Supplemental Materials

for

Conservation of an RNA Regulatory Map between *Drosophila* and Mammals

Angela N. Brooks, Li Yang, Michael O. Duff, Kasper Daniel Hansen, Jung W. Park, Sandrine Dudoit, Steven E. Brenner and Brenton R. Graveley

Supplemental Methods.

Expression and purification of recombinant PS

A full-length PS cDNA was cloned into pTrc-His2, expressed in *E. coli* and purified using Ni-NTA chromatography (QIAGEN) followed by DEAE chromatography.

Gel shift assay

RNAs containing either three UCAU or GCAU repeats were radiolabeled with ^{32}P-UTP during *in vitro* transcription. These RNAs were incubated with recombinant PS at room temperature for 20-30 minutes and the mixture was loaded onto 5% non-denaturing gel electrophoresis. The gel was dried and analyzed by phosphorimager and Optiquant software (Perkin Elmer, CA).

Identifying changes in overall gene expression

FPKM values were obtained for all genes in the Untreated and *ps(RNAi)* samples using Cufflinks (Trapnell et al. 2010). A Fisher's test was applied to identify genes with significantly different expression between untreated and *ps(RNAi)* given a Benjamini-Hochberg corrected p-value ≤ 0.05 for 7,834 tests. No genes containing a KH or RRM domain found to significantly changing; however, RNA binding proteins *eIF-4a* and *CG10630* had significant, relatively small changes in gene expression (15% increase and 30% decrease, respectively) compared to the 60% decrease in gene expression of

Pasilla. *eIF4a* had an FPKM of 2,683 in Untreated and 2,366 in PS-RNAi. *CG10630* had an FPKM of 165 in Untreated and 236 in the knockdown. It is unclear if the small changes in gene expression relative to the difference in Pasilla would have an effect on splicing.

JuncBASE analysis details

Following are descriptions of how JuncBASE analyzes alternative splicing events. All alternative splicing events and associated p-values can be found in Supplemental Dataset 1.

Cassette Exons

For every exon in our transcript annotation set, we searched for splice junctions that gave evidence for the skipping of the exon—termed exclusion junctions. An exclusion junction has splice sites flanking an intron that fully contains the exon; thus, when this intron is spliced the exon is necessarily skipped. Splice junction reads that aligned to the cassette exon were considered inclusion junctions. Reads mapping to all inclusion junctions and reads fully contained within the cassette exon were used for the inclusion counts. If there was more than one exclusion junction (in the case that either of the splice sites corresponding to the exclusion junction are also part of an alternative 5' or 3' splice site), the sum of counts for all exclusion junctions was used for the total exclusion counts.

Mutually Exclusive Exons

Groups of two or more non-overlapping exons that formed junctions with the same upstream and downstream splice site were identified. None of the exons within a group could have a junction read connecting them. If a junction supported skipping of all exons in the event, it was still classified as mutually exclusive.

If there were three or more exons in a group, all possible inclusion and exclusion isoforms were treated as separate events. For example, if Exon 3, Exon 4, and Exon 5 are mutually exclusive, one event would treat Exon 3 as the inclusion isoform and the amalgamation of Exons 4 and 5 as the exclusion isoform. Another event would treat Exon 4 as the inclusion isoform and the amalgamation of Exons 3 and 5 as the exclusion isoform. A third event would treat Exon 5 as the inclusion isoform and the amalgamation of Exons 3 and 4 as the exclusion isoform.

Junction reads and reads fully contained within the exons were used for counts.

Coordinate Cassette Exons

Groups of two or more consecutive and non-overlapping exons that were fully contained in an intron and wholly spliced out (inferred by a splice junction), and thus had evidence of skipping, were identified. From these, we identified cases where all consecutive exons within a group had a junction read connecting them, and classified these as coordinate cassette exons. The sum of all connecting splice junctions and reads completely contained within the exons were counted toward the inclusion isoform. Reads aligning to the exclusion junction were counted toward the exclusion isoform.

Alternative 5' Splice Site, Alternative 3' Splice Site, Alternative First Exon, and Alternative Last Exon

We identified instances where two or more introns (inferred from splice junctions) had the same start (or end) position. Introns that were already part of a cassette exon event, a mutually exclusive event, or a coordinate cassette exon event were removed from this list, unless the exon itself had alternative 5' or alternative 3' splice sites. Next, the strand of the event was identified by the annotated gene region that contained the intron. Introns were classified as part of an alternative 5' splice site or alternative 3' splice site event based on the strand orientation of the intron. If the adjacent exons of the alternative 5' (or 3') splice sites were non-overlapping and at least one adjacent exon was a first (or last) exon, then the event was categorized as an alternative first (or last) exon; otherwise, the event was classified as an alternative 5' (or 3') splice site.

If there were three or more introns involved in an event, each intron was treated as the inclusion isoform from a separate event (except the longest exclusion intron, which is not tested as an inclusion isoform), with all the others treated as exclusion (even if they actually included more sequence). For example, if there were three introns involved in an alternative 3' splice site choice, they will be treated as two separate alternative 3' splice site events. The 3 junctions would have the same start position, but 3 different end positions: *a,b,c*, where *c* is the splice site that would remove the longest intron. One alternative 3' splice site event would add reads supporting splice site *a* to the inclusion isoform count and counts supporting *b* and *c* to the exclusion isoform count. Another alternative 3' splice site event would add reads supporting splice site *b* to the inclusion isoform and counts supporting *a* and *c* to the exclusion isoform. For

alternative first and last exons, splice junction reads and reads fully contained within alternative first or last exons were used. For alternative 5' and 3' splice sites, splice junction reads were used for counts. In addition, reads fully contained within the alternative region of the inclusion isoform (the portion of the exon that is unique to the inclusion isoform) were used for the inclusion counts (see Figure 2, alternative 5' splice site and alternative 3' splice site). Reads that aligned to the boundary of the alternative region and the constitutive region of the event, were added to counts for the inclusion isoform and are indicated as red colored bars with a line in Figure 2.

Fisher's Exact Test to identify significantly affected alternative splicing events

Read counts from either the inclusion or exclusion isoforms in the untreated or RNAi sample were used to create a 2 x 2 contingency table (e.g., Figure 3, and Supplemental Figure 8). A Fisher's exact test was performed for each event. A total of 2,324 test were performed from the seven event classes described above. A cutoff corresponding to a Benjamini-Hochberg corrected p-value of 0.05 was applied to each of the above seven event classes. Between a choice of a Bonferroni or a Benjamini-Hochberg (BH) mutiple testing correction, the less stringent BH cutoff incorporated additional cassette exon events that appeared valid based on visual inspection of the read alignments.

The shift in the proportion of inclusion reads (inclusion reads/(exclusion + inclusion reads)) in each sample was used to determine the shift in the direction of the splicing event. For example, if there was an increase in the percent of reads that supported the inclusion isoform in the RNAi sample, then that event was considered to be normally repressed by PS.

Identifying significantly affected retained intron events

Every confident junction in our data was examined for evidence of intron retention. For quantifying the retention (inclusion) of the intron, we did not use all reads falling within the intron coordinates, because the intron could contain, for example, internal cassette exons. We found that the most informative reads to quantify the inclusion of an intron were reads that aligned to the intron-exon junction of both the 5' and 3' end of the intron. However, this could lead to some introns with both alternative 5' and 3' splice sites to be misclassified as retained introns. The splice junction reads formed from the splicing of the intron were counted towards the exclusion isoform. We found that simply summing the inclusion counts from the 5' and 3' end of the intron-exon junction could be confounded with alternative 5' and 3' splice sites. Therefore, we applied Fisher's exact test separately at the 5'-end and the 3'-end and combined the resulting two p-values as depicted in Supplemental Figure 6. All events with a Benjamini-Hochberg corrected p-value ≤ 0.05 (based on N = 1,568) were considered affected.

Identifying significantly affected junctions that are not classified in an event type.

To identify potentially unclassified alternative splicing events, read counts to each junction were compared to read counts to all other mutually exclusive junctions on the same strand. When an exon-exon junction is observed, an intron is removed and other introns that overlap the removed intron (an alternative intron selection) cannot also be removed; therefore, junctions whose implied introns are overlapping are mutually exclusive. Counts from each junction and its mutually exclusive junctions were used

from both the untreated and *ps(RNAi)* samples to perform Fisher's exact test to identify significantly changing events. Seventeen junctions that were significantly changing (Benjamini-Hochberg corrected p-value ≤ 0.05), but not part of any classified event, were identified (Supplemental Dataset 3).

Tandem 3' UTRs (alternative polyadenylation)

We attempted to identify tandem 3' UTRs by two methods, but did not find sufficient evidence for a genome-wide analysis of alternative polyadenylation. To detect novel alternative poly(A) cleavage sites, we identified reads that had at least 6 consecutive As (or Ts) at the end (or beginning) of the read, and which thus potentially represent direct sequencing the poly(A) taily. In total, 56,828 reads (before mapping) from all the *ps (RNAi)* samples contained evidence of a poly(A) tail and 58,636 reads from the Untreated samples. These were deemed insufficient for making significant conclusions, as as some of these reads may be genomically-encoded adenine stretches while others may not map uniquely; amongst those which do represent true mapppable poly-A sites, they will be distributed across all polyadenylated transcripts. In addition, tandem 3' UTR events do not contain exon or junction reads that are unique to the exclusion isoform, which makes it difficult for a reliable quantification of the two isoforms.

Obtaining a non-redundant set of alternative splicing events

Two or more splicing events were considered redundant if they contained overlapping intron coordinates. More specifically, *cassette exon, alternative 5' splice site, alternative 3' splice site, coordinate cassette exons, alternative first exon, alternative last exon, and*

retained intron events were considered redundant if their exclusion (skipping) introns overlapped. *Mutually exclusive exons* were considered redundant if any intron overlapped with another event. Events within each class of alternative splicing were considered redundant and the event with the lowest p-value was included in the non-redundant set.

Total number of alternative splicing events

The number of possible alternative splicing events was identified by using all junctions in the merged FlyBase r5.11 and MB5 annotation plus any additional confident novel junctions as input into *JuncBASE*. An alternative splicing event was considered "expressed" if at least one junction from any isoform was present in both the untreated and *ps(RNAi)* samples.

Calculation of conserved YCAY motif clusters

To ensure that we were implementing the correct YCAY conserved cluster score method, we reanalyzed the mouse exon sequences reported to be regulated by the NOVA splicing regulator (Ule et al. 2006). After personal communications with Jernej Ule and Robert Darnell, we realized that the YCAY conserved cluster score, calculated using predefined scores as described in the Supplemental Methods of Ule et al. (2006) ("predefined score method"), was slightly different from the score actually used for analyses and to generate figures in that paper ("counting method"). The software used for the analyses and figures in Ule et al. (2006), implementing the counting method, is available at http://splicing.rockefeller.edu/map/. That software counts the number of

times the following motifs occurred in a given window: YCAY(N)6YCAY, YCAYCAY, YCAY(N)2YCAY, YCAYCAY(N)6YCAY, YCAY(N)6YCAYCAY, YCAY(N)2YCAY(N)6YCAY, YCAY(N)6YCAY(N)2YCAY, YCAYCAYCAY, YCAY(N)2YCAY(N)2YCAY. Our implementation of both predefined scores and counting methods on the mouse NOVA-regulated exons from Ule et al. (2006) is included in Supplemental Figure 13. The method using predefined scores does not change the overall conclusions of the study; therefore, we decided to use the predefined scores method exactly as described in Ule et al. (2006) for our study on the Pasilla targets. *D. simulans, D. sechellia, D. yakuba,* and *D. erecta* were considered closely related species and *D. pseudobscura, D. ananassae, D. grimshawi, D. mojavensis, D. persimilis, D. virilis,* and *D. willistoni* as distant species. Thus, the conserved cluster score for a given window near PS-regulated events was:

Conserved S = 2*MIN S(*D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta*) + AVERAGE S(*D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta*) + AVERAGE S (*D. pseudobscura, D. ananassae, D. grimshawi, D. mojavensis, D. persimilis, D. virilis, D. willistoni*)

Where S in a given window is the $\log_{10}$ of the sum of occurrences of the predefined motifs described in Ule *et al.* (2006).

Calculating a net conserved YCAY conserved cluster score

In Ule et al. 2006, a net conserved YCAY cluster score, S, for a particular exon was given by:

Net conserved S = 1/2(MAX(NISE1, NISE2, NISE3, SUM(NISE2, NISE3)*2/3)-MAX (NISS1, NISS2, NESE))

Where NISE1, NISE2, and NISE3 were NOVA intronic splicing enhancer regions, NISS1, NISS2 were NOVA intronic splicing silencer regions, and NESE was a NOVA exonic splicing enhancer region.

Given the YCAY cluster scores in windows surrounding the PS-affected cassette exons, we only identified significant enrichment at positions analogous to NISE1, NISE2, NISS1, and NISS2. We therefore modified the net conserved cluster score to correspond with only positions that were significantly enriched with conserved YCAY clusters near PS-affected. Our modified cluster score is

Net conserved S = 1/2(MAX(NISE1:-180, NISE2:+40, NISE2:+60, NISE2:+80)-MAX (NISS1:+120, NESS1:+40))

The relative positions correspond to the locations of significant enrichment indicated by an asterisk in Figure 4A. Scores were calculated using only introns > 400nt.

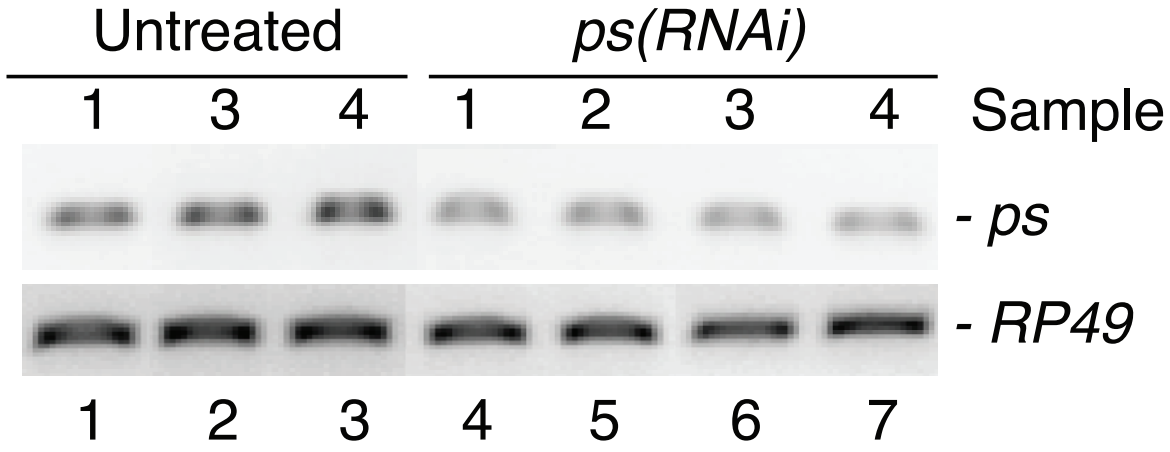
Identifying *Drosophila* orthologs of NOVA-regulated genes

A list of cassette exons regulated by NOVA, in mouse, were obtained from Supplemental Table 1 in Ule *et al.* Nature, 2006. Each gene was queried in TreeFam (Li et al. 2006; Ruan et al. 2008) to identify the orthologous *Drosophila* gene(s). Out of 47 genes regulated by NOVA, 33 had at least one *Drosophila* ortholog in the TreeFam database. To determine if these *Drosophila* orthologs were expressed in the S2-DRSC

cell line, we identified genes that had at least one confident junction as our threshold for expression.

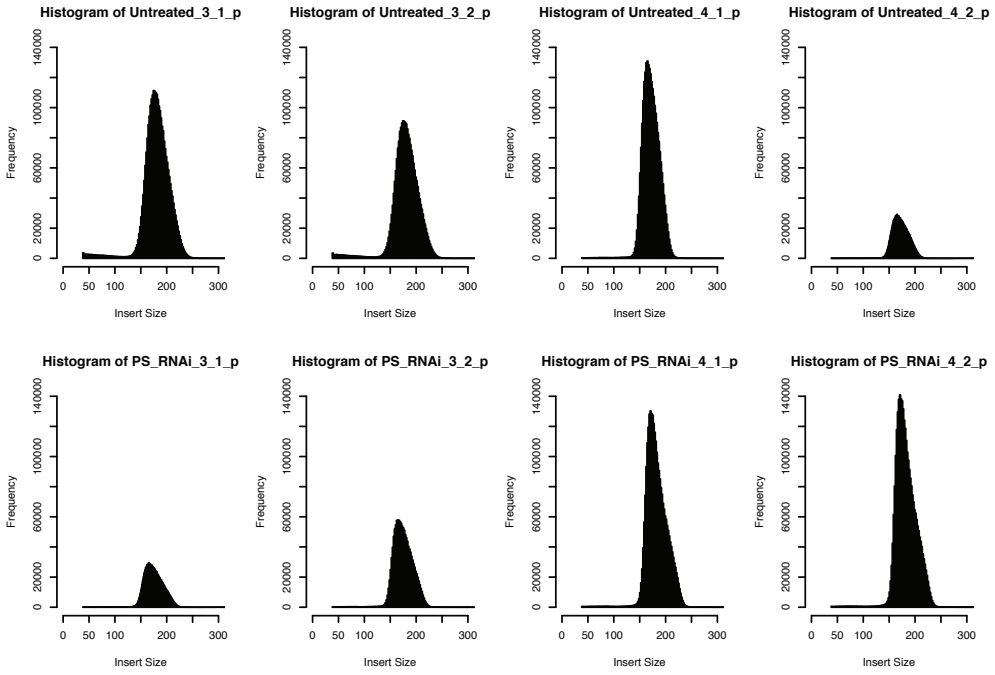
GO enriched terms

Funcassociate 2.0 (Berriz et al. 2009) was used to identify GO terms that were enriched in our set of PS target alternative splicing events. FlyBase CG names from all 323 PS-affected genes were used as our query set. A "gene space" set can be used in Funcassociate, to control for genes that might be enriched due to a bias in expression of the gene in the cell type. For this background set of genes, we used all genes that had at least one confident junction as our threshold for expression.

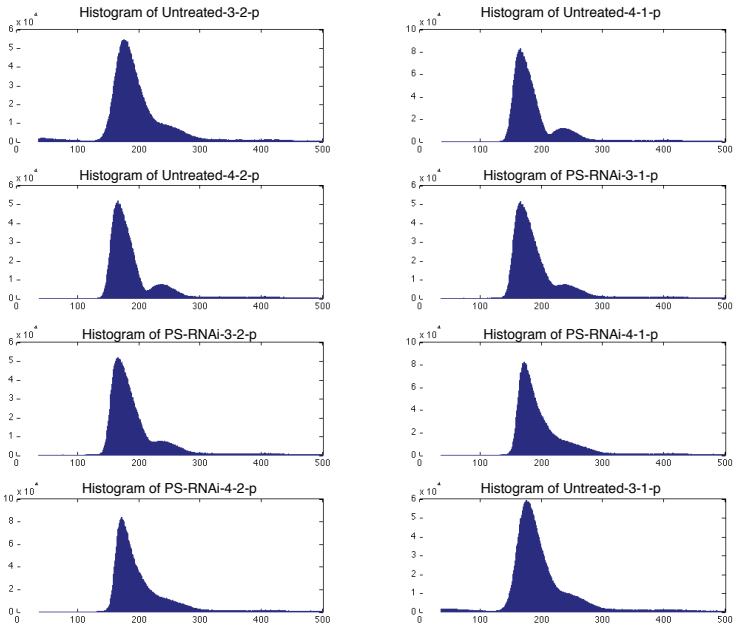


Supplemental Figure 1. Validation of RNAi depletion by RT-PCR. Semi-quantitative RT-PCR to examine the mRNA levels of *ps* in the three biological replicates of Untreated cells (lanes 1-3) and the four biological *ps(RNAi)* replicates (lane 4-7). The mRNA levels are consistently reduced in the *ps(RNAi)* cells. RT-PCR is performed using primer to the *RP49* gene as a control.

A.

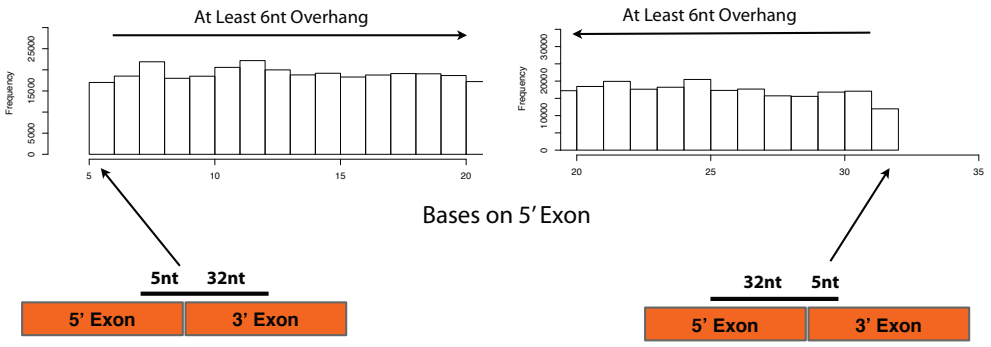


B.

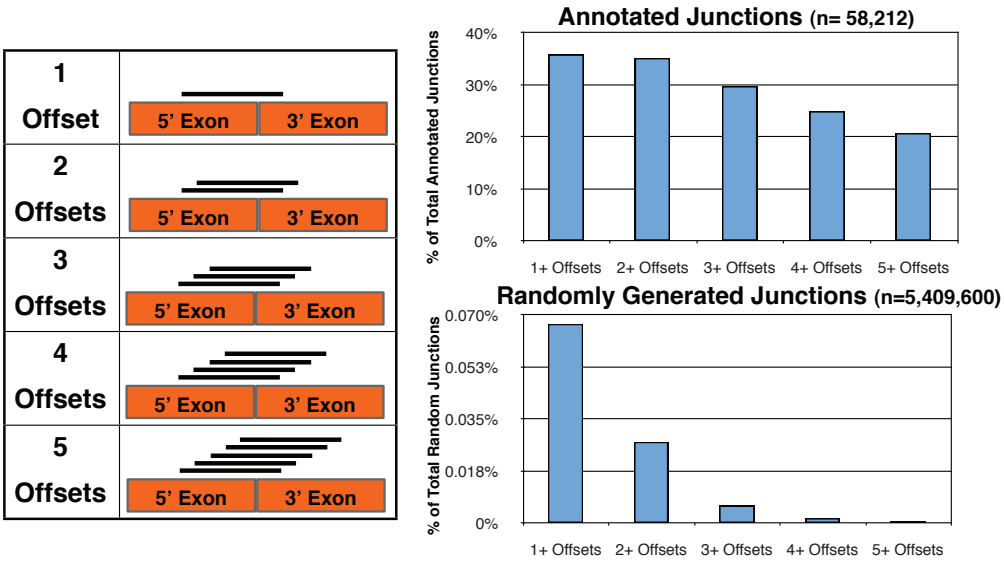


Supplemental Figure 2. Distribution of insert size for paired-end libraries. (A) Distribution of the insert size of the libraries as determined by alignment to a transcriptome (merge of FlyBase r5.11 and MB5) database. (B) Distribution of the insert size of the libraries as determined by alignment to the *D. melanogaster* genome.

A.



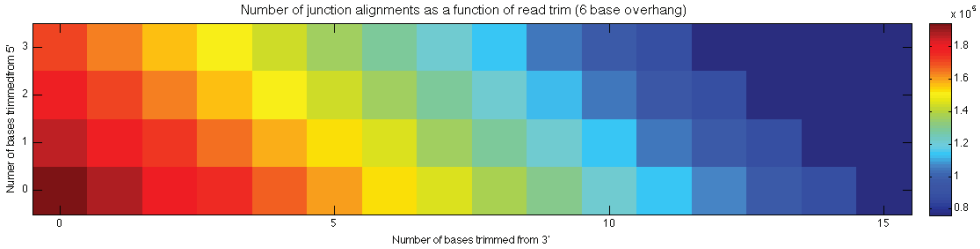
B.



Supplemental Figure 3. Analysis of optimal overhang and mismatch for splice junction alignments. (A) Distribution of Overhang Positions ≥ 5nt. A histogram of the number of uniquely aligned reads across all annotated junctions is shown. An even distribution of read alignments across all base positions occurs if at least a 6nt overhang is enforced. (B) Distinguishing true junctions from false positive alignments. To reduce the number of false positive junctions, as determined by randomly generated junctions, a total of 3 alignment start positions (offsets) were required to consider a junction to be truly present.

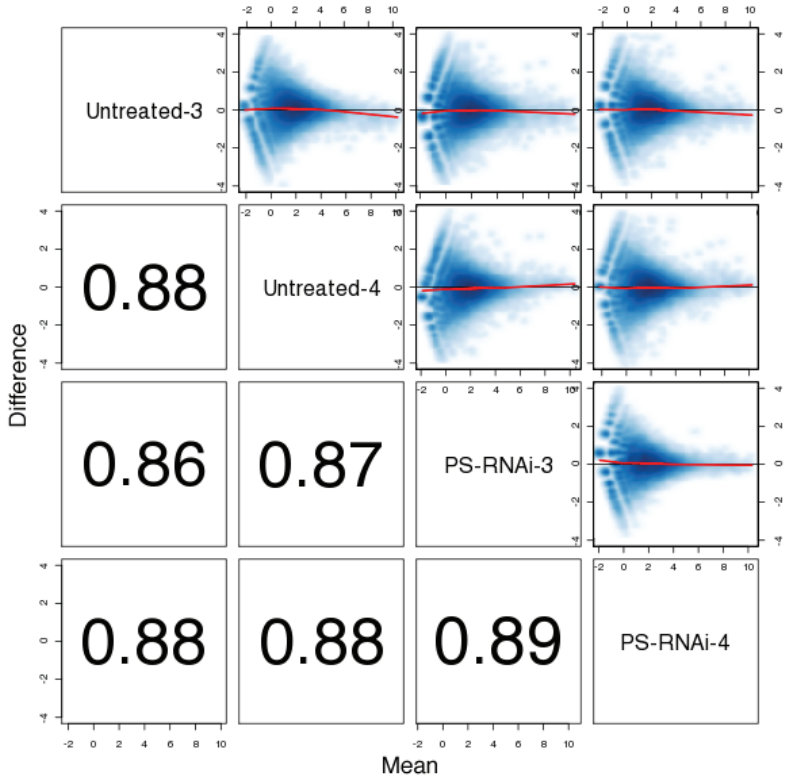
A.

B.

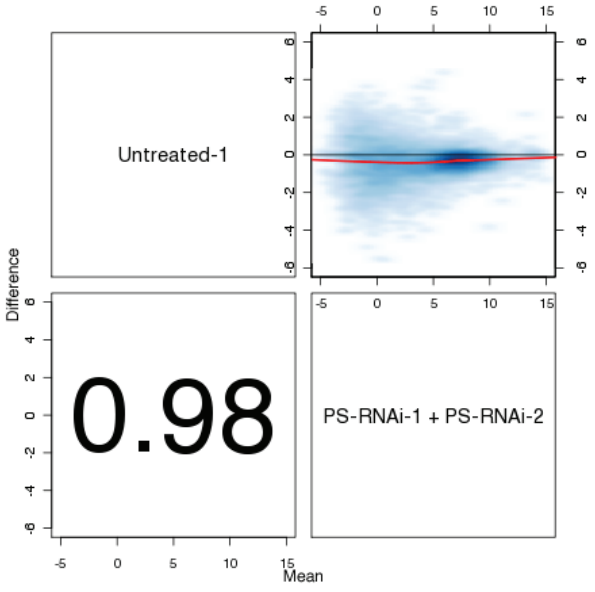


Supplemental Figure 4. Analysis of the impact of trimming on the number of reads that align to the genome and splice junctions. One lane of data was simultaneously aligned to the genome and splice junction database using a variety of parameter and the number of reads that aligned uniquely to the genome and junctions were calculated. In each round of alignment, the reads were trimmed from the 5' end by either 0, 1, 2, or 3 bases, and at the 3' end from between 0 to 15 bases in all possible pair-wise combinations. The number of reads aligning to both the genome and junction were then calculated and plotted as a heat map. (A) Reads aligned to the genome. (B) Reads aligned to splice junctions.

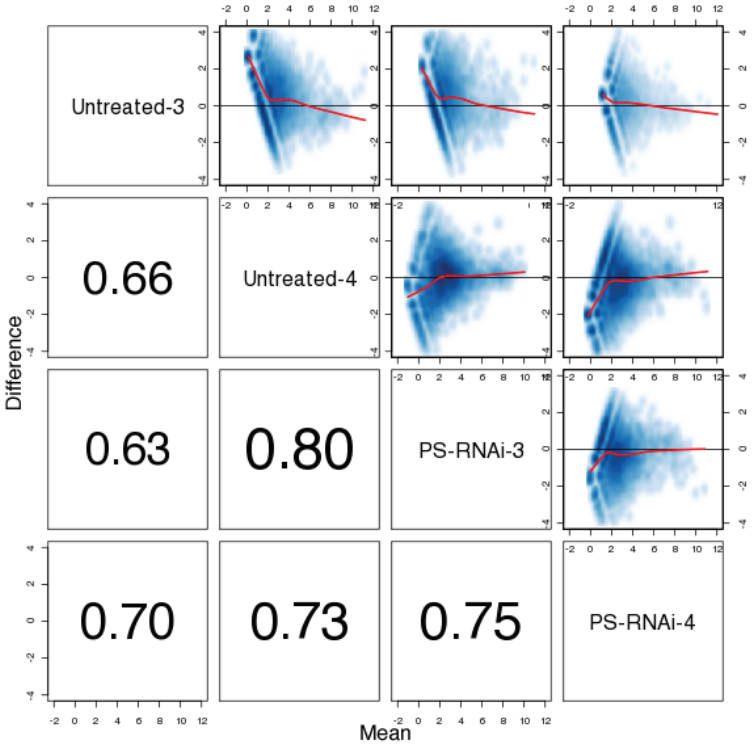
A.


Untreated-3
Untreated-4
PS-RNAi-3
PS-RNAi-4
0.88
0.86
0.87
0.88
0.88
0.89
Difference
Mean

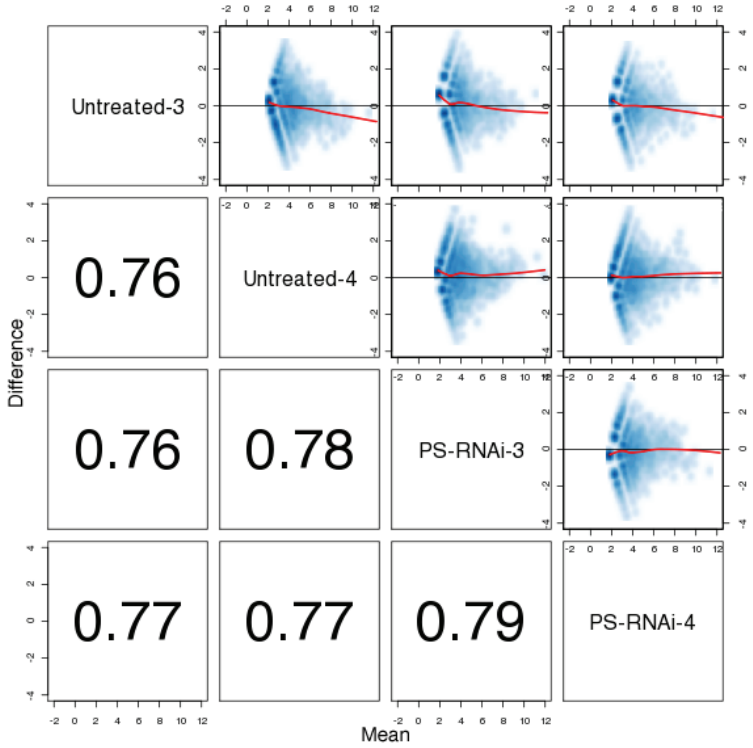
B.


Untreated-1
PS-RNAi-1 + PS-RNAi-2
0.98
Difference
Mean

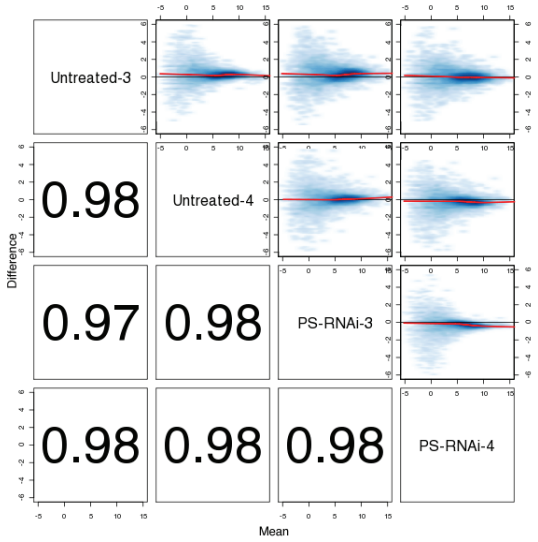
C.


Untreated-3
Untreated-4
PS-RNAi-3
PS-RNAi-4
0.66
0.63
0.80
0.70
0.73
0.75
Difference
Mean

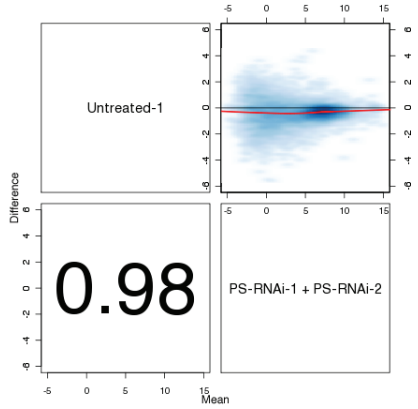
D.


Untreated-3
Untreated-4
PS-RNAi-3
PS-RNAi-4
0.76
0.76
0.78
0.77
0.77
0.79
Difference
Mean

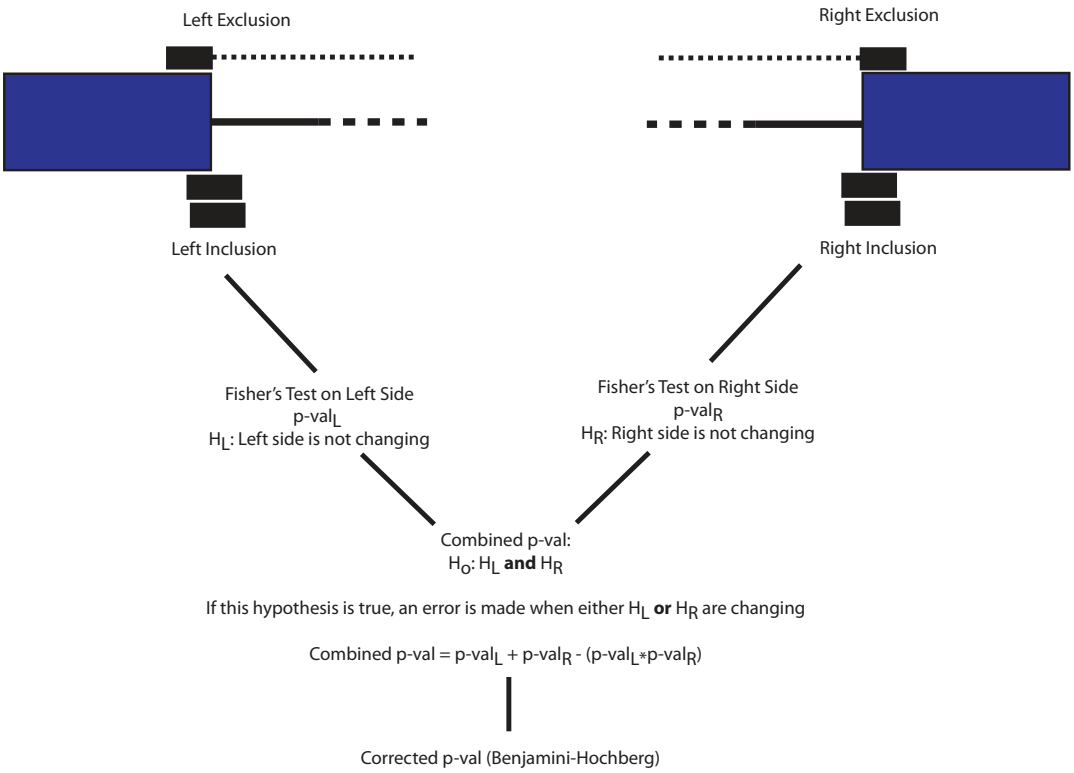
E.



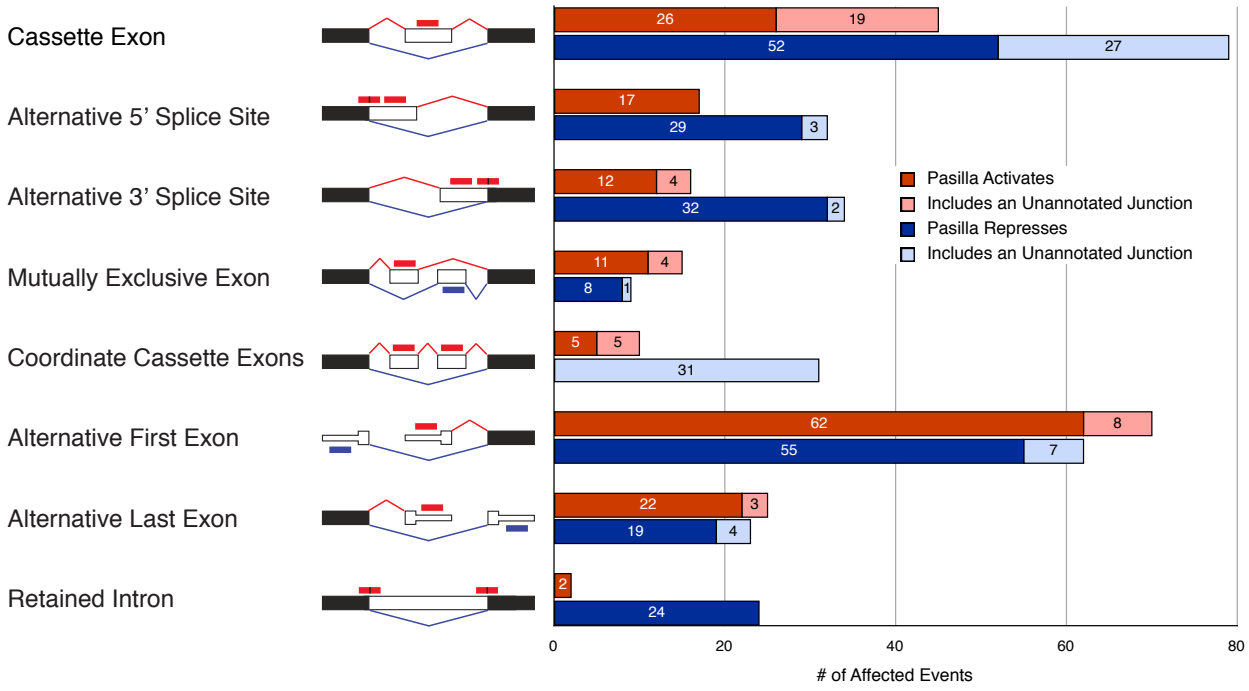
F.



Supplemental Figure 5. Mean-Difference plots for untreated and *ps(RNAi)* samples in both single and paired-end reads. The scale for all plots is $\log_2$. Lowess curves are shown in red. Labels refer to the biological sample number. If each sample was sequenced in multiple lanes, the sum of the counts for each junction and gene were used. For junctions, an RPM (reads per million) value was calculated for each junction. For genes, the RPKM (reads per kilobase per million) value (Mortazavi et al. 2008) was calculated. The value on the lower panels indicates the Pearson's correlation coefficent between samples. (A) RPM of junctions, paired-end samples. (B) RPM of junctions, single read lanes. (C) RPM of junctions, read 1s salvaged from paired-end alignments that were treated as single reads. (D) RPM of junctions, read 2s salvaged from paired-end alignments that were treated as single reads. (E) RPKM of genes, paired-end samples. (F) RPKM of genes, single read samples.

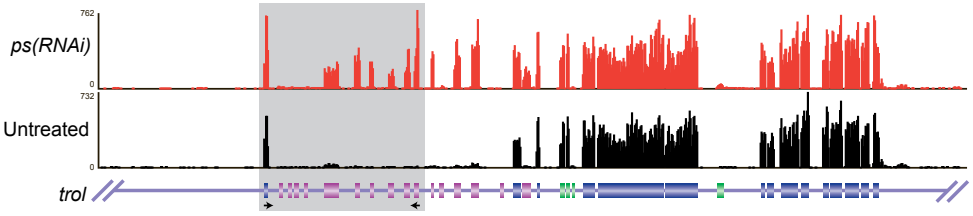


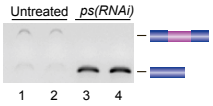
Supplemental Figure 6. Method used to identify retained intron events.



Supplemental Figure 7. 494 Splicing Events Affected by Pasilla. A set of non-redundant splicing events are shown in Figure 2. These splicing events had a corrected p-val 0.05 and are available in the Supplemental Materials.

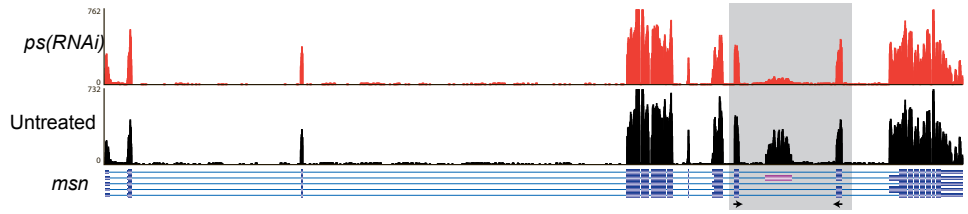
A) Novel, Coordinately Regulated Exons

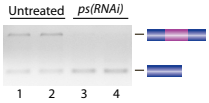




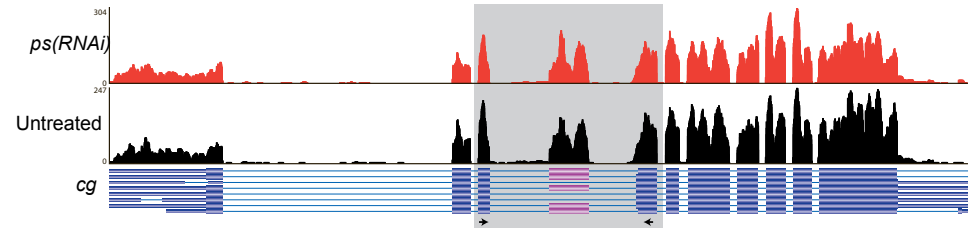
	Untreated	*ps(RNAi)*
Inclusion	136	3,573
Exclusion	52	0

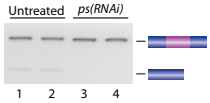
B) Cassette Exons



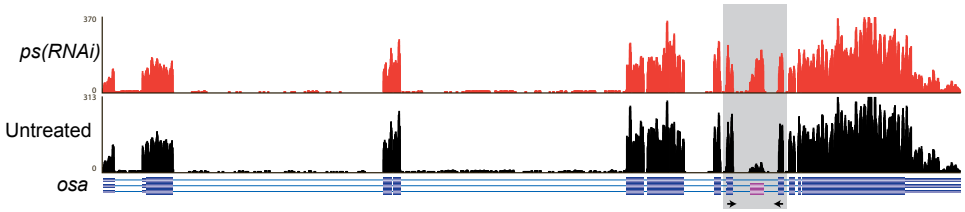


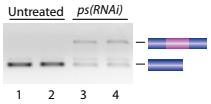
	Untreated	*ps(RNAi)*
Inclusion	4,384	821
Exclusion	130	208





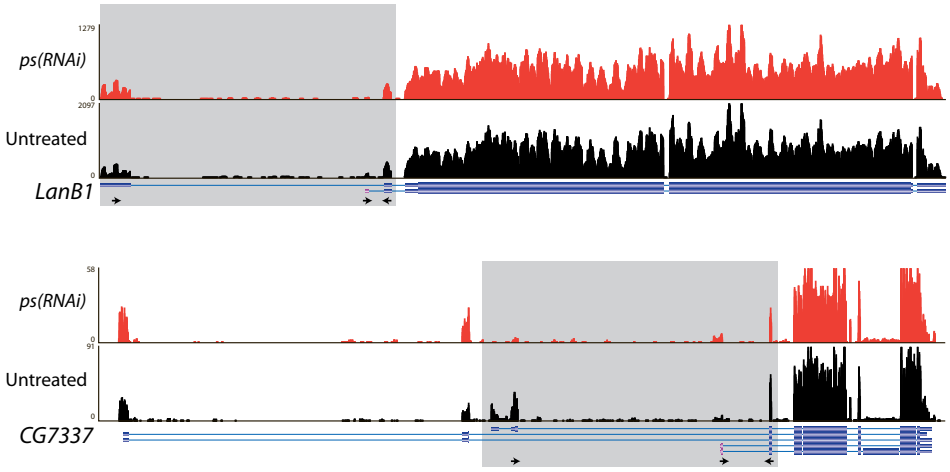
	Untreated	*ps(RNAi)*
Inclusion	865	831
Exclusion	9	0





	Untreated	*ps(RNAi)*
Inclusion	177	1,253
Exclusion	140	13

C) Alternative First Exons

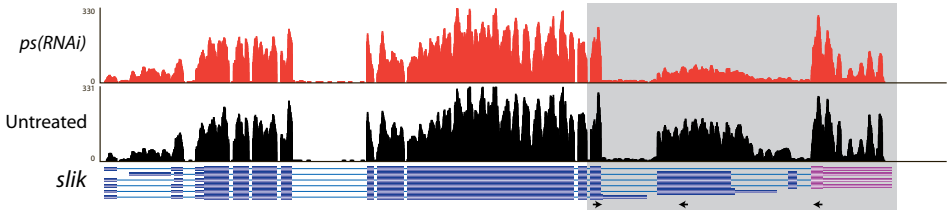


	Untreated	ps(RNAi)
Proximal	81	10
Distal	1,417	1,821

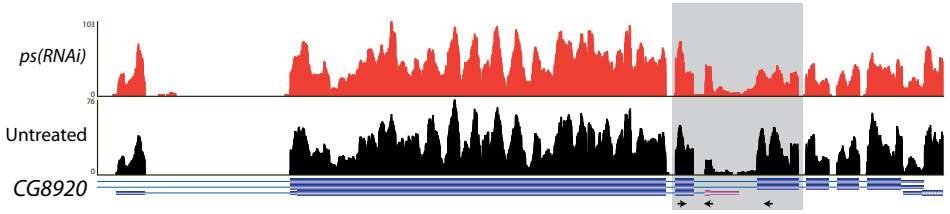
	Untreated	ps(RNAi)
Proximal	276	28
Distal	53	72

Alternative Last Exons

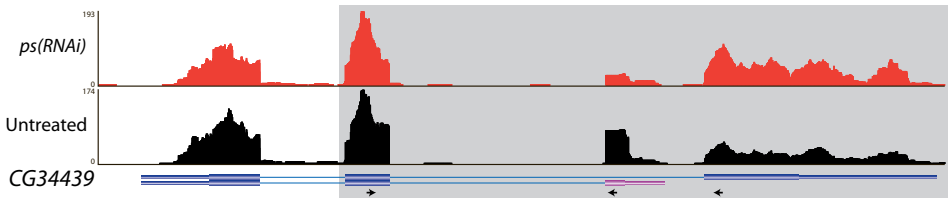
D) Alternative Last Exons



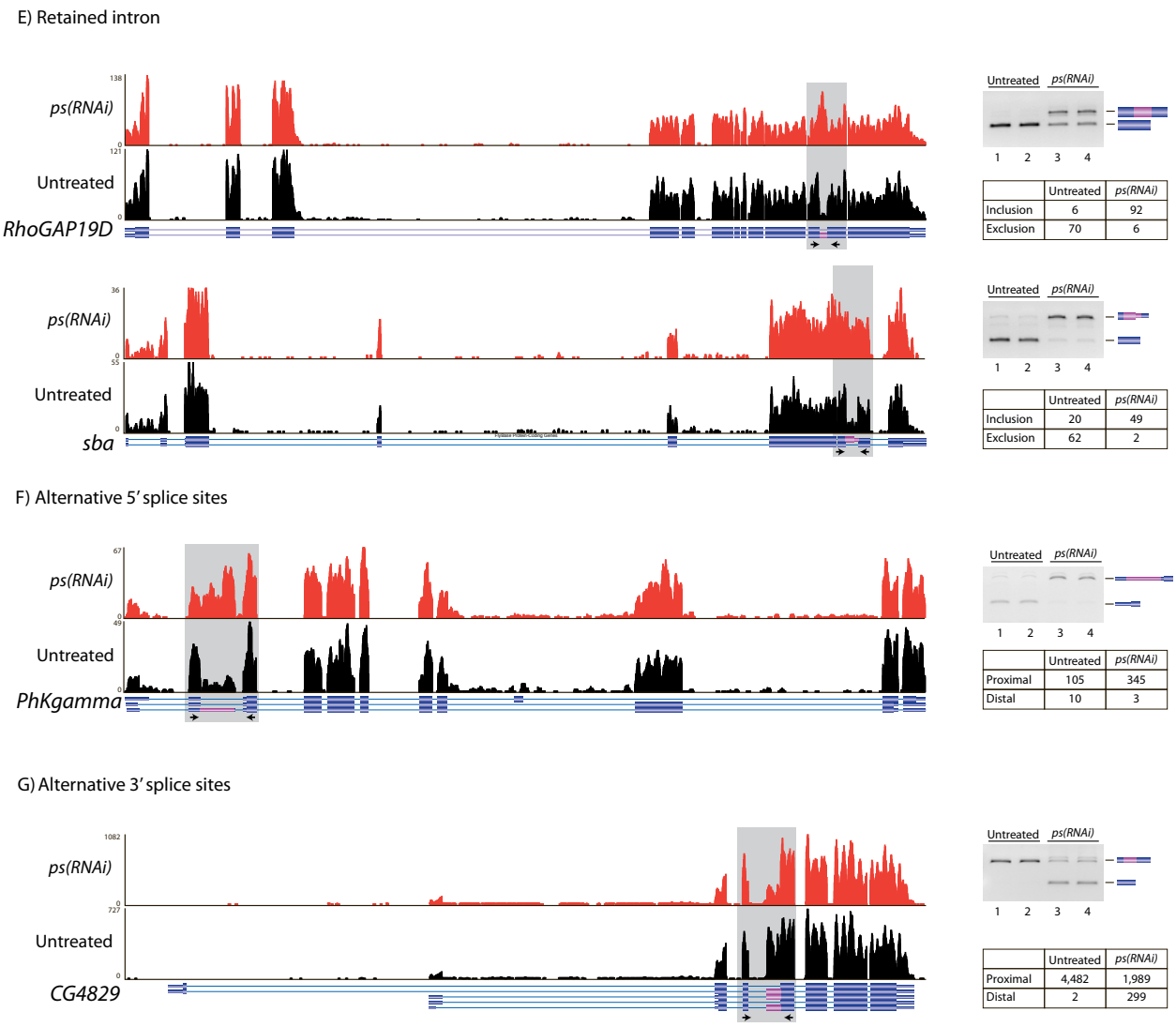
	Untreated	ps(RNAi)
Proximal	3,742	1,177
Distal	3,753	3,156



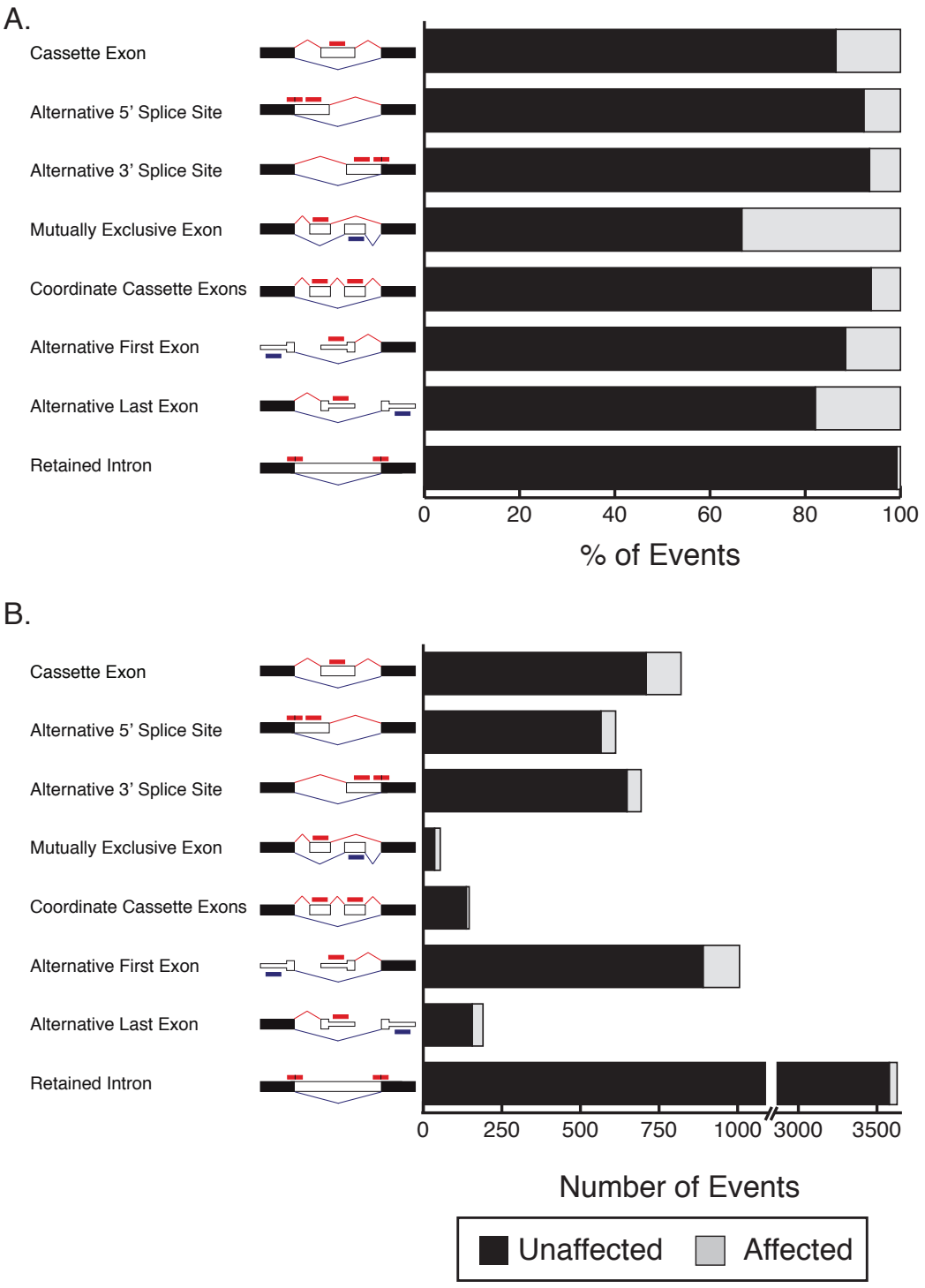
	Untreated	ps(RNAi)
Proximal	16	82
Distal	246	188



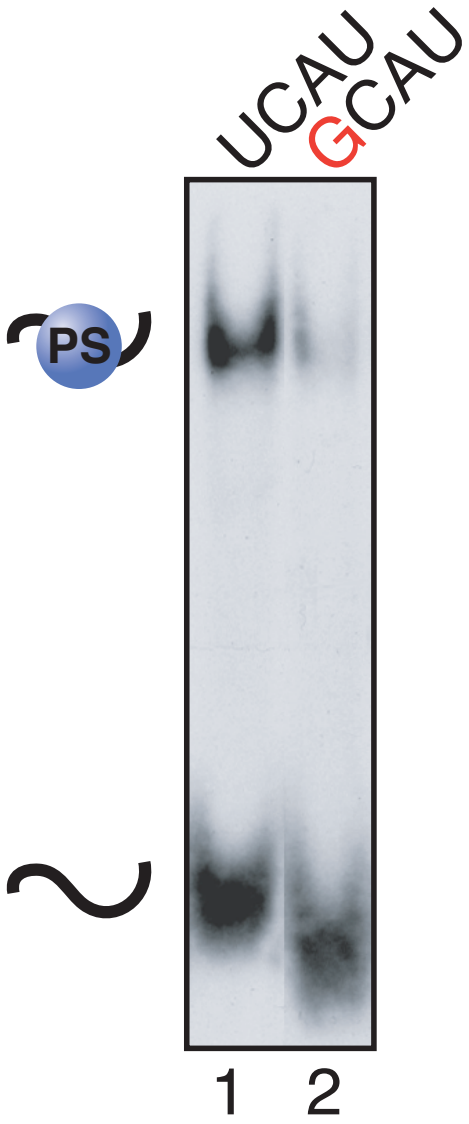
	Untreated	ps(RNAi)
Proximal	102	60
Distal	290	464



Supplemental Figure 8. Validation of PS-Regulated Splicing Events Identified by RNA-Seq. A) *trol*, an example of a splicing event that is annotated as a set of 9 constitutive exons, which we find to be coordinately skipped in untreated S2-DRSC cells, but coordinately activated in *ps(RNAi)* cells. B) *msn, cg* and *osa* are three examples of genes containing cassette exon affected by PS. LanB1 and CG7337 (C) are genes containing alternative first exons while *slik, CG8920,* and *CG34439* (D) are genes containing alternative last exons that are affected by PS. (E) *RhoGAP19D* is a gene containing a PS-regulated retained intron. *PhKgamma* (F) and *CG489* (G) are genes containing PS-regulated alternative 5' and 3' splice sites, respectively.

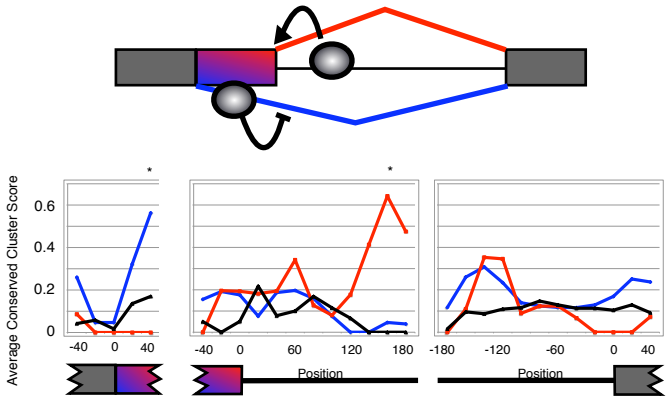


Supplemental Figure 9. Types of Splicing Events Affected by Pasilla. (A) The percentage of each class of expressed spicing event that are unaffected and affected in the *ps(RNAi)* sample is graphed. (B) The number of splicing events of each class that are unaffected and affected in the *ps(RNAi)* sample is graphed.



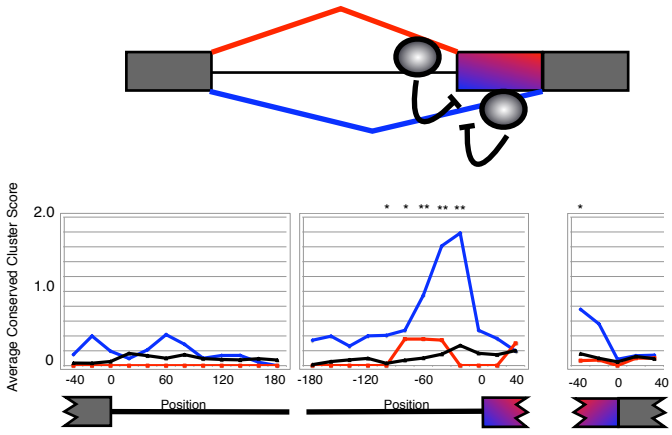
Supplemental Figure 10. Pasilla binds to YCAY containing RNA. RNAs containing three repeats of UCAU or GCAU were incubated in the presence of recombinant PS and the reactions resolved on a non-denaturing polyacrylamide gel. The locations of the unbound RNA and the RNA-protein complex are indicated.

A.



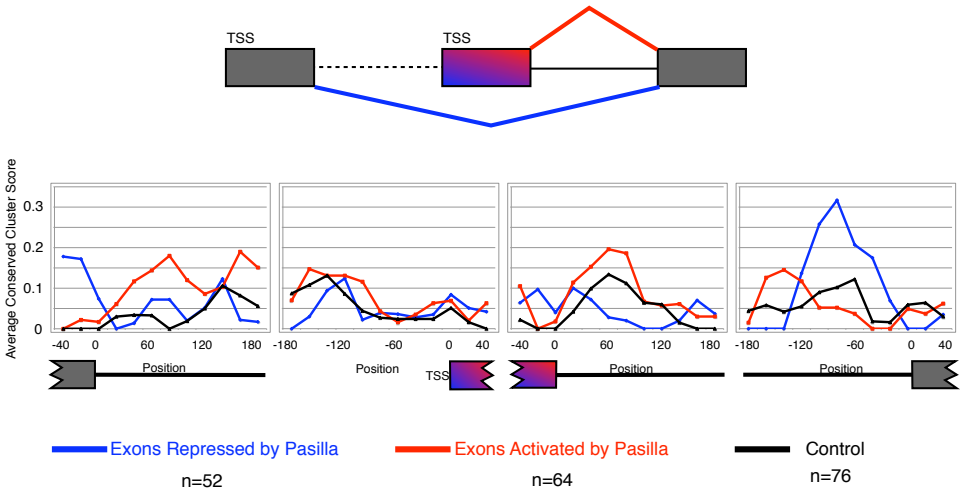
Exons Repressed by Pasilla n=30 | Exons Activated by Pasilla n=17 | Control n=97

B.

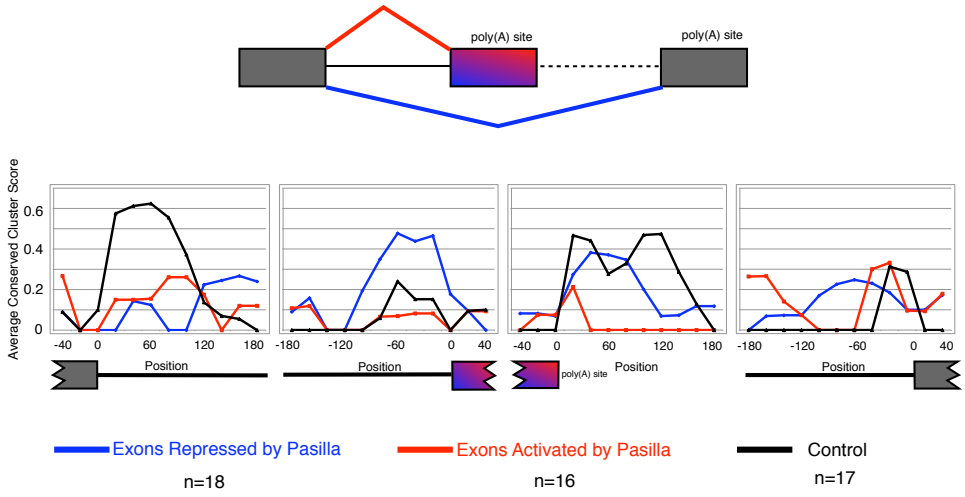


Exons Repressed by Pasilla n=29 | Exons Activated by Pasilla n=16 | Control n=172

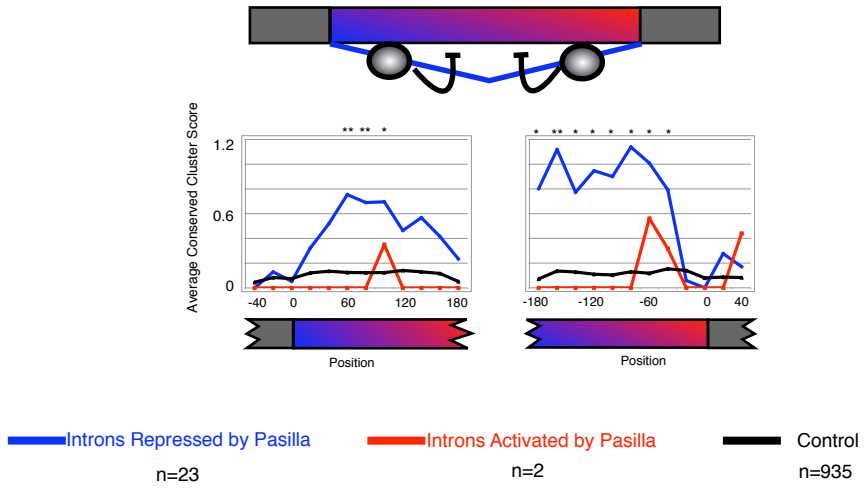
C


TSS
TSS
Average Conserved Cluster Score
0.3
0.2
0.1
0
-40 0 60 120 180
-180 -120 -60 0 40
Position
TSS
Exons Repressed by Pasilla
n=52
Exons Activated by Pasilla
n=64
Control
n=76

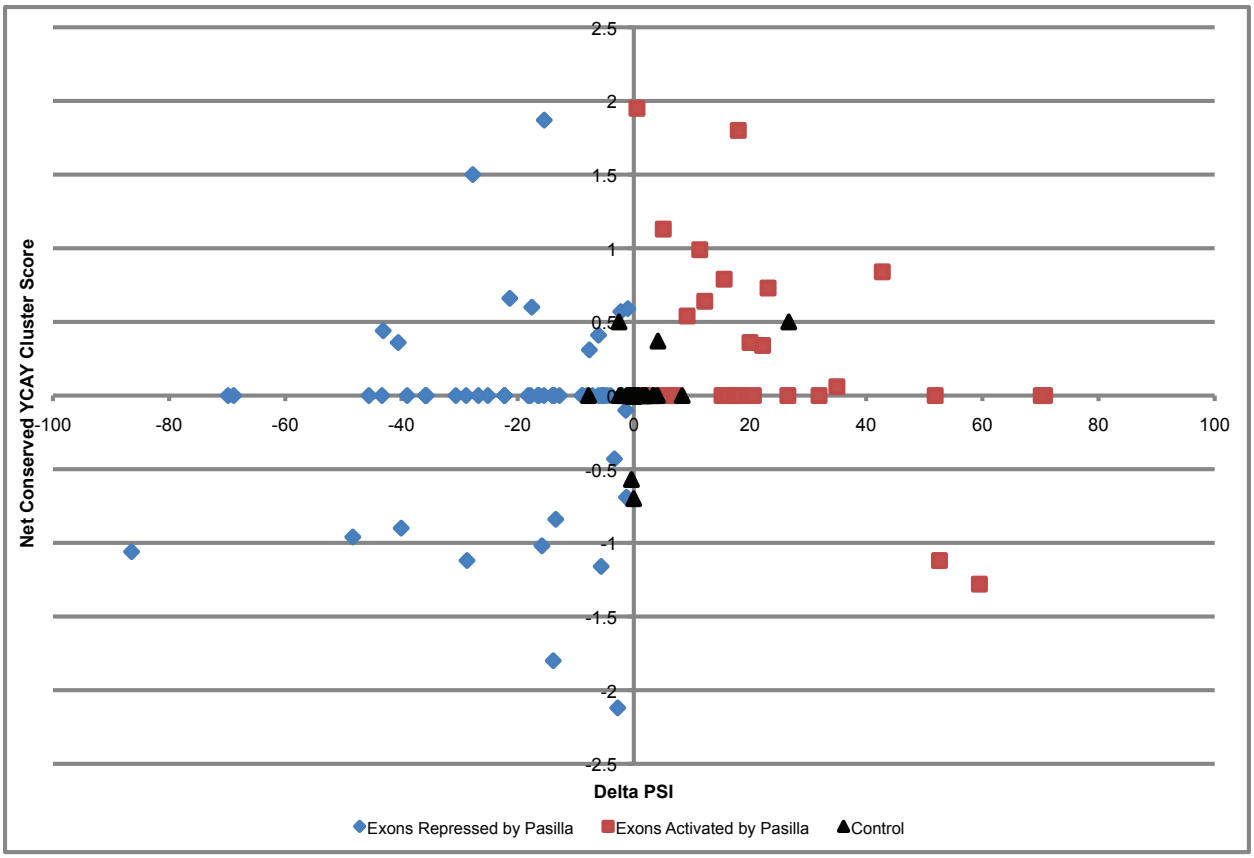
D


poly(A) site
poly(A) site
Average Conserved Cluster Score
0.6
0.4
0.2
0
-40 0 60 120 180
-180 -120 -60 0 40
Position
poly(A) site
Exons Repressed by Pasilla
n=18
Exons Activated by Pasilla
n=16
Control
n=17

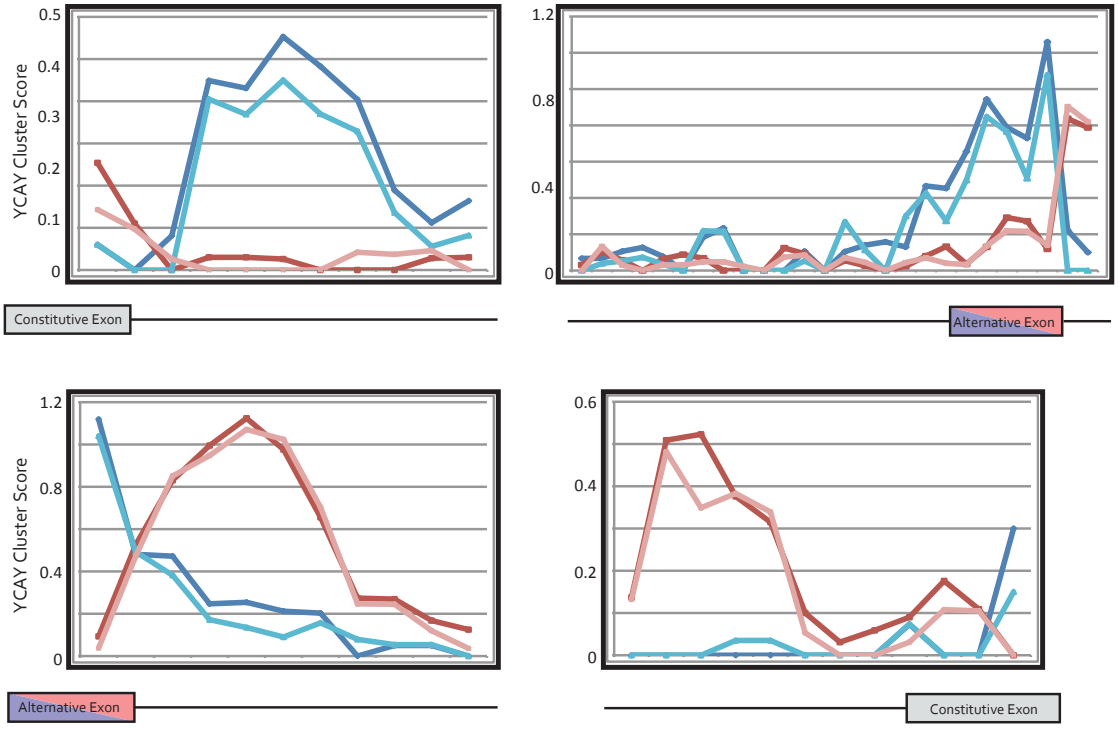
E.



Supplemental Figure 11. YCAY Cluster Maps for (A) alternative 5' splice sites, (B) alternative 3' splice sites, (C) alternative first exons, (D) alternative last exons, and (E) retained introns. Each position in the graph represents the average conserved YCAY cluster score, within a centered sequence window of 45 nt. The conserved YCAY cluster score was calculated for exons that are activated by PS, repressed by PS, and a set of control exons (Fisher's test, corrected p-value ≥ 0.95). Only regions adjacent to introns >400 nt were used for scoring. Positions with high cluster scores are indicated by an asterisk (Wilcoxon-rank sum test; uncorrected p-val < 0.01) or double asterisk if a position has an enrichment of YCAY clusters given a more stringent cutoff (Bonferroni-corrected p-val < 0.05).



Supplemental Figure 12 Net conserved YCAY cluster score versus change in percent spliced in (delta PSI) for all cassette exon events that were repressed by Pasilla (n=71) or activated by Pasilla (n=40). The control exons (n=90) are cassette exons that had no change in splicing (Benjamini-Hochberg corrected p-val > 0.95)



Inclusion Skipping

Predefined score method: YCAY cluster score as described in Supplemental Methods from Ule *et al.* 2006

Counting method: YCAY cluster score as used to create figures in Ule *et al.* 2006
(http://splicing.rockefeller.edu/map; personal communication, Jernej Ule and Robert Darnell)

YCAY cluster score used in figures for Ule *et al.* 2006 and as described in personal communication (counting method):

YCAY(N)6YCAY(N)12YCAY(N)3YCAYCAYNYCAY

1 YCAYCAY
1 YCAYCAY(N)6YCAY
1 YCAY(N)6YCAYCAY
1 YCAY(N)6YCAY(N)2YCAY
4 YCAY(N)6YCAY
1 YCAY(N)2YCAY

Total Score = log(2*(1+1+1+1+4+1)) = $\log_{10}(18) = 1.3$

YCAY cluster score as described in Ule *et al.* 2006 (predefined score method):

$$\frac{1 \quad 4 \quad 2 \quad 8}{\text{YCAY(N)6YCAY(N)12YCAY(N)3YCAYCAYNYCAY}}$$

Total Score = log(1+4+8+2) = $\log_{10}(15) = 1.2$

Supplemental Figure 13. Comparing the "Scoring" method and "Counting" method for calculating a YCAY cluster score on reported (Ule et al. 2006) mouse NOVA-regulated exons.

Supplemental Datasets.

Supplemental Dataset 1. List of Pasilla-Affected Classified Alternative Splicing Events.

Supplemental Dataset 2. Readme file for Supplemental Dataset 1.

Supplemental Dataset 3. List of Pasilla-Affected Unclassified Alternative Splicing Events.

Supplemental Dataset 4. GO Enrichment Terms for PS-Affected Splicing Events.

Supplemental Dataset 5. Annotation Files used for Splicing Analysis.

Supplemental Table 1: Sequencing Data by Lane and Sample ID Description

Sample ID	Biological Sample Number	Run Date	Flow Cell ID	Lane Number	Single Read/Paired-End
PS-RNAi-1-1-s	PS-RNAi-1	7/15/08	308T2AAXX	4	Single Read
PS-RNAi-1-2-s	PS-RNAi-1	7/18/08	308UEAAXX	2	Single Read
PS-RNAi-1-3-s	PS-RNAi-1	8/15/08	30AYWAAXX	6	Single Read
PS-RNAi-1-4-s	PS-RNAi-1	8/15/08	30AYWAAXX	7	Single Read
PS-RNAi-1-5-s	PS-RNAi-1	8/15/08	30AYWAAXX	8	Single Read
PS-RNAi-1-6-s	PS-RNAi-1	8/21/08	308A0AAXX	7	Single Read
PS-RNAi-2-1-s	PS-RNAi-2	8/21/08	308A0AAXX	8	Single Read
PS-RNAi-3-1-p	PS-RNAi-3	12/23/08	30M2BAAXX	5	Paired End
PS-RNAi-3-2-p	PS-RNAi-3	12/23/08	30M2BAAXX	6	Paired End
PS-RNAi-4-1-p	PS-RNAi-4	12/23/08	30M2BAAXX	7	Paired End
PS-RNAi-4-2-p	PS-RNAi-4	12/23/08	30M2BAAXX	8	Paired End
Untreated-1-1-s	Untreated-1	7/15/08	308T2AAXX	1	Single Read
Untreated-1-2-s	Untreated-1	7/15/08	308T2AAXX	6	Single Read
Untreated-1-3-s	Untreated-1	8/15/08	30AYWAAXX	2	Single Read
Untreated-1-4-s	Untreated-1	8/15/08	30AYWAAXX	3	Single Read
Untreated-1-5-s	Untreated-1	8/15/08	30AYWAAXX	4	Single Read
Untreated-1-6-s	Untreated-1	8/15/08	30AYWAAXX	5	Single Read
Untreated-3-1-p	Untreated-3	11/14/08	30MNEAAXX	2	Paired End
Untreated-3-2-p	Untreated-3	12/23/08	30M2BAAXX	2	Paired End
Untreated-4-1-p	Untreated-4	12/23/08	30M2BAAXX	3	Paired End
Untreated-4-2-p	Untreated-4	12/23/08	30M2BAAXX	4	Paired End

Supplemental Table 2. Primer sequences used in this study.

Supplemental Table 2a. Standard Primers

Primer	Sequence
T7	TAATACGACTCACTATAGGG
SP6	ATTTAGGTGACACTATAG
M13 Forward	GTAAAACGACGGCCAG
M13 Reversse	CAGGAAACAGCTATGAC

Supplemental Table 2b. Gene-specific primers for dsRNA cloning

Primer	Sequence
CG8144dsFP	ATGTCCAAGTCCTGAAACCG
CG8144dsRP	CGCTCCTGTGTCTGTTTTGA

Supplemental Table 2c. Gene-specific primers for validating the efficiency of RNAi

Primer	Sequence
CG8144FP	CCACGGAGGCTATCATGGTC
CG8144RP	ACGACACATCCGTGGGCTTC

Supplemental Table 2d. Gene-specific primers for validating splicing changes

Primer	Sequence
CG5295FP	GAAGGCTCTGTCGAAGATCAC

Primer	Sequence
CG5295RP	GATCACTTAGCACACTAGCATC
CG16973FP	TGGCGATCGGACTTTGATCATG
CG16973RP	AGGACGCTGCTGGTCCTTGTC
CG7467FP	TCCGGTAGCAAGTGGACCACAG
CG7467RP	GATATTGTCCTGGACCCTGTC
CG1828FP	AGGCTGCTGAGGAAGACCGCAAC
CG1828RP	CTTGCGATCCTTGTCCTTGC
CG8367FP	GGTGGTCAAGAAGCAGGAC
CG8367RP	GGACAACTCTAATGTGCCTG
CG7123F1P	ATTCGTTGGACGCGGATGC
CG7123F2P	AGTGTTTGCCGCATCAACG
CG7123RP	GTCCCACTGCCAGCTGAGCAG
CG4829FP	GAAGAAGCTAACCATCGTGC
CG4829RP	CACATTGGGCAAGACGTCGTC
SesAntFP	CACGCGGCTTCGTCTGATTCTG
AntRF	CACCTGCAGAATGAGCTTCAC
SesRP	GAAGACCTGCTTGTACTTGTCC
CG1412FP	ATGCTGCTGCACAACCTGTC
CG1412RP	CGATGAGCTGAACTTGGTAG
CG13598FP	GTGGGATCTACAAGCATTCG
CG13598RF	CCAATGGCAAAGGATCGTGG
CG7337FP1	CAGTCAGCCCTAAAGTATGCAG
CG7337FP2	CAAGTTCCCAGCGAATTACGAG
CG7337RP	GTTTTGCGTTGAAGAGCACCAC

Primer	Sequence
CG1830FP1	GAGAGGATCCACTTCCAATCG
CG1830RP	CTGCCCAGAATCTCCTTGG
CG4527FP	TGGAGCACGAACACTCCAAG
CG4527RP1	AGAATGCGTGAGGATCGCTC
CG4527RP2	CTGCATAGCTGTCATAGCTG
CG8920FP	GGAGGTCATATGCTTCATCG
CG8920RP1	ACTTATAGGGCACCAGTCAC
CG8920RP2	ATCCTTGAGCATGAGAGACG
CG34439FP	TCAGCCGAGTTATGTACCAG
CG34439RP1	CTATCACTTCCGAACTGCAC
CG34439RP2	TCCTGTCACGAGCACAGGGTTG

Supplemental Table 3: Read counts per sample

	Untreated-1	Untreated-3	Untreated-4	Untreated Total	ps(RNAi)-1	ps(RNAi)-2	ps(RNAi)-3	ps(RNAi)-4	ps(RNAi) Total
Total sequenced single reads	32,989,325			32,989,325	35,158,667	2,112,638			37,271,305
Total sequenced paired-end reads		10,575,821	12,230,102	22,805,923			12,263,470	12,467,158	24,730,628
Uniquely aligned single reads†	24,140,581	930,830	3,161,078	28,232,489	20,866,815	965,365	3,898,495	1,393,333	27,124,008
Uniquely aligned paired-end reads		7,781,735	7,076,985	14,858,720			6,084,523	9,272,469	15,356,992

†Includes both single read sequence and paired-end reads with an unalignable mate.

BR - Biological Replicate