competing splice sites (Braunschweig et al., 2013). Our results provide evidence that certain transcription factors also regulate AS networks by binding RNA sequences adjacent to target exons while also impacting the expression of splicing regulators that control the same target exons. These observations add to a growing body of evidence indicating that proteins lacking canonical RNA binding domains, including those linked to transcription and chromatin regulation, can interact with RNA (Baltz et al., 2012; Castello et al., 2012; G Hendrickson et al., 2016; Kwon et al., 2013).

(C) Representative RT-PCR validations of AS events identified by SPAR-seq and/or RNA-seq that are co-regulated by Zfp871 and Srrm4, that are regulated independently of Srrm4 (i), or that are not regulated by either protein (c-). Additional AS events are shown in Figure S5A.

(D) GO-enrichment analysis of genes with exons and microexons that show AS changes following knockdown of Zfp871. Abscissa shows the enrichment odds ratio (see STAR Methods for details).

(E) Change in mRNA expression of splicing factors upon knockdown of Zfp871 or Srrm4. Factors that were knocked down are indicated in dark gray, and blue indicates significant differences (FDR < 0.05).

(F) Average iCLIP signal of Flag-Zfp871 in N2A cells around exons with increasing, unchanged, or decreasing inclusion upon knockdown of Zfp871, as analyzed by RNA-seq. Inserts show median and inter-quantile range of mean intronic CLIP signals. Asterisks indicate p values of one-sided Mann-Whitney U tests. See also Figure S5.

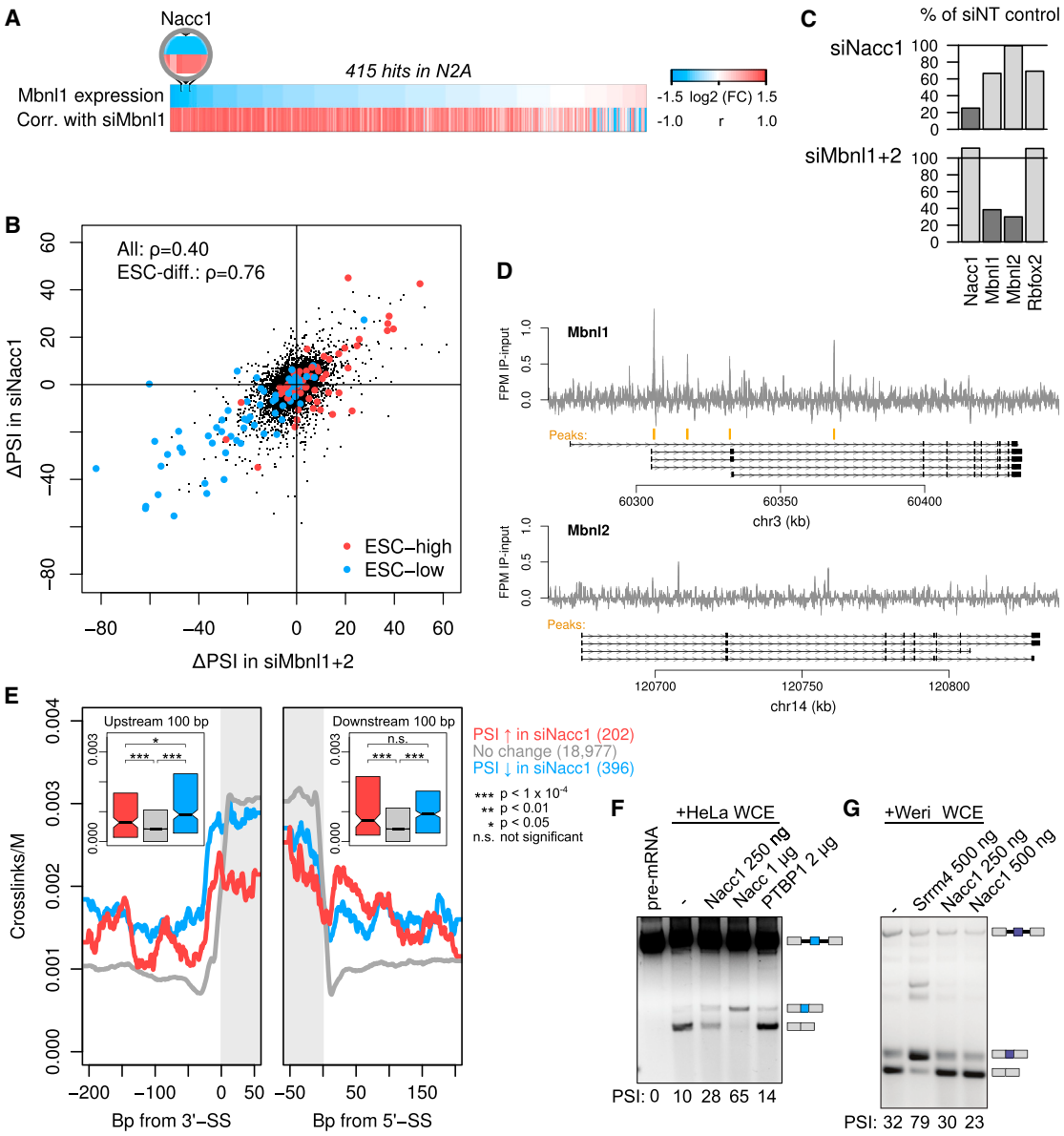


Figure 6. Nacc1 Regulates a Network of ES-Cell-Differential AS Events

(A) Effects of N2A screen hit knockdowns (|SSMD| > 4.75) on Mbnl1 mRNA expression, and correlation of AS changes with those observed upon knockdown of Mbnl1, sorted according to changes in Mbnl1 expression.

(B) Splicing changes of alternative exons observed upon knockdown of Mbnl or Nacc1. Annotated ES-cell-differential events (Han et al., 2013) are highlighted.

(C) mRNA expression changes upon knockdown of Nacc1 or Mbnl in N2A cells analyzed by RNA-seq. Dark gray represents genes that are targeted by knockdown.

(D) ChIP-seq profiles of Nacc1 in N2A cells at Mbnl1 and Mbnl2 genes. FPM, ChIP fragments per million reads.

(E) Average Nacc1 iCLIP signal in N2A cells around exons with increased, unchanged, or decreased inclusion upon knockdown of Nacc1, as analyzed by RNA-seq. Inserts show median and inter-quantile range of mean intronic CLIP signals. See also Figures S6C and S6D.

(F) In vitro splicing of *Myo9b* minigene reporter transcripts in HeLa whole-cell extracts, with or without the addition of recombinant Nacc1 or PTBP1 as negative control. PSI quantitation is shown below.

(G) In vitro splicing of *Daam1* minigene reporter transcripts (as a specificity control) (Calarco et al., 2009) in Weri whole-cell extracts, with or without the addition of recombinant Srrm4 or Nacc1. See also Figure S6.

Remarkably, knockdown of more than 50% of ZnF genes analyzed in our screen resulted in pronounced effects on different subsets of AS events. Specific classes of ZnF proteins, such as those possessing CCCH-type motifs, include members with established roles in RNA binding and regulation (e.g., U2AF1 and MBNL proteins) (Konieczny et al., 2014; Singh and

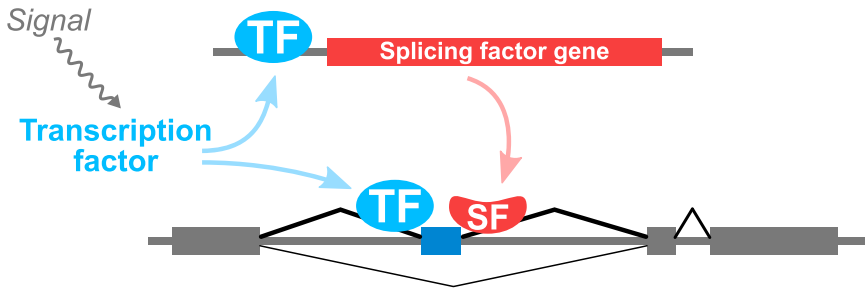


Figure 7. Model of Dual Roles of Transcription Factors in Regulating AS
Changes in the activity of a transcription factor (TF) results in increased expression of splicing factors (SFs), which, in turn, mediate AS changes. Additionally, binding of the transcription factor to pre-mRNA coordinately regulates overlapping AS events.

Valcárcel, 2005). In contrast, while C2H2 ZnF proteins constitute the largest class of annotated DNA-binding proteins, only a small number of these proteins have been investigated for such roles. Examples include Gtf3a/TFIIIA (Layat et al., 2013), HZF/Znf385a (Iijima et al., 2005), CTCF (Saldaña-Meyer et al., 2014), YY1 (Sigova et al., 2015), and WT1 (Hastie, 2001). Our results provide evidence for a more widespread role for this class of ZnF protein in splicing control. It is interesting to consider that the expansion of this class of proteins during vertebrate evolution, which occurred in part to fulfill roles in transcriptional regulation and the suppression of rapidly evolving transposable elements (Stubbs et al., 2011), may also have arisen to regulate increasingly complex patterns of AS associated with the evolution of organs such as the mammalian brain (Barbosa-Morais et al., 2012).

Finally, we show that the paradigm of DNA binding and transcription factors multitasking at the level of RNA to regulate AS extends to other classes of proteins, such as the BTB/POZ domain protein Nacc1. The remarkable duality in regulatory capacities of the two factors investigated in more detail in the present study suggests that additional transcription factors possess coordinated regulatory functions that operate indirectly and directly to regulate AS (Figure 7). The SPAR-seq screen thus highlights annotated transcription factors and DNA-binding proteins as a cache of previously unknown regulators of AS networks, including those with important roles in the control of cell fate. Further exploration of the splicing regulators identified in this study holds promise for the discovery of mechanisms and networks of AS regulation with critical roles in development and disease. Moreover, the flexibility of the SPAR-seq system opens the door to the comprehensive elucidation of RNA regulatory networks in diverse other mechanistic and biological contexts.

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
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines and Cell Culture
- METHOD DETAILS
  - High-Throughput siRNA Knockdown and RNA Purification
  - Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing
  - Events Monitored in the Screen
  - Selection of Candidate Splicing-Related Genes
  - siRNA Knockdown for RNA-Seq Experiments
  - ChIP-Seq Experiments
  - iCLIP Experiments
  - Crosslinking-Immunoprecipitation Experiments
  - In Vitro Splicing Assays
  - Protein Extraction and Western Blotting
  - Co-Immunoprecipitation Assay
  - RNA Extraction and (q)RT-PCR assays
  - Cloning and Plasmids
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Pipeline for High-Throughput Screen Data Analyses
  - Correlation Network Analyses and Functional Cluster Identification
  - Hit Frequencies of Proteins with Certain Domains
  - Principal Component Analysis
  - Analysis of Correlation within CORUM Complexes
  - Analysis of RNA-Seq Data
  - Analysis of ChIP-Seq Data
  - Analysis of iCLIP Data
- DATA AND SOFTWARE AVAILABILITY
  - Data Access
  - Software

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.01.011.

AUTHOR CONTRIBUTIONS

H.H., U.B., and B.J.B. designed the study with input from J.M., J.L.W., and J.F.G. H.H. developed and performed SPAR-seq screens, with contributions from U.B., F.V., and A.D., and performed follow-up experiments. U.B. analyzed SPAR-seq, RNA-seq, iCLIP-seq, and ChIP-seq data, with input from H.H., R.J.W., T.G.-P., T.S.-W., K.C.H.H., and B.J.B. T.G.-P. and S.N.-S. performed iCLIP experiments and in vitro splicing assays. R.J.W. performed correlation and functional network analyses. C.L.H. and D.O. performed screen-hit characterization. E.R. performed ChIP experiments. J.W. performed RT-PCR validation experiments. Q.P. designed primers. D.R., H.Z., and T.R.H. generated recombinant proteins. L.M and C.L.C. contributed to experiments characterizing Arglu1. B.J.B., H.H., and U.B. wrote the manuscript with input from the other authors.

ACKNOWLEDGMENTS

T.G.-P. and R.J.W. contributed equally to this study. We thank Razvan Nutiu, Nuno Barbosa-Morais, and Kevin Brown for assistance with the development

of SPAR-seq; Dax Torti, Danica Leung, Tanja Durbic, and Graham O'Hanlon in the Donnelly Sequencing Centre for sequencing; Troy Ketela, Patricia Mero, and Thomas Sun for assistance with robotics; and Tara Paton, Jo-Anne Herbrick, and Steve Scherer at the Hospital for Sick Children Toronto for access to PCR machines. This research was supported by grants from the CIHR (B.J.B, J.M., J.L.W., T.R.H., and J.F.G.), NIH (T.R.H., B.J.B., and J.F.G.), and Stem Cell Network (B.J.B.). H.H. was supported by an Ontario Graduate Scholarship and a University of Toronto Open Fellowship; U.B. was supported by EMBO and Human Frontier Science Program Long Term Fellowships; T.G.-P. was supported by EMBO and OIRM fellowships; R.J.W. was supported by CIHR postdoctoral and Marie Curie IOF fellowships; K.C.H.H. was supported by an Ontario Graduate Scholarship and a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship. B.J.B. holds the University of Toronto Banbury Chair in Medical Research.

Received: July 7, 2016
Revised: November 16, 2016
Accepted: January 5, 2017
Published: February 2, 2017

REFERENCES

Albulescu, L.O., Sabet, N., Gudipati, M., Stepankiw, N., Bergman, Z.J., Huffaker, T.C., and Pleiss, J.A. (2012). A quantitative, high-throughput reverse genetic screen reveals novel connections between Pre-mRNA splicing and 5′ and 3′ end transcript determinants. PLoS Genet. *8*, e1002530.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., and Therrien, M. (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251–262.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.

Berriz, G.F., Beaver, J.E., Cenik, C., Tasan, M., and Roth, F.P. (2009). Next generation software for functional trend analysis. Bioinformatics *25*, 3043–3044.

Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. Bioinformatics *27*, 2463–2464.

Boucher, L., Ouzounis, C.A., Enright, A.J., and Blencowe, B.J. (2001). A genome-wide survey of RS domain proteins. RNA *7*, 1693–1701.

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.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. *24*, 1774–1786.

Calarco, J.A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinas, 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.

Cassiday, L.A., and Maher, L.J., 3rd (2002). Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. *30*, 4118–4126.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. *10*, 741–754.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcárcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nat. Struct. Mol. Biol. *14*, 620–629.

Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. PLoS Comput. Biol. *6*, e1001016.

Daguenet, E., Dujardin, G., and Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO Rep. *16*, 1640–1655.

De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Lührmann, R., and Pena, V. (2015). The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. Nat. Struct. Mol. Biol. *22*, 138–144.

Eldridge, A.G., Li, Y., Sharp, P.A., and Blencowe, B.J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc. Natl. Acad. Sci. USA *96*, 6125–6130.

Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS Genet. *5*, e1000325.

Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science *315*, 972–976.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. *15*, 689–701.

G Hendrickson, D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. Genome Biol. *17*, 28.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132–146.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods *5*, 235–237.

Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241–245.

Hastie, N.D. (2001). Life, sex, and WT1 isoforms–three amino acids can make all the difference. Cell *106*, 391–394.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

Heyd, F., and Lynch, K.W. (2011). Degrade, move, regroup: signaling control of splicing proteins. Trends Biochem. Sci. *36*, 397–404.

Hirsch, C.L., Coban Akdemir, Z., Wang, L., Jayakumaran, G., Trcka, D., Weiss, A., Hernandez, J.J., Pan, Q., Han, H., Xu, X., et al. (2015). Myc and SAGA rewire an alternative splicing network during early somatic cell reprogramming. Genes Dev. *29*, 803–816.

Hubbard, K.S., Gut, I.M., Lyman, M.E., and McNutt, P.M. (2013). Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. F1000Res. *2*, 35.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274–287.

Iijima, T., Imai, T., Kimura, Y., Bernstein, A., Okano, H.J., Yuzaki, M., and Okano, H. (2005). Hzf protein regulates dendritic localization and BDNF-induced

translation of type 1 inositol 1,4,5-trisphosphate receptor mRNA. Proc. Natl. Acad. Sci. USA *102*, 17190–17195.

Irimia, M., and Blencowe, B.J. (2012). Alternative splicing: decoding an expansive regulatory layer. Curr. Opin. Cell Biol. *24*, 323–332.

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.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. Nat. Rev. Genet. *12*, 715–729.

Kolle, G., Shepherd, J.L., Gardiner, B., Kassahn, K.S., Cloonan, N., Wood, D.L., Nourbakhsh, E., Taylor, D.F., Wani, S., Chy, H.S., et al. (2011). Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. Genome Res. *21*, 2014–2025.

Konieczny, P., Stepniak-Konieczna, E., and Sobczak, K. (2014). MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res. *42*, 10873–10887.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

Lallena, M.J., Chalmers, K.J., Llamazares, S., Lamond, A.I., and Valcárcel, J. (2002). Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. Cell *109*, 285–296.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Layat, E., Probst, A.V., and Tourmente, S. (2013). Structure, function and regulation of Transcription Factor IIIA: from Xenopus to Arabidopsis. Biochim. Biophys. Acta *1829*, 274–282.

Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. *11*, 75–87.

Lin, S., and Fu, X.D. (2007). SR proteins and related factors in alternative splicing. Adv. Exp. Med. Biol. *623*, 107–122.

Lu, D., Searles, M.A., and Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. Nature *426*, 96–100.

Lu, X., Göke, J., Sachs, F., Jacques, P.E., Liang, H., Feng, B., Bourque, G., Bubulya, P.A., and Ng, H.H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. Nat. Cell Biol. *15*, 1141–1152.

Lu, Y., Loh, Y.H., Li, H., Cesana, M., Ficarro, S.B., Parikh, J.R., Salomonis, N., Toh, C.X., Andreadis, S.T., Luckey, C.J., et al. (2014). Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. Cell Stem Cell *15*, 92–101.

McCracken, S., Longman, D., Marcon, E., Moens, P., Downey, M., Nickerson, J.A., Jessberger, R., Wilde, A., Caceres, J.F., Emili, A., and Blencowe, B.J. (2005). Proteomic analysis of SRm160-containing complexes reveals a conserved association with cohesin. J. Biol. Chem. *280*, 42227–42236.

Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., et al. (2012). Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. Mol. Cell. Biol. *32*, 954–967.

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.

Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M., et al. (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nat. Biotechnol. *33*, 555–562.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Ohta, S., Nishida, E., Yamanaka, S., and Yamamoto, T. (2013). Global splicing pattern reversion during somatic cell reprogramming. Cell Rep. *5*, 357–366.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. *40*, 1413–1415.

Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. Mol. Cell *57*, 7–22.

Pelham, H.R., and Brown, D.D. (1980). A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc. Natl. Acad. Sci. USA *77*, 4170–4174.

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.

Rao, S., Zhen, S., Roumiantsev, S., McDonald, L.T., Yuan, G.C., and Orkin, S.H. (2010). Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. Mol. Cell. Biol. *30*, 5364–5380.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, W193-200.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes–2009. Nucleic Acids Res. *38*, D497–D501.

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.

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proc. Natl. Acad. Sci. USA *107*, 10514–10519.

Savage, K.I., Gorski, J.J., Barros, E.M., Irwin, G.W., Manti, L., Powell, A.J., Pellagatti, A., Lukashchuk, N., McCance, D.J., McCluggage, W.G., et al. (2014). Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. Mol. Cell *54*, 445–459.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978–981.

Singh, R., and Valcárcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nat. Struct. Mol. Biol. *12*, 645–653.

Sonnhammer, E.L., and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. *43*, D234–D239.

Stoilov, P., Lin, C.H., Damoiseaux, R., Nikolic, J., and Black, D.L. (2008). A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. Proc. Natl. Acad. Sci. USA *105*, 11218–11223.

Stubbs, L., Sun, Y., and Caetano-Anolles, D. (2011). Function and evolution of C2H2 zinc finger arrays. Subcell. Biochem. *52*, 75–94.

Tadepally, H.D., Burger, G., and Aubry, M. (2008). Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. BMC Evol. Biol. *8*, 176.

Tange, T.O., Nott, A., and Moore, M.J. (2004). The ever-increasing complexities of the exon junction complex. Curr. Opin. Cell Biol. *16*, 279–284.

Tejedor, J.R., Papasaikas, P., and Valcárcel, J. (2015). Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. Mol. Cell *57*, 23–38.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Uren, P.J., Bahrami-Samani, E., Burns, S.C., Qiao, M., Karginov, F.V., Hodges, E., Hannon, G.J., Sanford, J.R., Penalva, L.O., and Smith, A.D. (2012). Site identification in high-throughput RNA-protein interaction data. Bioinformatics *28*, 3013–3020.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth Edition (Springer).

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.

Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B., et al. (2013). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nat. Commun. *4*, 2480.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B., and Carstens, R.P. (2009). ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol. Cell *33*, 591–601.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., et al. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766–770.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394.

Zhang, X.D. (2007). A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. Genomics *89*, 552–561.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, D., Jiang, P., Xu, Q., and Zhang, X. (2011). Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. J. Biol. Chem. *286*, 17746–17754.

Zheng, S., Damoiseaux, R., Chen, L., and Black, D.L. (2013). A broadly applicable high-throughput screening strategy identifies new regulators of Dlg4 (Psd-95) alternative splicing. Genome Res. *23*, 998–1007.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-Nacc1	Abcam	ab29047; RRID: AB_870608
Critical Commercial Assays		
OneStep RT-PCR Kit	QIAGEN	210210
Deposited Data		
SPAR-Seq	This study	GEO: GSE80196
RNA-Seq	This study	GEO: GSE80204
ChIP-Seq	This study	GEO: GSE80203
CLIP-Seq	This study	GEO: GSE80202
Experimental Models: Cell Lines		
Mouse: CGR8	ECACC	07032901
Mouse: N2A	ATCC	CCL-131
Sequence-Based Reagents		
siGENOME mouse genomic siRNA SMARTpool library	Dharmacon	N/A
iCLIP reverse transcription primer Rt1clip: /5Phos/NNAAC CNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	Huppertz et al., 2014	N/A
iCLIP reverse transcription primer Rt9clip: /5Phos/NNGCC ANNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	Huppertz et al., 2014	N/A
iCLIP reverse transcription primer Rt10clip: /5Phos/NNGAC CNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	Huppertz et al., 2014	N/A
iCLIP reverse transcription primer Rt13clip: /5Phos/NNTCC GNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC	Huppertz et al., 2014	N/A
Software and Algorithms		
Bowtie 2.2.6	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/
Tophat 2.1.0	Trapnell et al., 2009	https://ccb.jhu.edu/software/tophat/
MACS 2.1.0	Zhang et al., 2008	https://github.com/taoliu/MACS
Piranha 1.2.1	Uren et al., 2012	http://smithlabresearch.org/software/piranha/
vast-tools 1.0	Braunschweig et al., 2014; Irimia et al., 2014	https://github.com/vastgroup/vast-tools

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Prof. Benjamin J. Blencowe, Donnelly Centre, University of Toronto, b.blencowe@utoronto.ca.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines and Cell Culture

CGR8 mouse embryonic stem cells (ESCs) were cultured as described previously on gelatin-coated plates (Gabut et al., 2011; Han et al., 2013). Mouse neuroblastoma (N2A) cells were grown in DMEM supplemented with 10% FBS, sodium pyruvate, MEM non-essential amino acids, and penicillin/streptomycin. All cell lines were maintained at 37°C with 5% CO_2.

METHOD DETAILS

High-Throughput siRNA Knockdown and RNA Purification

Knockdown and control treatments were performed in both mouse ESCs (CGR8) and neuroblastoma cells (N2A), in two biological replicates, at the Lunenfeld-Tanebaum Research Institute (LTRI) SMART robotics facility. Treatments comprised SMARTpool

siRNAs (siGENOME, Dharmacon) targeting 1416 genes that function in diverse aspects of gene regulation, 32 positive controls (siRNA targeting Mbnl1 and Mbnl2, siMbnl), and 88 negative controls (non-targeting siRNA, mock transfection, and untreated cells).

An automated pipeline was developed for high-throughput cell plating, siRNA transfection, and RNA purification (Biomek FX Laboratory Automation Workstation, Beckman Coulter). Twenty-four hours prior to transfection, CGR8 and N2A cells were seeded in 96-well plates, using 3000 and 5000 cells per well, respectively. Cells were transfected with the SMARTpool siRNAs at 50 nM final concentration using DharmaFECT1 reagent (Dharmacon), as recommended by the manufacturer. Forty-eight hours post-transfection, total RNA was purified from cultured cells using the RNeasy Plus 96 Kit (QIAGEN), as per the manufacturer's instructions. In total, 6144 (sixty-four 96-well plates) RNA samples (~65 μL per well) were prepared from CGR8 and N2A cells from two replicates.

Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing

Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode Sequencing (SPAR-seq) was developed for the parallel analysis of dozens of endogenous alternative splicing (AS) events in response to thousands of knockdown and control treatments. For each treatment, a multiplex RT-PCR assay was applied to simultaneously amplify 50 transcript regions that span multiple exons to assess AS and gene expression in a single reaction. In some cases, more than one AS event was monitored from the same region. Optimized event-specific primers with 5′ universal adaptor sequence were used (Figure 1A; Figure S1A). Primers for AS events were designed to anneal near splice junctions in order to monitor different spliced variants (Table S1 for primer sequences). The multiplex RT-PCR reaction was carried out in 96-well plates using the OneStep RT-PCR kit (QIAGEN) as recommended by the manufacturer, with the following changes: reactions were performed in a volume of 20 μL with 2 μL of the purified total RNA as input, and a mixture of 50 pairs of primers was added to each reaction at a final concentration of 0.025 μM for each individual forward and reverse primer. Four identical Veriti 96-well Thermal Cyclers (Applied Biosystems) were used with the following program: 50°C for 30 min, 95°C for 15 min, 30 cycles of 94°C for 40 s, 58°C for 1 min (slow ramp rate), 72°C for 3 min, and a final extension step at 72°C for 10 min.

For multiplex barcode sequencing, unique, dual-index barcodes were designed, including 16 forward 8-base barcodes (minimum Hamming distance of 4) and 768 reverse 8-base barcodes, comprising a subset of previously reported barcode sequences (Hamady et al., 2008). Reverse barcodes were selected to further increase the distance between barcodes, and minimize hairpin structures and primer dimers. To multiplex 768 samples per sequencing lane in the current screen (see below), unique reverse barcodes were used for each sample, while forward barcodes were used to mark each half (48 samples) of a 96-well plate to provide additional redundancy. These two sets of barcodes were incorporated into forward and reverse primers, respectively, after the universal adaptor sequences and were added to the amplicons in the second PCR reaction, which was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific), as per the manufacturer's instructions. For each 20 μL of reaction, 1 μL of the multiplex RT-PCR reaction product was used as template. The thermal cycling conditions were as follows: 98°C for 30 s, 15 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 30 s, and a final extension step at 72°C for 5 min.

To achieve high-throughput preparation of barcode sequencing libraries for the screen, a Versette Automated Liquid Handler (Thermo Scientific) was programmed to set up both the multiplex RT-PCR reactions and the second Phusion PCR reactions in 96-well plates. The resulting libraries were pooled and sequenced at the Donnelly Sequencing Centre. The entire screen monitored 52 AS events from 6144 samples, as described above. An Illumina HiSeq 2500 flow cell with 8 lanes was used for the barcode sequencing, and 768 samples from eight 96-well plates were multiplexed per lane. In total, approximately 2.1 billion 101-bp paired-end reads and two separate 8-bp index reads were generated.

Events Monitored in the Screen

The majority of the monitored AS events are ESC-differential AS events defined by large-scale RNA-Seq analyses (Han et al., 2013) and other studies (Gabut et al., 2011; Kolle et al., 2011; Ohta et al., 2013; Rao et al., 2010; Salomonis et al., 2010; Venables et al., 2013), some of which have been shown to be important for ESC pluripotency, differentiation, and somatic cell reprogramming. Multiple non-ESC-specific AS events (e.g., neural-regulated exons and microexons) (Irimia et al., 2014; Raj et al., 2014) and gene expression only events (e.g., splicing factor and internal control genes) were also monitored (see Table S1 for further details).

Selection of Candidate Splicing-Related Genes

Candidate genes for knockdowns (Table S2) were selected using multiple approaches. Genes that encode proteins with known or putative roles in splicing and RNA processing were curated manually from the literature and comprised all known splicing factors, including spliceosome-associated proteins as well as known and predicted RNA binding proteins (RBPs). In addition, genes encoding chromatin-related proteins and transcription factors were selected using four approaches: (1) Genes whose involvement in splicing regulation has been reported previously; (2) Genes with the GO Slim annotation 'chromosome', to capture chromatin-associated proteins; (3) Genes encoding proteins with domains that are involved in chromatin modification/binding/remodeling, based on literature and PFAM domain descriptions; (4) Genes encoding proteins with a domain composition similar to known chromatin proteins. The latter set was identified as follows: the presence of PFAM domains in the protein encoded by the transcript with most domains of each gene was tabulated. Genes were labeled as positive if associated with the GOslim term 'chromosome', and negative if not associated with it but associated with at least one term related to cytoplasm, extracellular region, plasma membrane, or translation. Two-thirds of the labeled set were used to train a support vector machine with ten-fold cross-validation using the

svm() function in the R package e1071 (https://cran.r-project.org/web/packages/e1071/). After parameter optimization, a set of parameters that yielded a true positive rate of 0.479, false positive rate of 0.028, and accuracy of 0.945 on the training set was used to train a final model on the full labeled set of genes to predict association with 'chromosome' for unlabeled genes. Genes positively labeled by this prediction were associated with the term 'nucleus' > 3.5 times more often than negatively labeled genes. Of 1,925 pre-selected genes from all four streams, 836 were finally selected based on minimum expression cut-off, measured using RNA-Seq data) in ESCs or N2A cells and availability of SMARTpool siRNAs. Lastly, signaling components and post-translational factors that have been previously linked to splicing regulation were also included.

For all categories, a total of 1,416 genes were selected for knockdown. The four groups shown in Figure 1C and Table S2 represent a post hoc categorization based on the approaches described above but prioritizing RNA binding domains, association with the spliceosome, experimental detection as part of the mRNA interactome (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013), GO annotation "RNA binding" or "RNA processing" for the splicing factor/RBP group, and the occurrence of a C2H2 domain.

siRNA Knockdown for RNA-Seq Experiments

Cells were transfected with SMARTpool siRNAs (Dharmacon) using Lipofectamine RNAiMAX (Invitrogen), as recommended by the manufacturer. A non-targeting siRNA pool (siNT) was used as a control. Cells were harvested 48 hr post transfection, and total RNA was extracted with TRI Reagent (Sigma) or RNeasy columns (QIAGEN).

ChIP-Seq Experiments

Chromatin immunoprecipitation (ChIP) was performed as previously described (Najafabadi et al., 2015). Briefly, ~20 million N2A cells were crosslinked in 1% formaldehyde. Following sonication of DNA fragments, Nacc1 was immunoprecipitated from the lysate with 4 μg of Nacc1 antibody (Abcam ab29047) followed by crosslink reversal and DNA precipitation. Libraries were sequenced on the Illumina HiSeq 2500 to a depth of 20 million 51-nucleotide single end reads.

iCLIP Experiments

Individual nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) was performed as described previously (Huppertz et al., 2014). Zfp871 was immunoprecipitated from N2A cells engineered to express Flag-tagged Zfp871 under the control of doxycyclin using the PiggyBac system (Woltjen et al., 2009). Flag-Zfp871 expression was induced for 24 hr prior to crosslinking (0.15 J/cm^2) at 254 nm with a Stratalinker 1800. Two replicates from two independent single-cell clones expressing Flag-tagged Zfp871 were used for generating a total of four iCLIP samples. Lysates generated from the crosslinked cells were treated with Turbo DNase (Ambion) and RNase I (1:100; Ambion) for 5 min 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 μL of protein G Dynabeads (Life Technologies) and 10 μg of anti-Flag (Sigma) antibody. Following stringent high salt washes, the immunoprecipitated RNA was 5′ end-labeled using radioactive ^{32}P isotopes followed by on-bead-ligation of pre-adenylated adaptors to the 3′ end. 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 enzyme 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 on PCR purification columns (QIAGEN), eluted DNA was mixed at a ratio of 1:5:5 from the low, middle, and high fractions and submitted for sequencing.

Nacc1 iCLIP was performed in the same way, except for the following changes: Nacc1 was immunoprecipitated from two independent N2A cell pellets using anti-Nacc1 antibody (Abcam ab29047). Prior to immunoprecipitation, cells were UV cross-linked with 0.4 J/cm^2 at 254 nm with a Stratalinker 1800 and lysates were treated with Turbo DNase and RNase I (1:500) for 5 min at 37°C. A total of 2% input material was saved to prepare size-matched control libraries. Furthermore, we used a recently reported enhanced CLIP ligation method (Van Nostrand et al., 2016) for adaptor ligation. The immunoprecipitated material as well as the input were separated with SDS-PAGE and transferred to a nitrocellulose membrane (Protran). For the input sample, the membrane was cut matching the size of the immunoprecipitated material. Consecutive steps proceeded as described above.

For Flag-Zfp871 the barcoded primers used were: Rt1clip: /5Phos/NNAACCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC; Rt9clip: /5Phos/NNGCCANNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC; Rt10clip: /5Phos/NNGACCNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC; and Rt13clip: /5Phos/NNTCCGNNNAGATCGGAAGAGCGTCGTGgatcCTGAACCGC.

For Nacc1 the barcoded primers used were Rt1clip, Rt10clip, and Rt13clip.

Crosslinking-Immunoprecipitation Experiments

To investigate the region(s) of Nacc1 protein responsible for RNA binding, we divided the full-length protein into three fragments: the N-terminal fragment (F1, amino acids 2-132) including the POZ/BTB domain, the middle fragment (F2, amino acids 133-351), and the C-terminal fragment (F3, amino acids 352-514) including the BEN domain. Using the PiggyBac system (Woltjen et al., 2009), N2A cells were engineered to overexpress Flag-tagged full-length or different truncated forms of Nacc1 under the control of doxycyclin. Subsequently, crosslinking was performed in the same way as the initial steps of iCLIP. Lysates generated from the crosslinked cells were treated with Turbo DNase and RNase I (1:500) for 5 min at 37°C. RNA-protein complexes were immunoprecipitated using protein G

Dynabeads and anti-Flag antibody. Following high salt washes, the immunoprecipitated RNA was 5′ end-labeled using radioactive ^{32}P isotopes. The immunoprecipitated complexes were separated with SDS-PAGE and transferred to a nitrocellulose membrane.

In Vitro Splicing Assays

Preparation of whole cell splicing extracts and purification of recombinant proteins have been previously described in detail (Calarco et al., 2009). The Myo9b in vitro splicing reporter was constructed by amplifying the mouse genomic DNA region, including the specific alternative exon, its flanking introns, as well as its constitutive exons. In vitro splicing assays performed in a volume of 20 μL contained 1.5 mM ATP, 5 mM creatine phosphate, 5 mM DTT, 3 mM $MgCl_2$, 2.6% PVA, 30 units of RiboLock RNase inhibitor (Thermo Scientific), 20 ng of splicing substrate, 50-60 μg of splicing extract, and up to 12 μL of splicing buffer (20 mM HEPES-KCl pH 7.9, 100 mM KCl, 0.2 mM EDTA, 10% glycerol, 1 mM DTT) with or without the addition of recombinant proteins. All recombinant Nacc1 proteins contained an N-terminal GST-tag and were expressed and purified from *Escherichia coli* C41 (Lucigen) using standard methodologies. Reactions were incubated at 30°C for one hour. RNA was extracted using TRI Reagent (Sigma) and then resuspended in 10 μL of DEPC-treated water. Spliced products were amplified by RT-PCR assays using 2 μL of the recovered RNA and primers specific for Myo9b upstream and downstream constitutive exons. RT-PCR products were resolved on a 3% agarose gel.

Protein Extraction and Western Blotting

Cell pellets were lysed in radioimmunoprecipitation assay (RIPA) buffer by brief sonication. 30-150 μg of protein lysate was separated on a 10% SDS-polyacrylamide gel and transferred to a PVDF membrane. The membranes were blotted with the following antibodies: anti-Flag M2 (1:1500, Sigma), anti-Nacc1 (1:7500, Abcam), and anti-α-tubulin (1:5000, Sigma). Secondary antibodies (GE Healthcare) and chemiluminescence reagents (Perkin Elmer) were used as per the manufacturer's instructions.

Co-Immunoprecipitation Assay

293T cells were transiently transfected with FLAG-Srsf2, HA-Arglu1, or both using Lipofectamine 2000 (Life Technologies). Cells were lysed in 0.5% TNTE buffer containing 250 U/μL Benzonase (Novagen) and incubated on ice for one hour. After pre-clearing with Protein G-Dynabeads (Life Technologies), lysates were incubated with anti-Flag M2 antibody (Sigma) bound to Protein G-Dynabeads for 1 hr at 4°C. Subsequently, immunoprecipitates were washed 5X with 0.1% TNTE buffer, subjected to SDS-PAGE, transferred onto a nitrocellulose membrane, and immunoblotted with the anti-HA antibody (Roche) or anti-Flag M2 antibody (Sigma).

RNA Extraction and (q)RT-PCR assays

Total RNA was extracted using RNeasy Mini Kit (QIAGEN) or TRI Reagent (Sigma), and RT-PCR assays were performed using the OneStep RT-PCR kit (QIAGEN), as per the manufacturer's instructions. 20 ng of total RNA was used per 10 μL of reaction. The number of amplification cycles was 22 for Gapdh, and 27-32 for all other transcripts analyzed. Reaction products were separated on 1%–3% agarose gels. Quantification of isoform abundance was performed as previously described (Han et al., 2013).

For qRT-PCR, first-strand cDNAs were generated from 1-3 μg of total RNA using Maxima H Minus First Strand cDNA synthesis Kit (Thermo Scientific), as per the manufacturer's recommendations, and diluted to 20 μg/μL. qPCR reactions were performed in a volume of 10 μL using 1 μL of diluted cDNA and FastStart Universal SYBR Green Master (Roche Applied Science). Primers used for PCR reactions are available upon request.

Cloning and Plasmids

Flag-tagged *Zfp871* was cloned into the PiggyBac destination vector from cDNA derived from N2A cells using the Gateway system. The primers used for cloning are:

forward 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCGAGTCAGTGG CCTTTGAGGATGTG - 3′ and reverse 5′ - GGGGA CCACTTTGTACAAGAAAGCTG GGTCCTACTAAAAAACACTGGAATCCAGGGTAGTG - 3′.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pipeline for High-Throughput Screen Data Analyses

De-multiplexing and Mapping

Sets of reads from each sequencing lane, consisting of forward/reverse event reads and forward/reverse barcode reads, were assigned to one of 768 samples by matching the forward and reverse barcodes to the expected combinations. Up to two mismatches were allowed if and only if these allow a match to a single barcode. ~79% of the total read sets from 8 lanes were successfully assigned to the 6,144 samples. Forward and reverse event reads were then mapped to custom junction libraries representing all expected splice variants (available on GEO: GSE80196), using bowtie with settings–best -v 3 -k 1–trim3 26–trim5 20 (Langmead et al., 2009). Trimming of the first 26 and last 20 bases was performed to remove lower-quality and uninformative ends and thereby increase the rate of mappable reads. Junction libraries were first constructed from NCBIm37/mm9 gene annotations, and then refined based on the results of de novo mapping of the reads from one full Illumina lane from each cell line using TopHat (default settings with -i 40 and providing Ensembl transcript annotations for NCBIm37) (Trapnell et al., 2009), after quality trimming (minimum MAPQ of 33) from the 3′ end. Initially, 64 events within 50 genes were considered. In total, ~20% of the forward and reverse reads

were mapped uniquely to one splice variant. Across all 6,144 samples, the median number of reads per gene was ~300 in each direction, with ~91% reaching > 20 reads (Figure S1C).

AS Quantification and SSMD Calculation

Percent Spliced In (PSI) values were calculated for each alternative exon, or part thereof in the case of alternative 5′ or 3′ splice sites ('event'), independently from forward and reverse reads as the percentage of reads supporting inclusion divided by the total number of reads for the event. For alternative 5′ or 3′ splice site events that were part of an alternative exon [Foxm1 (A2), Mta1 (A1), Uspl1 (A3)], the PSI was instead calculated with reference to the total reads supporting inclusion of that exon in order to assess independent regulation of alternative splice site usage. The average of PSI values obtained from forward or reverse reads was calculated, except in cases where some forward or reverse reads were ambiguous with respect to either inclusion or exclusion due to insufficient read length, in which case only one read was used. PSI values showed low-level batch effects per 96-well plate, which were reduced by subtracting a weighted plate median in which negative controls (siNT and mock treatment) were given 20x more weight than other samples. To derive variances, and because PSI values are not nearly normally distributed but roughly follow a beta distribution, we elected to fit a beta distribution to each pair of replicate treatments, as well as to all negative controls from both replicates, using maximum-likelihood fitting as implemented in the fitdistr() function from the R package MASS (Venables and Ripley, 2002). Iterative optimization of shape parameters was initiated with settings x = PSI, shape1 = 1, shape2 = 1, method = "L-BFGS-B," lower = 0.01, and upper = mean number of reads supporting each PSI value. A modified Strictly Standardized Mean Difference (SSMD) (Zhang, 2007) was then calculated such that:

$$\mathrm{SSMD} = (\mu_t - \mu_c)/\mathrm{SQRT}(\mathrm{var}_t + \mathrm{var}_c)$$

where μ_t and μ_c are the means (corresponding to the PSI), and var_t and var_c are the variances of the beta distributions fitted to the treatment replicates and negative controls, respectively. Events for which not all reads from at least one direction were informative [Mff (A2), Tead1 (A2)] were excluded from further analysis, as were events in individual treatments with less than 20 reads in one or both replicates (~9% of all events x treatment combinations). Additionally, the following events were removed from further splicing analysis: gene expression only events or constitutive exons, Fgf4, Gapdh, Sall4, Srpk2, and Srrm4; events with consistently low read counts, Atg13 (A1), H2afy (A2), and Tcf7l1 (N2A only), and Dnmt3b (N2A only); or events where the measured inclusion was biased by differential length of isoforms (Fgfr1). Additionally, Dnmt3b in CGR8 cells and Uspl1 (A3) were excluded from most analyses due to missing values in > 10% of all treatments, resulting in a list of 52 filtered events. SSMD scores are provided in Table S3.

Differential Expression Analysis

Read counts from all splice variants of each gene were used to estimate relative mRNA expression levels, reads per million reads (RPMs). Differential expression analysis based on raw read counts was performed using the generalized linear model workflow from the R package edgeR (Robinson et al., 2010). Because biases stemming from batch effects and position of the well on the plate affected the expression analysis more than the PSI (which is the ratio of two measurements), we used plate, position (edge/interior), and treatment as design factors (where all treatments of each type of control were treated as replicates.) Models were fitted using estimateDisp() to estimate the common, trended, and tag-wise dispersion for the CGR8 and N2A data separately. Subsequently, differences attributable to treatment, plate, or position contrasts were extracted with glmLRT() and represented as log2-fold changes with associated FDR. This approach showed that it was important to model plate and position explicitly. Fold-changes for treatments (Table S4) are relative to the siNT and mock controls, while untreated controls were treated like experimental knockdowns and used to monitor the efficiency of normalization.

Correlation Network Analyses and Functional Cluster Identification

To identify the appropriate SSMD cut-off, a cross-validation resampling approach was undertaken based on identification of functional groups from Enrichr (Chen et al., 2013). This approach randomly selected 30% of the events exceeding a cut-off and clustered the results using affinity propagation clustering (Bodenhofer et al., 2011; Frey and Dueck, 2007). To select the number of clusters, a cut-off of h = 0.4 was used for the merging objects. Each cluster was individually assessed for GO term and complex enrichment using Enrichr. For each cut-off this procedure was repeated 1000 times independently for both N2A and CGR8 samples. The reproducibility of the top functional groups at each cut-off was calculated. This identified a cut-off of 2.25 within the CGR8 samples and 4.75 within the N2A samples as providing the strongest enrichment of associated terms.

Events with SSMD scores above the identified cut-off were used. These events were clustered based on the ΔPSI values using affinity propagation clustering (Bodenhofer et al., 2011; Frey and Dueck, 2007). A signed Pearson correlation was used to calculate pairwise correlation coefficients to dampen effects of diverging means and variances between samples. Gene enrichment for each cluster was assessed using the gene enrichment tool gprofiler (Reimand et al., 2007) using the following databases: GO, CORUM, REACTOME, and KEGG. A Benjamini-Hochberg FDR multiple correction with a p value cut-off of 0.05 only including genes sets with maximum set size of 1000.

To annotate events that differentiate each cluster, two-sided Mann-Whitney U-tests were carried out comparing the event PSI values of the genes knocked down in the cluster with the rest of the knockdowns represented in the figure. Differential events with Benjamini-Hochberg FDR < 0.05 are shown.

Hit Frequencies of Proteins with Certain Domains

The percentages of knockdowns that exceeded the threshold of ± 3.0 SSMD for any AS event in N2A cells, among all proteins containing a certain PFAM domain was calculated. Only knockdowns with missing values for less than half of all events were considered. Separation into 'known' and 'new' AS regulators was based on association with any GO category whose name contained either of the strings 'mRNA splic' or 'spliceosom'.

Principal Component Analysis

The R function *prcomp* was used to derive principal components of either CGR8 or N2A SSMD values (uncentered, unscaled) from experimental knockdowns but not controls in order to avoid dominating effects of positive controls. Treatments with more than 10 missing values in each cell line were excluded. The full datasets including controls were then projected onto the obtained principal components. Outlines shown in Figure S4C represent the convex hull surrounding all points in the group.

Analysis of Correlation within CORUM Complexes

To score complexes for which knockdown of subunits resulted in correlated AS changes, all human and mouse complexes annotated in CORUM (Ruepp et al., 2010) were considered. For complexes isolated in both organisms, only the mouse variant was considered. When a complex was identified only in human cells, the mouse orthologs defined in InParanoid (Sonnhammer and Östlund, 2015) were used. Complexes in which less than three components were represented in the screen were discarded. Then, the average pairwise correlation of SSMD values between all components in a complex was calculated, and significance was assessed with a Mann-Whitney U-test between the correlations among the components in the complex and the pairwise correlations of all non-control knockdowns in the cell line. The false discovery rate was adjusted using the Benjamini-Hochberg method.

Analysis of RNA-Seq Data

AS and Gene Expression Analysis

RNA-Seq data were processed using our AS and gene expression analysis pipeline, vast-tools version 1.0 (Braunschweig et al., 2014; Irimia et al., 2014), which is available on github (https://github.com/vastgroup/vast-tools). From primary output, events with poor coverage or junction balance were filtered out (vast-tools quality column *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 assessed through the vast-tools *diff* module available with the main pipeline. This strategy utilizes Bayesian inference, employing a Binomial likelihood function where the count of inclusion reads (K) follows $K \sim \text{Binomial}(\Psi, N)$. Ψ represents PSI or PIR, and N is the total junction reads per-event. We apply an uninformative conjugate prior distribution (uniform Beta where $\alpha = 1$, $\beta = 1$), and apply Bayes theorem to obtain the posterior distribution over $\Psi \sim \text{Beta}(K + \alpha, (N-K) + \beta)$. We combine biological replicates by sampling empirical posterior distributions of each replicate and fitting a new posterior Beta using maximum-likelihood (MLE) estimation with 'fitdistr' from the MASS package in R. The difference between two biological conditions, modeled as two posterior distributions $X \sim$Beta, and $Y \sim$Beta, follows in the form $P(X-Y > 0)$. This probability can be estimated from the difference of empirical distributions sampled between X and Y such that $P(X-Y > 0) = \Sigma^{n}_{i=1}(X_i - Y_i > 0) / N$. Significantly differential events were additionally required to have a PSI difference > 10. Gene expression differences were calculated based on vast-tools raw read counts per gene. In cases with a single replicate (siNacc1, siMbnl, and siNT control), counts were converted to read-per-million (RPM) and changes calculated as $\log_2((1 + \text{RPM[siSpecific]}) / (1 + \text{RPM[siNT]}))$, and genes were required to have a vast-tools cRPKM ≥ 3 and a raw read count of ≥ 10 in at least one of the compared samples. In cases with multiple replicates (siZfp871, siSrrm4, and siNT control), differential expression was assessed with the R package edgeR.

Gene Ontology Analyses

FuncAssociate (Berriz et al., 2009) was used to find over-represented GO terms associated with genes with changes in CE and MIC events that were significant and greater than 10 PSI. As a background, all genes with measured PSI values that survived filtering were used. Only terms with a minimum odds ratio of 4 and less than 1,000 associated genes were plotted. If two categories mutually overlapped by more then 70% of associated genes, only the category with stronger enrichment was shown. Adjusted P values represent P values derived by iterative simulation in FuncAssociate.

Analysis of Features Associated with Zfp871-Regulated Events

SVM-BPfinder (Corvelo et al., 2010) was used to assign the most likely branchpoint and associated polypyrimidine tract length of introns upstream of regulated and non-regulated exons. MaxEntScan (Yeo and Burge, 2004) was used to assess splice site strength.

Analysis of ChIP-Seq Data

Illumina adaptor sequences were removed from 3′ ends of 51-nt reads and remaining reads were mapped to the mouse genome, NCBIm37/mm9, using bowtie2 (Langmead and Salzberg, 2012) with default settings. After removal of duplicate reads, peaks were called jointly on immunoprecipitated and input samples with MACS 1.4 (Zhang et al., 2008).

Analysis of iCLIP Data

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 trimmed to 49 nt from the 3′ end, and duplicates were discarded. Reads were de-multiplexed, and the

random positions, barcodes, and any 3′-bases matching Illumina adaptors were removed. Remaining reads longer than 25 nt were mapped to the mouse genome/transcriptome (Ensembl annotation of NCBIm37) 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 borders were generated as described previously for ChIP-seq data (Braunschweig et al., 2014), except that reads were first reduced to their first position, which is adjacent to the crosslink position, and no normalization against a control was performed. A 21-bp running window average was used for display only, and average signals across replicates are shown.

For numeric analysis of intronic signal, the number of crosslinks per position and million of sequenced reads was calculated per replicate and averaged across replicates. Mean intronic signals were compared between groups of AS events using the one-sided Mann-Whitney U-test.

Overlap of AS events affected by Nacc1 with Nacc1 RNA binding was scored as follows: Piranha (Uren et al., 2012) was used to obtain CLIP clusters separately in each replicate, using aligned RNA-Seq reads from N2A cells as a covariate. Both sets of clusters were merged, and for each new cluster, significance of enrichment over input was assessed with a binomial test in which the number of successes was the number of (pooled) CLIP reads overlapping the cluster (only the most 5′ base of each read was considered); the number of trials was the sum of the numbers of reads in CLIP and input; and the probability was the fraction of uniquely aligned reads in the CLIP libraries over the number of reads from CLIP and input combined. Only clusters with a Benjamini-Hochberg corrected FDR < 0.1 were retained.

DATA AND SOFTWARE AVAILABILITY

Data Access

SPAR-seq (GEO: GSE80196), RNA-Seq (GEO: GSE80204), ChIP-Seq (GEO: GSE80203), and iCLIP-Seq (GEO: GSE80202) data are available through the Gene Expression Omnibus. Series record: GEO: GSE80205.

Software

All scripts were written in Python, Perl, or R and are available upon request.