Regulation of alternative splicing in *Drosophila* by 56 RNA binding proteins

Angela N. Brooks[1,5,6], Michael O. Duff[3], Gemma May[3], Li Yang[3], Mohan Bolisetty[3], Jane Landolin[4],

Ken Wan[4], Jeremy Sandler[4], Benjamin W. Booth[4], Susan E. Celniker[4],

Brenton R. Graveley[3*], Steven E. Brenner[1,2*]

[1]Department of Molecular and Cell Biology, [2]Department of Plant and Microbial Biology, University of

California, Berkeley, CA

[3]Department of Genetics and Genome Sciences, Institute for Systems Genomics, University of

Connecticut Health Center, Farmington, CT

[4]Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, CA

[5]Broad Institute, Cambridge, MA,

[6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston MA

Corresponding authors: Brenton R. Graveley: email: graveley@uchc.edu, Steven E. Brenner:

email:brenner@compbio.berkeley.edu

Keywords: *Drosophila*, alternative splicing, modENCODE, RNA-seq, RNAi

Abstract

Alternative splicing is regulated by RNA binding proteins (RBPs) that recognize pre-mRNA sequence elements and activate or repress adjacent exons. Here, we used RNA interference and RNA-seq to identify splicing events regulated by 56 *Drosophila* proteins, some previously unknown to regulate splicing. Nearly all proteins affected alternative first exons suggesting that RBPs play important roles in first exon choice. Half of the splicing events were regulated by multiple proteins, demonstrating extensive combinatorial regulation. We observed that SR and hnRNP proteins tend to act coordinately with each other, not antagonistically. We also identified a cross-regulatory network where splicing regulators affected the splicing of pre-mRNAs encoding other splicing regulators. This large-scale study substantially enhances our understanding of recent models of splicing regulation and provides a resource of thousands of exons that are regulated by 56 diverse RBPs.

Introduction

Most metazoan genes contain introns that are removed from their primary transcripts (pre-mRNA) by the spliceosome – a macromolecular machine composed of hundreds of proteins and five small RNAs – which joins flanking exons together to generate a mature mRNA. Pre-mRNAs are alternatively spliced when the spliceosome uses different splice sites, thus creating different mRNAs, and frequently proteins, from a single gene. Alternative splicing is an important aspect of gene regulation that is used to produce different transcript isoforms in a tissue- (Pan et al. 2008; Wang et al. 2008; Barbosa-Morais et al. 2012; Merkin et al. 2012; Brown et al. 2014) or temporal-specific manner (Graveley et al. 2011). Alternative splicing is primarily regulated by proteins that bind to the pre-mRNA and act to enhance or inhibit spliceosome assembly, typically at nearby splice sites (Nilsen and Graveley 2010). To date, only ~50 vertebrate RNA binding proteins have been identified as regulators of alternative splicing (Gabut et al. 2008) which is surprising given that there are ~100,000 alternative splicing events known in humans (Pan et al. 2008; Wang et al. 2008). To obtain a comprehensive understanding of how splicing is regulated, it is essential to identify splicing regulatory proteins and the specific splicing events they affect.

The best-characterized splicing regulatory proteins are members of the SR and hnRNP protein families (Chen and Manley 2009). These proteins play key roles in both constitutive and alternative splicing and act by several different mechanisms. SR proteins have been shown to help recruit the spliceosome to splice sites and to promote exon inclusion (Chen and Manley 2009), while hnRNPs are thought to primarily repress exon inclusion by either antagonizing SR proteins, directly inhibiting the spliceosome, or through a "looping out" mechanism of regulation (Chen and Manley 2009). Several studies have identified targets of *Drosophila* SR and hnRNP proteins and found very few overlapping effects between the two types of proteins (Blanchette et al. 2005; Blanchette et al.

2009). A recent study of *Drosophila* SR proteins found that most regulatory targets of SR proteins were co-regulated by another SR protein and that they coordinately promoted exon inclusion and skipping (Bradley et al. 2015). A study of hnRNP proteins in human kidney cells identified significant overlaps in the targets of hnRNPs (Huelga et al. 2012). This study also found examples of cross regulation where one hnRNP can regulate the splicing of pre-mRNAs encoding other hnRNPs (Huelga et al. 2012).

In addition to SR and hnRNP proteins, other RNA binding proteins including RBFOX2, NOVA1/2, and PTBP1 among others (Ule et al. 2006; Boutz et al. 2007; Yeo et al. 2009) have been shown to regulate alternative splicing in a tissue- or developmentally-regulated manner. These proteins, like SR and hnRNP proteins, contain one or more RRM or KH domains that typically interact with RNA in a sequence-specific manner (Perez-Canadillas and Varani 2001).

Unexpectedly, proteins that are core components of the spliceosome have been shown to modulate splicing in yeast and *Drosophila* in a substrate-specific manner (Park et al. 2004b; Pleiss et al. 2007). In the *Drosophila* study (Park et al. 2004b), the impact of core spliceosomal components was examined on only a few alternative splicing events. Thus, the genome-wide impacts of perturbing core components of the spliceosome in *Drosophila* remains to be determined.

Another set of proteins that appear to play an unexpected role in alternative splicing are components of the exon junction complex (EJC). The EJC proteins EIF4A3, TSU/RBM8A, and MAGO have been shown to regulate splicing (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010; Michelle et al. 2012). EJC components are deposited near the exon-junction after splicing where they assist in transport, localization, and translation of the processed mRNA (Tange et al. 2004; Isken and Maquat 2008). Additionally, EJC proteins are important in facilitating nonsense-mediated decay (NMD) of transcripts containing a premature stop codon (Tange et al. 2004; Chang et al. 2007).

4

To gain a more comprehensive understanding of the regulation of alternative splicing, we used RNA interference (RNAi) to deplete 56 proteins, including known splicing regulators such as SR and hnRNP proteins, or putative splicing regulators that contain RNA binding domains, core spliceosomal proteins, and EJC components, in *D. melanogaster* S2-DRSC tissue culture cells and identified transcriptome-wide changes by RNA sequencing (RNA-seq). We identified thousands of targets for these proteins, providing the most extensive genome-wide characterization of targets of splicing regulators in any organism to date.

Results

Identification of transcriptome-wide regulatory targets of 56 proteins

To obtain a global understanding of splicing regulation in *D. melanogaster* we performed RNA-seq of poly(A)+ RNA isolated from S2-DRSC cells that were individually depleted of 56 candidate splicing regulatory proteins. We selected the proteins based on their previously known roles in regulating alternative splicing, homology to known regulators (Mount and Salz 2000; Park et al. 2004a; Barbosa-Morais et al. 2006), or their potential to function as a splicing regulator due to the presence of one or more RRM or KH domains. Each protein was categorized as an SR protein, hnRNP protein, a core component of the spliceosome, part of the exon junction complex (EJC), having evidence of splicing regulation, or having no prior evidence of splicing regulation (Supplemental Table 1), though these categories are not mutually exclusive.

After confirming depletion of the targeted mRNA by RT-PCR (Supplemental Figure 1), we prepared poly(A)+ RNA-seq libraries from two biological replicates of each RNAi-depleted sample as well as from untreated cells, and sequenced each library to generate a total of 18-59 million uniquely aligned, single-end, 75-76 bp reads per sample (Supplemental Table 1). All reads were trimmed to 75

bp and aligned to the genome and a set of annotated and novel splice junctions (see Supplemental Methods) using Bowtie (Langmead et al. 2009). To obtain a confident set of novel junctions to use in subsequent analyses, we scored alignments to novel splice junctions using a Shannon entropy measure (Graveley et al. 2011). We used Cufflinks (Trapnell et al. 2010) to quantitate gene and transcript levels in each sample, and JuncBASE (Brooks et al. 2011) to identify and quantitate alternative splicing events.

We identified 23,079 alternative splicing events that were expressed in our dataset (Figure 1A), 2,876 of which were significantly altered by depletion of at least one splicing regulatory protein (Figure 1B). JuncBASE incorporates unannotated splice junctions when examining splicing alterations. Only 35% of splicing alterations involved solely splice junctions annotated in FlyBase 5.12 (Figure 1B). Most (55%) splicing changes involved a splice junction reported in a large modENCODE study (MDv1) of the *Drosophila* developmental transcriptome (Figure 1B) and 10% of events involved splice junctions novel to this study. By using all observed splice junctions in our analysis, we were able to gain a more global picture of splicing changes, which would have been missed if we relied on only annotated events. A splicing event was considered to be altered if the difference in the "percent spliced in" (PSI or Ψ) value, the fraction of a gene's mRNAs that contain the exon, in the depleted sample was >10% of a virtual reference and the difference was significant given a false discovery rate (FDR) of 5% (see Methods). We performed RT-PCR validation experiments for 39 events identified as being significantly regulated in the RNA-seq data and found a significant correlation (*P*-value < 0.001) between ΔΨ values determined by JuncBASE and by RT-PCR (Supplemental Figure 2A). Additionally, our set of exons significantly affected by knockdown of *B52*, *Rbp1*, *Rbp1-like*, *Rsf1*, *SC35*, *Srp54*, and *x16* had concordant ΔΨ values with an independent study of these same proteins (Bradley et al. 2015, Supplemental Figure 2B).

Between 65 and 780 events were affected by depletion of each protein, though each had distinct effects on the number, magnitude, and type of affected splicing events (Figure 1C). Depletion of core spliceosomal components such as *Rm62*, *snRNP-U1-70K*, and *pea (Prp22)* impacted the largest number of splicing events (over 500 each). Other proteins that had a large impact on splicing include the SR protein B52 and the RNA binding protein CAPER, which both affected over 400 splicing events. The 12 proteins we examined that had no previously known role in alternative splicing affected many splicing events. In fact, depletion of *eIF3ga* and *shep* each affected over 300 events, more than any hnRNP proteins and most SR proteins. EIF3GA is orthologous to human EIF3G, a component of the eIF3 complex involved in translation (Lasko 2000; Hinnebusch 2006). Many splicing regulators have been shown to also affect translation (e.g., Maslon et al. 2014; Romanelli et al. 2013); therefore, other examples of multi-functional RNA-binding proteins exist. SHEP is an RNA binding protein that contains two RRM domains, is highly expressed in the nervous system of the fly (Brown et al. 2014), and is involved in gravity sensing (Armstrong et al. 2006). Our results strongly suggest that EIF3GA and SHEP are splicing regulatory proteins; however, additional biochemical studies will be important to further support this function.

Each protein also had distinct effects on the magnitude of individual splicing events. Overall, while the magnitude of some (3%) splicing changes was strong ($\Delta\Psi$ >50%), most were moderate ($\Delta\Psi$ 25-50%) or weak ($\Delta\Psi$ 10-25%). The largely weak effects on splicing events upon depletion of the core components of the spliceosome (*Rm62*, *snRNP-U1-70K*, and *pea*) suggest that cells may be fairly insensitive to changes in the levels of spliceosomal components. We confirmed depletion of the targeted mRNAs from the RNA-seq data (Supplemental Figure 3) and have shown protein depletion for genes with available antibodies in previous studies using the same dsRNA protocols (Park et al. 2004; Olsen et al., 2007). Therefore, we believe the weak splicing effects are not due to inefficient

protein depletion. However, as deletions of most genes encoding spliceosomal components are lethal in the fly, it is likely that more complete depletion of these proteins would show larger effects than we observe.

Fifty-two of the proteins examined had a significant preference in the type of alternative splicing event that was affected (chi-square test, corrected *P*-value <0.05). For example, knockdowns of SNRNP-U1-70K, CG1646 (PRP39), and ROX8 preferentially affected alternative 5' splice sites, consistent with their roles as integral U1 snRNP proteins (snRNP-U1-70K and CG1646 (PRP39)) or U1 snRNP interacting proteins (ROX8) (Figure 1C). Although the branchpoint recognition protein SF1 does affect alternative 3' splice site events, cassette exons were most affected by SF1 depletion, consistent with a recent study of human SF1 (Corioni et al. 2011). The observed preferential effects suggest possible mechanisms for proteins without well defined regulatory mechanisms. For example, *glorund* (*glo*) encodes an hnRNP protein known to regulate translation (Kalifa et al. 2006) and preferentially affects alternative 5' splice sites and cassette exons, suggesting a role in the regulation 5' splice site recognition or activation.

Alternative first exon usage cannot be explained by secondary effects through transcription factors

Though observed in previous studies of splicing regulators (Brooks et al. 2011; Huelga et al. 2012), we were surprised that a considerable proportion (6-30%) of alternative splicing events affected by all proteins are alternative first exons (AFE), as this may involve a change in promoter use (Figure 1C). To increase our confidence that these were true first exons, we only considered annotated exons (r5.32 (McQuilton et al. 2012)) that are also supported by the modENCODE CAGE datasets (Brown et al. 2014). To determine if these observations could be explained by splicing or

gene expression changes of transcription factors, we used the modENCODE ChIP-seq data for 41 transcription factors (Negre et al. 2011; Boyle et al. 2013) to predict their downstream target genes based on evidence of strong binding around gene promoters. Although we observed 266 cases where a splicing or gene expression change in a transcription factor co-occurred with an AFE change in their downstream targets, none of the AFE changes occurred consistently with effects on the transcription factor across the samples. Moreover, it is unclear if or how the splicing or gene expression changes observed in the transcription factors impact protein levels or isoforms, nor whether the interactions observed by ChIP-seq near the promoters of the affected AFEs are functional. Thus, we cannot attribute the observed AFE changes to secondary effects caused by changes in splicing or gene expression of transcription factors. Intriguingly, a recent study (Ji et al. 2013) revealed that SRSF2 associates with the promoter and enhances transcription elongation by recruitment of CDK9 from the 7SK RNA complex. Thus, there are precedents for RNA binding proteins impacting promoter choice and our results suggest that this may be more widespread than previously recognized. However, we currently do not understand the exact mechanism for our observed changes in widespread alternative first exon changes.

Gene expression changes are distinct from splicing changes

As many splicing factors not only regulate splicing, but also control other RNA processing steps including transcription, export, and decay (Braunschweig et al. 2013), we examined the impact of depleting each protein on mRNA levels. We identified between 197 and 861 genes in each sample (4194 genes in total) with at least a two-fold change in expression in comparison to the untreated sample (Supplemental Figure 4, q-value < 0.05). However, 1,670 of the genes that were affected

were expressed at low abundance (FPKM <1) in the untreated sample indicating that ~40% of these changes cannot be reliably measured.

Proteins affecting the most splicing events also caused the most changes in gene expression (Supplemental Figure 5), for example PEA and RM62. There are, however, exceptions to this trend. The greatest exception is the small ribosomal subunit protein RPS3 which had the second largest affect on mRNA levels (802 genes), but the 23rd largest affect on splicing (188 events). Interestingly, this protein also has been implicated in DNA damage repair via deoxyribophosphodiesterase activity (Wilson et al., 1993). The changes in gene expression may involve a pathway response, instead of a direct effect of RPS3 on RNA.

Although the relative number of splicing and gene expression changes observed was similar for most proteins, we found no significant overlap of genes with gene expression changes and splicing changes. Thus, the majority of gene expression changes we observed cannot be immediately explained by secondary effects of splicing changes in that gene. Further supporting multi-functional effects, 61% of intronless genes were differentially expressed upon knockdown of at least one RBP, which is more than expected given that 53% of intron-containing genes were differentially expressed (Fisher's exact test, *P*-value < 0.0001). However, we expect that the full set of splicing regulatory targets would be partially masked by NMD (Kawashima et al. 2014) and some of the differential gene expression may be secondary effects due to changes in transcription or RNA stability. Although additional studies are needed to understand the mechanism of observed gene expression changes, it is clear that looking at both splicing and gene expression changes are important to gain a global picture of transcriptome targets of each RBP.

Cross-regulation of splicing in transcripts of other regulator proteins

Several groups have previously reported that splicing regulators affect the splicing of pre-mRNAs encoding other splicing regulators (Kumar and Lopez 2005; Anko et al. 2012; Huelga et al. 2012). In some of these known cases, the cross-regulation upregulates expression of an NMD-targeted isoform, thus causing changes in gene expression of the target splicing regulator (Anko et al. 2012; Huelga et al., 2012). In *Drosophila*, a classic example of splicing regulators affecting the function of other splicing regulators is in the sex determination pathway where SXL regulates splicing of another splicing regulator, TRA, to create a functional, female-specific isoform (Salz and Erickson 2010). Although we did not observe extensive cross-regulation at the gene expression level (Supplemental Figure 6), our data does indicate cross-regulation of splicing among the 56 RBPs studied. Specifically, depletion of 31 of the RBPs affected the splicing of at least one of the 56 RBP pre-mRNAs and splicing of 26 of the RBP encoding pre-mRNAs were affected by depletion of at least one of the 56 RBPs (Figure 2). The greatest extent of cross-regulation was observed upon depletion of PSI and CAPER which each affected the splicing of 6 of the 56 RBP pre-mRNAs. On the other extreme, splicing of the *Syp* and *Rm62* pre-mRNAs were each affected by the depletion of 11 different RBPs. The cross-regulation of CAPER on splicing of other RBP genes and the splicing regulation of *Syp* by other RBPs was confirmed by RT-PCR (Supplemental Figure 2A).

Most SR and hnRNPs act coordinately rather than antagonistically

Many previous studies have shown that SR proteins tend to activate splicing while hnRNP proteins tend to repress splicing and, therefore, we examined the activity biases among the 56 proteins (Figure 3). We infer that when an exon is excluded upon knockdown of a protein, that protein functions to activate splicing of the exon, and conversely repression is inferred when increased inclusion of an exon is observed upon knockdown of a protein. Although SR proteins preferentially

11

promote exon inclusion—B52, RBP1, and SRP54 strongly promoted exon inclusion (binomial test, Bonferroni-corrected *P*-value < 0.05)—hnRNP proteins had more varied effects. For example, SQD (HRP40) and HRB27C (HRP48) had a significant preference to promote exon inclusion while CG30122 and HNRNP-K promote skipping (binomial test, *P*-value < 0.05). SF1 and SNRNP-U1-70K, core components of the spliceosome, strongly promoted cassette exon inclusion (binomial test, Bonferroni-corrected *P*-value < 0.05). In general, most proteins had a tendency to activate splicing, 10 proteins (mostly hnRNP proteins) repressed exons, and 6 showed no bias.

We next examined the extent to which individual splicing events are regulated by multiple proteins, and when multiple proteins affect a single splicing event, whether they act antagonistically or coordinately. The majority of splicing events (56%) are significantly affected by more than one protein (Supplemental Figure 7). One pair of proteins that had a large number of overlapping effects is SNRNP-U1-70K and PSI, which physically interact (Labourier et al. 2001) and work coordinately to regulate P-element splicing (Siebel et al. 1992). We observed that in nearly all cases, SNRNP-U1-70K and PSI work coordinately to regulate splicing of the target mRNA in the same direction. Among all pairs of proteins, we identified 81 pairs that had a statistically significant overlap in the cassette exon events that they affected (Fisher's exact test, Bonferroni-corrected *P*-value < 0.05; Figure 4, black squares). One example of a significant overlap in the exons affected by two proteins is HRB98DE (HRP38) and GLO, which also physically interact (Guruharsha et al. 2011). Another example involves the previously uncharacterized splicing regulators, EIF3GA and SHEP, which had a significant overlap in their regulatory targets. Additionally, both RNA-binding proteins had overlapping targets with multiple proteins in the EJC complex (particularly TSU), which have previously been shown to regulate splicing (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010; Michelle et al. 2012). These genetic interactions suggest possible physical interactions between EIF3GA, SHEP,

and the EJC complex; however, biochemical studies are necessary to confirm this. Coordinate regulation of splicing is known to be context specific (Ke and Chasin 2011); therefore, we may be missing additional co-regulated splicing events that occur in a different cell type.

Our observation of cross-regulation complicates our analysis of co-regulated targets. It is difficult to assess whether observed splicing changes are secondary effects of cross-regulation where knockdown of an RNA-binding protein affects splicing or expression of another RNA binding protein. To address this, we identified cases where RNAi depletion of one protein's gene (gene A) significantly down-regulated the gene expression or affected splicing of another RBP gene (gene B) (Figure 2, Supplementary Figure 7). We then asked if there was a significant overlap in the observed splicing changes upon knockdown of gene A and gene B (Figure 4, black squares). There were no cases where a significant overlap of splicing changes between knockdown of gene A and gene B could be explained by the down-regulation of gene expression or change in splicing of B. Therefore we do not believe that co-regulated splicing events are due to secondary effects from changes in the other RNA-binding proteins assayed; however, we cannot rule out secondary effects from other splicing regulators not targeted in this study.

In contrast to previous reports (Mayeda and Krainer 1992; Caceres et al. 1994), we found that in situations where a cassette exon was affected by both an hnRNP and an SR protein, the proteins tended to act coordinately to promote exon inclusion (Figure 4). SR proteins tended to act coordinately with each other to activate exons (Figure 4, top left corner), while hnRNP proteins also tended to act coordinately, but in many cases, either to activate or repress exons. For example, HRB98DE (HRP38) and GLO shared 16 events and they activated and repressed a similar number of exons, but did so in the same direction in 14 of those cases (Figure 4). There were only a few antagonistic SR and hnRNP protein pairs where the SR protein tended to activate and the hnRNP

protein tended to repress, such as the strong exon activators SRP54, RBP1, and RBP1-LIKE with the strong exon repressors CG30122 and HNRNP-K (Figure 3, Figure 4). Therefore, our results do not support a general model of antagonistic regulation between SR and hnRNP proteins.

DISCUSSION

Together, this work provides a valuable resource of splicing regulatory networks in *Drosophila*. We identified approximately 3,000 individual splicing events affected by knockdown of one or more of these 56 proteins. For 12 of the RNA binding protein genes (*Srp54*, *CG6227*, *Rm62*, *mub*, *qkr54B*, *Upf1*, *B52*, *Rbp1*, *elav*, *snRNP-U1-70K*, *Syp*, and *SC35*), a separate study found a significant overlap between the regulated splicing events reported here and binding targets of the same protein using RIP-seq (Stoiber et al. 2015). This suggests that the splicing events regulated by these 12 proteins are more often due to primary effects. Stoiber et al. also examined two other RNA binding proteins, TRA2 and FRM1, but they did not have a significant overlap of RIP-seq targets and regulated splicing events suggesting indirect effects or additional non-splicing related functions. For the remaining 42 proteins, additional studies are necessary to test whether each protein directly binds to the pre-mRNAs that are regulated to help establish primary versus secondary effects—particularly in light of the cross-regulation observed amongst RBPs and affects on AFE events.

This work, as well as others (e.g., Bradley et al. 2015; Huelga et al. 2012), further demonstrates that SR and hnRNPs both promote exon inclusion and exclusion; although in some cases we see a significant bias toward one mode of regulation. Moreover, our study does not support a general model of antagonistic regulation between SR and hnRNP proteins. We also identified specific splicing events affected by depletion of core components of the spliceosome and components of the EJC, further supporting their function in alternative pre-mRNA processing.

14

We would not expect the target genes of the *D. melanogaster* splicing regulators to be broadly conserved in distant eukaryotes; however, there is evidence that regulatory specificity of *Drosophila* splicing regulatory proteins extend to orthologs in other eukaryotes, such as human and mouse (Brooks et al. 2011; Irimia et al. 2011). We found multiple genes in our study, such as SHEP and EIF3GA, that have not been previously characterized as regulators of splicing. It is possible that the splicing regulatory functions of these genes are conserved across metazoans.

METHODS

RNAi depletion and RNA-seq

S2-DRSC cells were treated twice with 20 µg of dsRNAs individually targeting each RNA binding protein gene in biological replicates. After 5 days, total RNA was harvested and used to prepare poly(A)+ mRNA-seq libraries which were sequenced on an Illumina GAIIx to generate single end reads of 75-76 bp. Reads were simultaneously aligned to the genome and splice junctions using Bowtie (Langmead et al. 2009). Additional details are described in Supplemental Methods.

Splicing analysis

Identifying alternative splicing events

Although JuncBASE (Brooks et al. 2011) is designed to identify differential splicing events, it can also be used to identify all possible alternative splicing events from genes that were observed to be expressed. As coverage across a gene can vary in RNA-seq data (Hansen et al. 2010; Li et al. 2010; Kakaradov et al. 2012), an expression cutoff was made on individual splicing events (20 reads; total of reads supporting the inclusion isoform and exclusion isoforms).

Although a splice junction can be observed as constitutive in the observed RNA-seq data, it may have the potential to be alternatively spliced in other contexts; therefore, alternative splicing events were identified by including potentially unexpressed splice junctions that were annotated in MB8 or MDv1 (Graveley et al. 2011) that provide evidence that expressed constitutive exon-exon junctions can be alternatively spliced. For example, if an expressed junction differs in the corresponding 3' splice site of an annotated junction, this gives evidence that the expressed junction has the potential for alternative splicing. The coordinates of expressed splice junctions and additional annotated junctions giving evidence of alternative splicing were formatted into BED files, which are created in a preprocessing step of JuncBASE. The BED files were then directly used as input to JuncBASE to identify and classify all possible splicing events.

Briefly, JuncBASE uses splice junction alignments and exon coordinates to identify the following alternative splicing events: cassette/skipped exons (SE), alternative 5' splice sites (A5SS), alternative 3' splice sites (A3SS), mutually exclusive exons (MXE), coordinate cassette/skipped exons (CSE), alternative first exons (AFE), alternative last exons (ALE), and retained introns (RI) (Supplemental Figure 8).

Exon coordinate annotations from the MDv1 annotation (Graveley et al. 2011) and novel exons identified through Cufflinks v0.9.3 (Trapnell et al. 2010) were used as input to JuncBASE. Each alignment file from the RNAi samples and the untreated sample was run through Cufflinks to identify *de novo* transcript annotations using the option -I 200000.

16

JuncBASE restricts the classification of alternative 5' and 3' splice sites to events near annotated exons, however this restriction was not enforced in cases where an alternative junction was identified in MDv1.

Alternative first exons were identified by rerunning JuncBASE using only first exon annotations in FlyBase r5.32 with CAGE tag evidence. Alternative first exon events called from this run were combined with all other AS types from the JuncBASE run using the MDv1 annotation and *de novo* transcripts, described above.

Alternative splicing quantification

After the identification of each alternative splicing event, JuncBASE counts reads supporting the inclusion and exclusion isoform of each event. Isoform abundances are then calculated by dividing the read counts for the isoform by the length of the isoform. "Percent spliced in" (Ψ) values for each splicing event are derived from the isoform abundances:

$$\Psi = \frac{inclusion\ isoform\ abundance}{inclusion\ isoform\ abundance + exclusion\ isoform\ abundance}$$

Virtual reference calculation

To control for any splicing changes that may have been caused as an artifact of performing RNAi, a virtual reference sample was created to serve as a control. The assumption behind the creation of a virtual reference is that most splicing events will only be affected by a subset of proteins; therefore, comparing splicing in an RNAi sample with the median splicing across all events will identify splicing events specifically affected by each protein. The virtual reference exclusion and inclusion counts were calculated from the median expression and median Ψ values from all RNAi samples:

virtual reference inclusion count = median total count of event * median inclusion ratio

virtual reference exclusion count = median total count of event * median exclusion ratio

Differential splicing calculation

A Fisher's exact test was performed for each splicing event comparing the exclusion and inclusion read counts from each RNAi sample to the virtual reference inclusion and exclusion counts to identify samples with a significant difference in splicing. For each RNAi sample, a Benjamini-Hochberg multiple testing correction for each event type (e.g., cassette exon, alternative 5' splice sites, etc.) was performed to give a final set of affected splicing events with a false discovery rate (FDR) < 0.05.

Alternative splicing events with more than two isoforms were tested for significant differences in splicing by treating each of the isoforms as the inclusion isoform and all other combined isoforms as the exclusion isoforms. In the final reporting of significantly affected events, the inclusion isoform giving the lowest corrected *P*-value is reported.

Gene Expression

Cufflinks and CuffDiff (Trapnell et al. 2010) were used to quantitate gene expression levels using the MDv1 annotation (Graveley et al. 2011). To quantitate expression levels of the 56 targeted RNA binding proteins, the annotation was masked of dsRNA regions used for the RNAi depletion to disambiguate reads that may have originated from either dsRNA or the target gene.

Identifying target genes of transcription factors

To identify target genes of transcription factors, we used the peak-finding results from ChIP-seq datasets for 41 transcription-factor (TF) (Negre et al. 2011; Boyle et al. 2013). We identified transcription factors that had ChIP-seq peaks 2000 bp upstream to 100 bp downstream from the transcription start (according to the MDv1 annotation), and produced a TFBS score for each TF and transcript and gene based on peak coverage. From the distribution of TFBS scores, we selected a cutoff of high TFBS scores to define TF target sites.

Data access

The sequence data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSM461183-GSM627418.

Author Contributions

A.N.B., M.O.D., L.Y., B.R.G, and S.E.B designed the work. L.Y. and G.M. performed RNAi, prepared libraries, and performed sequencing. G.M. and M.B. performed validations. J.L., K.W., J.S., and S.E.C. provided sequencing support and quality control. A.N.B., M.O.D., B.W.B., and B.R.G.

performed computational analyses with input from S.E.B; A.N.B., B.R.G., and S.E.B. wrote the paper with contributions from co-authors.

Acknowledgments

References

Anko ML, Muller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. 2012. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13**(3): R17.

Armstrong JD, Texada MJ, Munjaal R, Baker DA, Beckingham KM. 2006. Gravitaxis in Drosophila melanogaster: a forward genetic screen. *Genes, brain, and behavior* **5**(3): 222-239.

Ashton-Beaucage D, Udell CM, Lavoie H, Baril C, Lefrancois M, Chagnon P, Gendron P, Caron-Lizotte O, Bonneil E, Thibault P et al. 2010. The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. *Cell* **143**(2): 251-262.

Barbosa-Morais NL, Carmo-Fonseca M, Aparicio S. 2006. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res* **16**(1): 66-77.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**(6114): 1587-1593.

Blanchette M, Green RE, Brenner SE, Rio DC. 2005. Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes Dev* **19**(11): 1306-1314.

Blanchette M, Green RE, MacArthur S, Brooks AN, Brenner SE, Eisen MB, Rio DC. 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Mol Cell* **33**(4): 438-449.

Boutz PL, Stoilov P, Li Q, Lin C, Chawla G, Ostrow K, Shiue L, Ares M, Black DL. 2007. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes and Development* **21**(13): 1636.

Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier L, Janette J, Jiang L et al. 2014. Comparative analysis of regulatory information and circuits across diverse species. *Nature* **512**(7515):453-456

Bradley T, Cook ME, Blanchette M. 2015. SR proteins control a complex network of RNA-processing events. *RNA* **21**(1):75-92

Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. 2013. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**(6): 1252-1269.

Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR. 2011. Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Res* **21**(2): 193-202.

Brown JB, Boley N, Eisman R, May GE, Stoiber M, Booth BW, Park S, Robinson G, Suzuki AM, Wan KH et al. 2014. Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**: 393-399.

Caceres JF, Stamm S, Helfman DM, Krainer AR. 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**(5179): 1706-1709.

Chang YF, Imam JS, Wilkinson MF. 2007. The nonsense-mediated decay RNA surveillance pathway. *Annual review of biochemistry* **76**: 51-74.

Chen M, Manley J. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology* **10**(11): 741-754.

Corioni M, Antih N, Tanackovic G, Zavolan M, Kramer A. 2011. Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res* **39**(5): 1868-1879.

Gabut M, Chaudhry S, Blencowe BJ. 2008. SnapShot: The splicing regulatory machinery. *Cell* **133**(1): 192 e191.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**: 473-479.

Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O et al. 2011. A protein complex network of Drosophila melanogaster. *Cell* **147**(3): 690-703.

Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* **38**(12): e131.

Hinnebusch AG. 2006. eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem Sci* **31**(10): 553-562.

Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**(2): 167-178.

Irimia M, Denuc A, Burguera D, Somorjai I, Martin-Duran JM, Genikhovich G, Jimenez-Delgado S, Technau U, Roy SW, Marfany G et al. 2011. Stepwise assembly of the Nova-regulated alternative splicing network in the vertebrate brain. *Proc Natl Acad Sci U S A* **108**(13): 5319-5324.

Isken O, Maquat LE. 2008. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet* **9**(9): 699-712.

Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R, Burge CB, Fu XD. 2013. SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. *Cell* **153**(4): 855-868.

Kakaradov B, Xiong HY, Lee LJ, Jojic N, Frey BJ. 2012. Challenges in estimating percent inclusion of alternatively spliced junctions from RNA-seq data. *BMC Bioinformatics* **13 Suppl 6**: S11.

Kawashima T, Douglass S, Gabunilas J, Pellegrini M, Chanfreau GF. 2014. Widespread use of non-productive alternative splice sites in Saccharomyces cervisiae. *PLoS Genet.* **10**(4):e1004249

Kalifa Y, Huang T, Rosen LN, Chatterjee S, Gavis ER. 2006. Glorund, a Drosophila hnRNP F/H homolog, is an ovarian repressor of nanos translation. *Dev Cell* **10**(3): 291-301.

Ke S, Chasin LA. 2011. Context-dependent splicing regulaiton: exon definition, co-occuring motif pairs and tissue specificity. *RNA Biol.* **8**(3):384-8

Kumar S, Lopez AJ. 2005. Negative feedback regulation among SR splicing factors encoded by Rbp1 and Rbp1-like in Drosophila. *EMBO J* **24**(14): 2646-2655.

Labourier E, Adams MD, Rio DC. 2001. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol Cell* **8**(2): 363-373.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.

Lasko P. 2000. The drosophila melanogaster genome: translation factors and RNA binding proteins. *J Cell Biol* **150**(2): F51-56.

Li J, Jiang H, Wong WH. 2010. Modeling non-uniformity in short-read rates in RNA-seq data. *Genome Biol* **11**(5): R50.

Maslon MM, Heras SR, Bellora N, Eyras E, Caceres JF. 2014. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *eLIFE* May 6:e02028. doi: 10.7554/eLife.02028

Mayeda A, Krainer AR. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**(2): 365-375.

McQuilton P, St Pierre SE, Thurmond J, FlyBase C. 2012. FlyBase 101--the basics of navigating FlyBase. *Nucleic Acids Res* **40**(Database issue): D706-714.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**(6114): 1593-1599.

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**(5): 954-967.

Mount SM, Salz HK. 2000. Pre-messenger RNA processing factors in the Drosophila genome. *J Cell Biol* **150**(2): F37-44.

Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R et al. 2011. A cis-regulatory map of the Drosophila genome. *Nature* **471**(7339): 527-531.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**(7280): 457-463.

Olsen S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley JM, Rio DC, Graveley BR. 2007. A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol* **14**(12):1134-40

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**(12): 1413-1415.

Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR. 2004a. Identification of alternative splicing regulators by RNA interference in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **101**(45): 15974-15979.

Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR. 2004b. Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc Natl Acad Sci USA* **101**(45): 15974-15979.

Perez-Canadillas JM, Varani G. 2001. Recent advances in RNA-protein recognition. *Current opinion in structural biology* **11**(1): 53-58.

Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. 2007. Transcript Specificity in Yeast Pre-mRNA Splicing Revealed by Mutations in Core Spliceosomal Components. *PLoS Biol* **5**(4): e90.

Roignant JY, Treisman JE. 2010. Exon junction complex subunits are required to splice Drosophila MAP kinase, a large heterochromatic gene. *Cell* **143**(2): 238-250.

Romanelli MG, Diani E, Lievens PM. 2013. New insights into functional roles of the polypyrimidine tract-binding protein. *Int J Mol Sci* **14**(11): 22906-32

Salz HK, Erickson JW. 2010. Sex determination in Drosophila: The view from the top. *Fly* **4**(1):60-70

Siebel CW, Fresco LD, Rio DC. 1992. The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes Dev* **6**(8): 1386-1401.

Stoiber MH, Olson S, May GE, Duff MO, Manent J, Obar R, Guruharsha KG, Artavanis-Tsakonas S, Brown JB, Graveley BR, Celniker SE. 2015. Extensive cross-regulation of post-transcriptional regulatory networks in *Drosophila. Genome Research accepted*

Tange TO, Nott A, Moore MJ. 2004. The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* **16**(3): 279-284.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech* **28**(5): 511-515.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**(7119): 580-586.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**(7221): 470-476.

Wilson DM, Deutsch WA, Kelley MR. 1993. Cloning of the Drosophila ribosomal protein S3: another multifunctional ribosomal protein with AP endonuclease DNA repair activity. *Nucleic Acids Res.* **21**(10):2516

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**(2): 130-137.

Figure Legends

Figure 1. Alternative splicing (AS) events affected by depletion of 56 RNA binding proteins. a. The proportion of each type of AS event that could potentially be regulated by each protein. **b**. The observed proportion of each type of splicing event affected by depletion of at least one protein is shown alongside the type of exon-exon junctions involved in the splicing event. **c**. The number, magnitude, type of AS, and type of exon-exon junction affected by each individual knockdown is shown. Each protein is categorized as SR, hnRNP, core splicing factor, exon junction complex (EJC), other known, or novel splicing regulator protein. The legend in the top right indicates the color codes used for the splicing event types (SE, skipped/cassette exon; A5SS, alternative 5' splice site; A3SS, alternative 3' splice site; MXE, mutually exclusive exons; CSE, coordinate skipped/cassette exon; AFE, alternative first exon; ALE, alternative last exon; RI, retained intron), the transcript annotation exon-exon junctions were found in (FB5.12, FlyBase 5.12; MDv1 (Graveley et al. 2011)), or the magnitude of splicing changes.

Figure 2: Cross-regulatory splicing networks involving the 56 RNA binding proteins. Based on validation experiments and manual inspection of RNA-seq data, we report observed cross-regulatory effects for moderate to strong splicing changes reported by JuncBASE. The cross-regulation of one protein affecting a splicing event of another protein's pre-mRNA is indicated in the matrix. Although splicing events are diagrammed as cassette exons, effects on all forms of alternative splicing are reported.

Figure 3: Bias of activation and repression of cassette exons. The number of cassette exons that are activated and repressed by each protein is shown. The proteins are ordered by increasing

proportion of activated exons. The significance of the bias to activate or repress exons is indicated with asterisks (two-sided binomial test, *P-value <0.05, **Bonferonni-corrected P-value < 0.05).

Figure 4. Antagonistic and coordinated effects of splicing regulators. Comparisons of cassette exons co-affected by each pair of proteins is represented as a matrix. Within each group (e.g., SR, hnRNP, Core, etc.) the proteins are ordered by the number of splicing events they affect as shown in Figure 1. Information in the upper and lower portions of the matrix is identical. The size of each circle gives the relative number of co-affected events and the color indicates if the pair acted antagonistically (one protein activates, the other represses) or coordinately (both proteins activate or both proteins repress). Overlaid spokes indicate what proportion of shared events are activated or repressed by each protein in the pair, thus giving more detailed information on the nature of the antagonistic or coordinated effects. Pairs of proteins that have a significant overlap in the number of co-affected events are indicated with black squares (Fisher's exact test, Bonferroni-corrected P-value < 0.05).
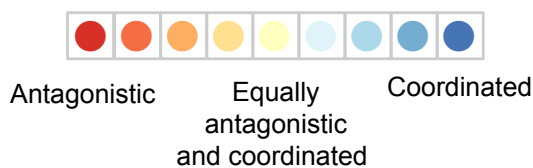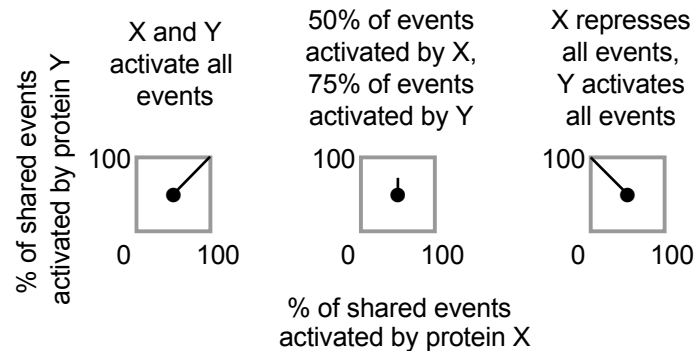
Figure 1

Figure 2

Figure 3

Figure 4

Regulation of alternative splicing in *Drosophila* by 56 RNA binding proteins

Angela N Brooks, Michael O Duff, Gemma May, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2015/08/26/gr.192518.115.DC1.html
P<P	Published online August 20, 2015 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the *Genome Research* Open Access option.
Creative Commons License	This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions