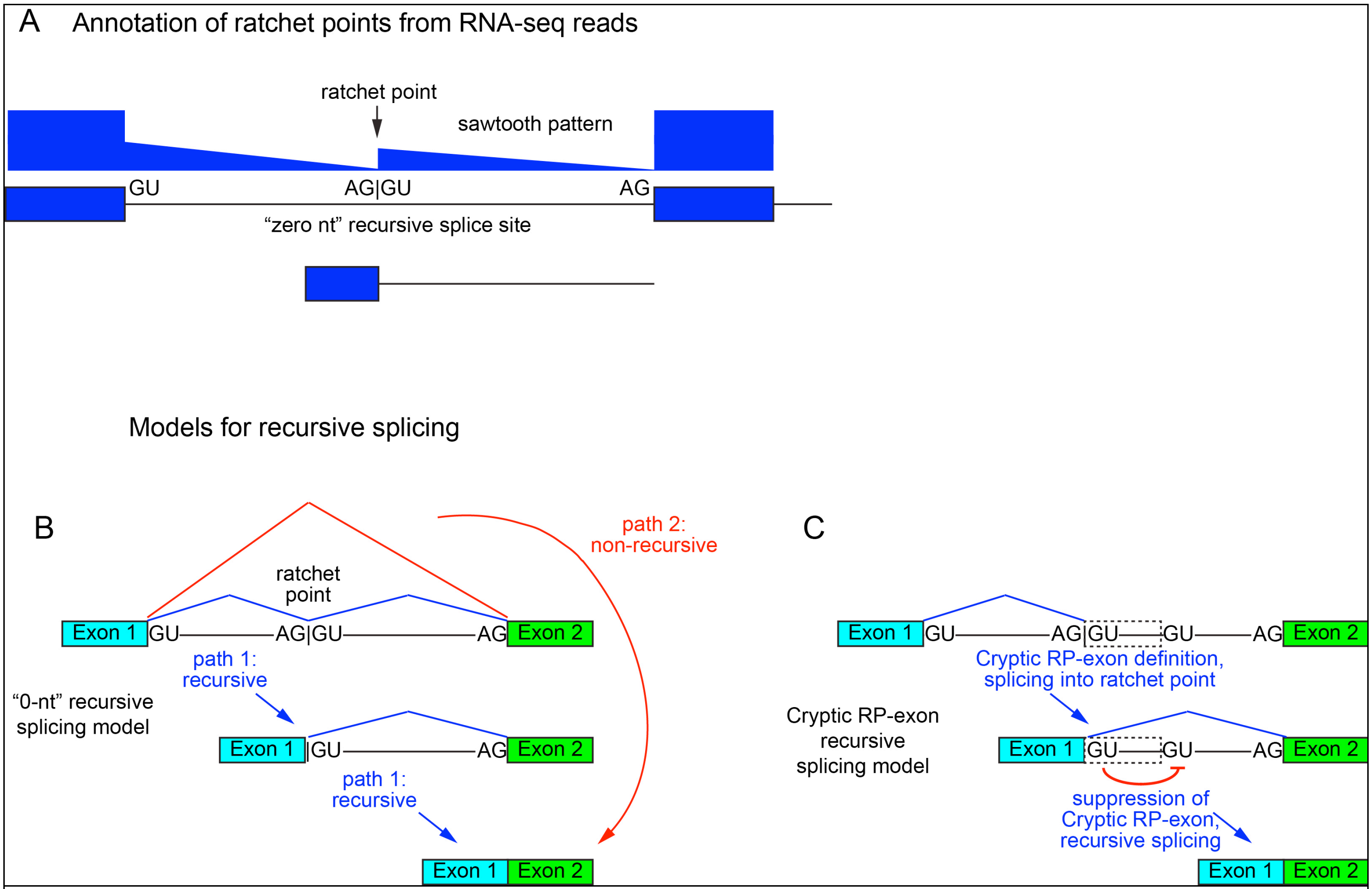
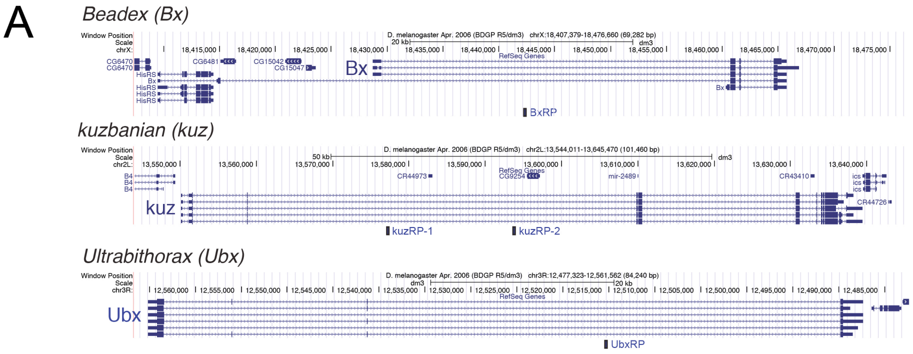

A
Annotation of ratchet points from RNA-seq reads
ratchet point
sawtooth pattern
GU
AG|GU
AG
"zero nt" recursive splice site
Models for recursive splicing
B
ratchet point
path 2: non-recursive
Exon 1
GU
AG|GU
AG
Exon 2
path 1: recursive
"0-nt" recursive splicing model
Exon 1
|GU
AG
Exon 2
path 1: recursive
Exon 1
Exon 2
C
Exon 1
GU
AG|GU
GU
AG
Exon 2
Cryptic RP-exon definition, splicing into ratchet point
Cryptic RP-exon recursive splicing model
Exon 1
GU
GU
AG
Exon 2
suppression of Cryptic RP-exon, recursive splicing
Exon 1
Exon 2

Supplementary Figure 1

Evidence and mechanistic models for recursive splicing.

(A) Sawtooth RNA-seq patterns are indicative of recursive splicing intermediates. Generally, RNA-seq coverage in introns is reflective of nascent transcription and resembles right-angled triangles, with highest coverage at the 5' end and lowest at the 3' end of the intron. However, introns that undergo recursive splicing consist of multiple intronic segments, each with its own right-angled triangle coverage, producing a sawtooth pattern. This property has been exploited to infer recursive splicing and annotate RPs. (B) Models for processing introns with RPs. It is conceivable that introns that contain RPs will be processed in one of two ways. First, the RP is utilized constitutively (path 1) and the intron is removed in two sequential steps. Second, the RP is skipped such that the entire intron is spliced out in one step (path 2). (C) A molecular model for recursive splicing. We propose that recursive splicing proceeds by first defining a cryptic RP-exon, which is specified by the RP splice acceptor and a downstream cryptic splice donor. Definition of the cryptic RP-exon allows removal of the first intron segment and production of the recursive intermediate. In the second splicing reaction, we propose that the regenerated RP splice donor outcompetes the cryptic splice donor, thereby removing the whole intron and ligating neighboring exons.

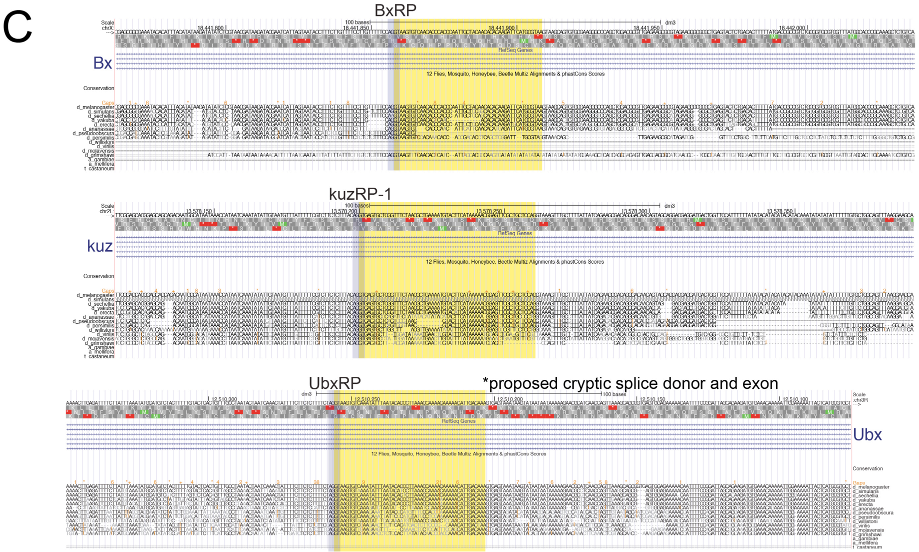


B

kuz[RP] - 6 nt deletion
WT: GAGCAGCAGACAATGGCATAATAAACCATAATCAAATATATTGTAATGTTTATTTTTCGTCTTCTCTTTAC**AGGTGAGT**GCTCGGTTTCTAA
MT: GAGCAGCAGACAATGGCATAATAAACCATAATCAAATATATTGTAATGTTTATTTTTCGTCTTCTCTTTACAG------GCTCGGTTTCTAA

Bx[RP] – 1 nt substitution, 2 nt insertion
WT: AATACCTTTCTTGTTTTCCTTGTTTTTCC**AGGT--AAGT**GTCAACACCCACCCAATTGCTACAACACACAAGAT
MT: AATACCTTTCTTGTTTTCCTTGTTTTTCCAGGAAAAAGTGTCAACACCCACCCAATTGCTACAACACACAAGAT

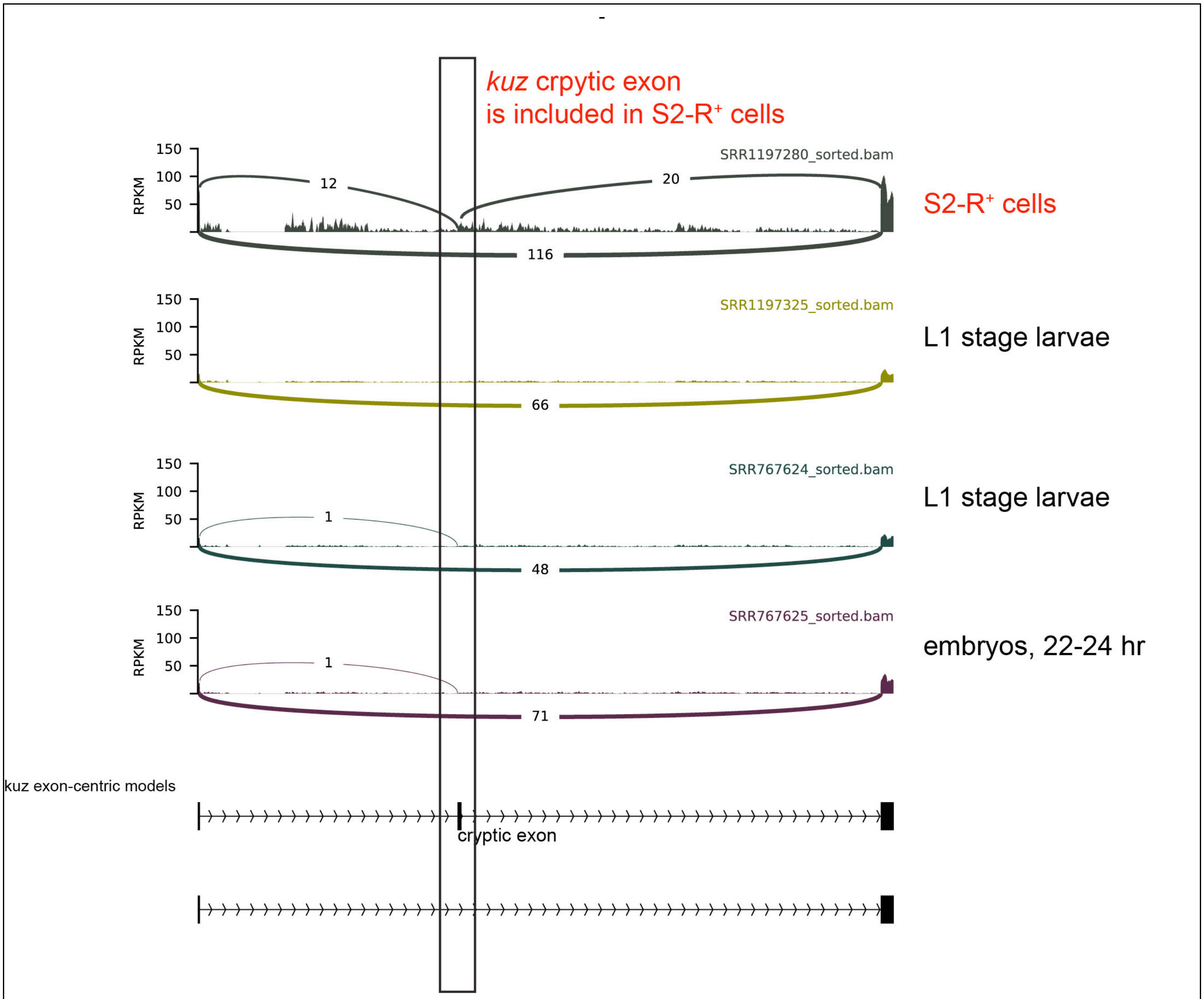
Ubx[RP] – 38 nt insertion
WT: TCAAACTATTTTCTTCTCTTTTCT**AG**--------------------------------------**GTAAGT**GTCAAATATTTAATACACCC
MT TCAAACTATTTTCTTCTCTTTTCTAGAATTCTGTCAAATATTTAATAACCCTTAAACCAAACAGGTAAGTGTCAAATATTTAATACACCC



Supplementary Figure 2

Genes with RPs were manipulated to identify cryptic exons.

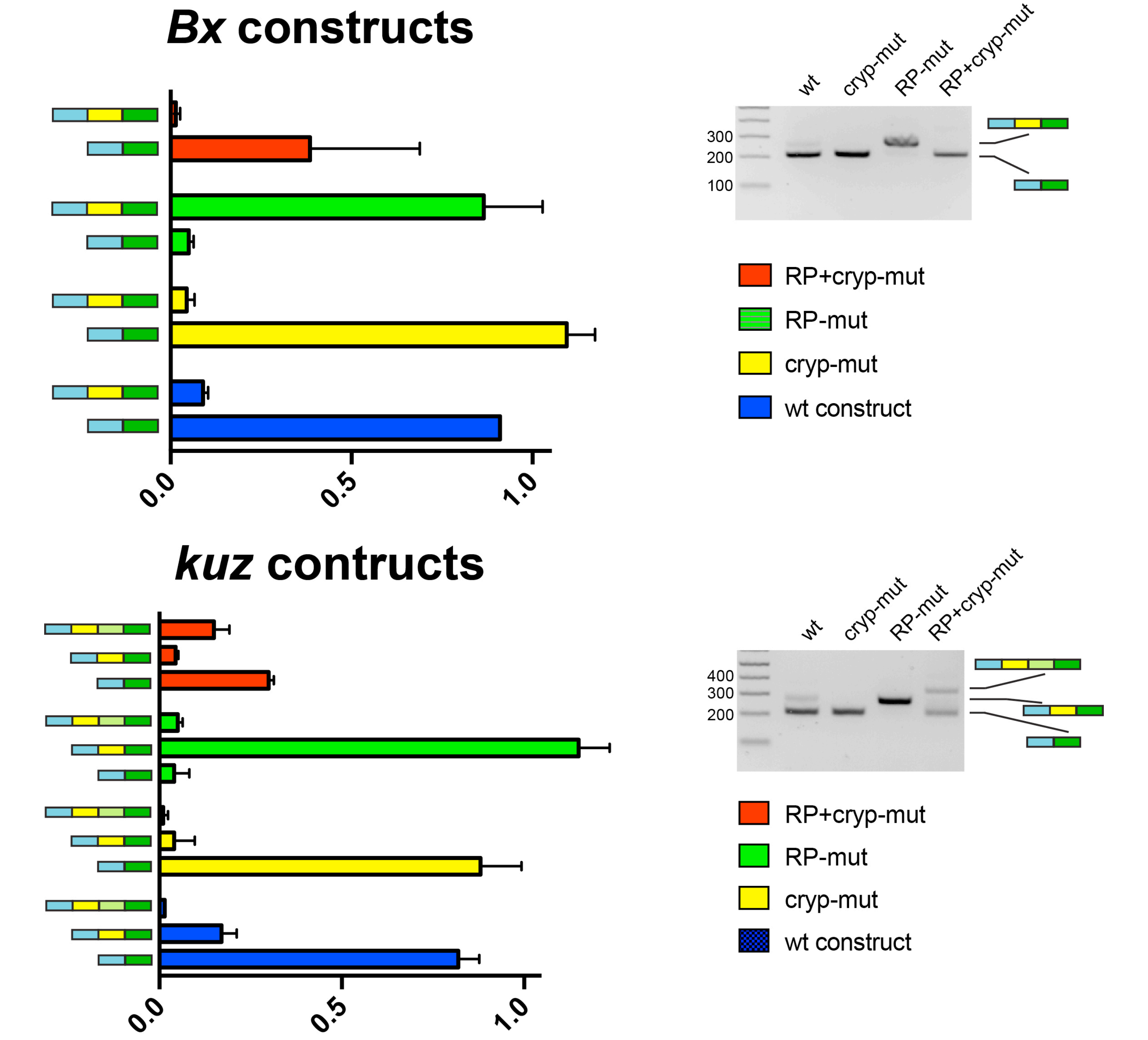
(A) UCSC genome browser screenshots display the three genes (*Bx*, *kuz,* and *Ubx)* manipulated in this study, including the approximate genomic locations of RPs within long host introns. (B) Nature of mutant alleles along with sequence alignments. (C) UCSC genome browser nucleotide-level screenshots of mutated RPs (grey highlight) along with cryptic exons detected in mutants (yellow highlight). **Ubx[RP]* is an insertion mutant, which separates the RP splice acceptor and donor sites by 38nt.This 38nt insertion is retained in mutant animals. However, bioinformatic analysis has identified a naturally occurring cryptic exon and splice donor as indicated.



Supplementary Figure 3

***kuz* cryptic exon is retained in S2-R$^+$ cells.**

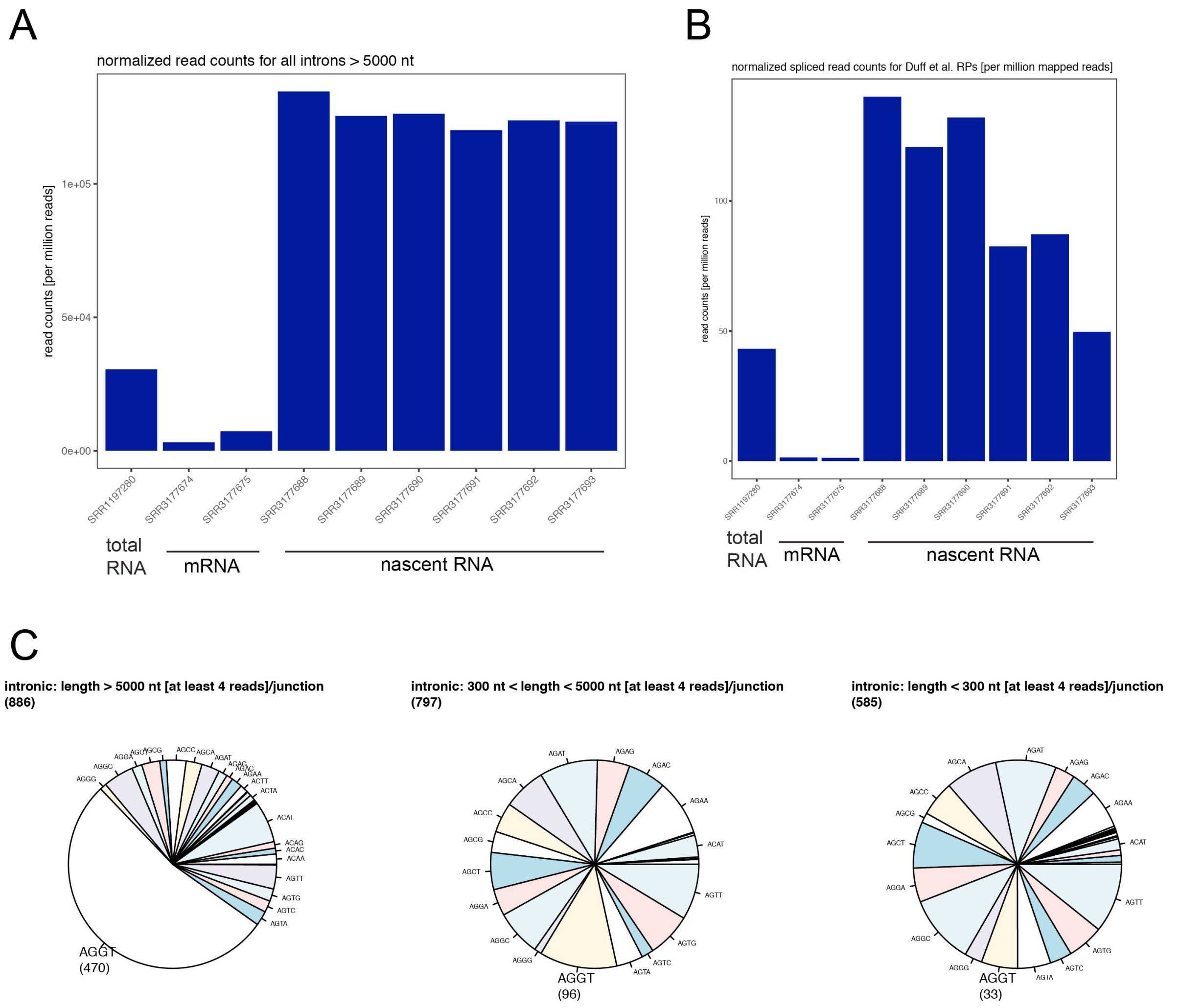
Sashimi plots were used to display the usage of the cryptic exon in modENCODE total RNA-seq data from S2-R$^+$ cells, L1 stage larvae and 22-24hr embryos. Spliced reads can only be detected into and out of the cryptic exons in S2-R$^+$ cells, suggesting selective inclusion and/or stabilization here.



Supplementary Figure 4

Quantification of relative cryptic exon inclusion ratios from wildtype and mutant recursive splicing minigenes.

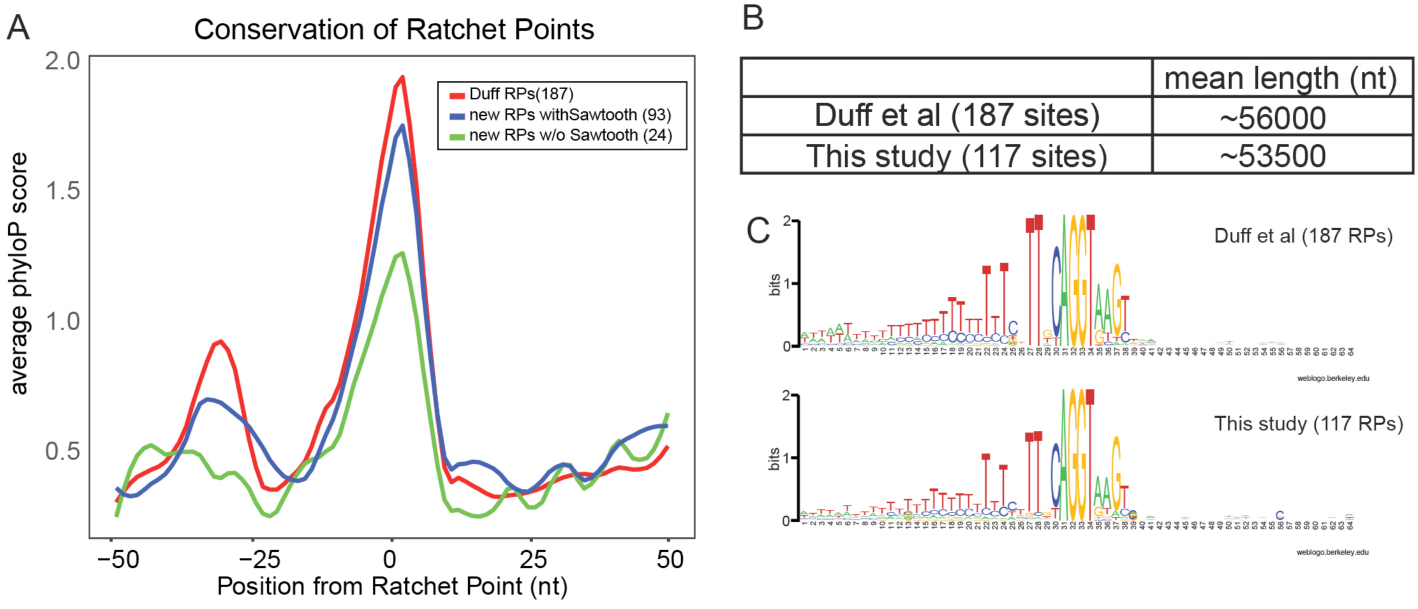
Bx and *kuz* minigenes that contained the indicated mutations (see Fig 2E, G) were transfected into S2 cells and subjected to rt-PCR analysis. Relative exon inclusion was calculated by normalizing the intensity of the mRNA band to all indicated bands in the same lane, and then scaled to total expression observed in wt lane. Representative gels are shown (see also main Figure 2).



Supplementary Figure 5

Nascent RNA-seq datasets are more suited for detection of RPs.

(A, B) Nascent RNA-, total RNA- and mRNA-seq datasets from S2 cells were evaluated for different criteria. (A) Nascent RNA datasets have higher coverage at intronic loci. Long intronic segments - with no overlapping genes - were identified and reads mapping to these regions were summed and normalized. (B) Junction spanning reads that mapped to RPs identified in Duff et al. were summed and normalized. (C) Junction spanning reads found from all nascent RNA-seq and GRO-seq were merged and those with 3' ends mapping to intronic regions were stratified by junction split (intron length) into three categories. For each category, pie charts were drawn to indicate tetranucleotide distributions at the 3' junction end. Note that AGGT, which resembles minimal ratchet point sequences are enriched only within the long intron category.

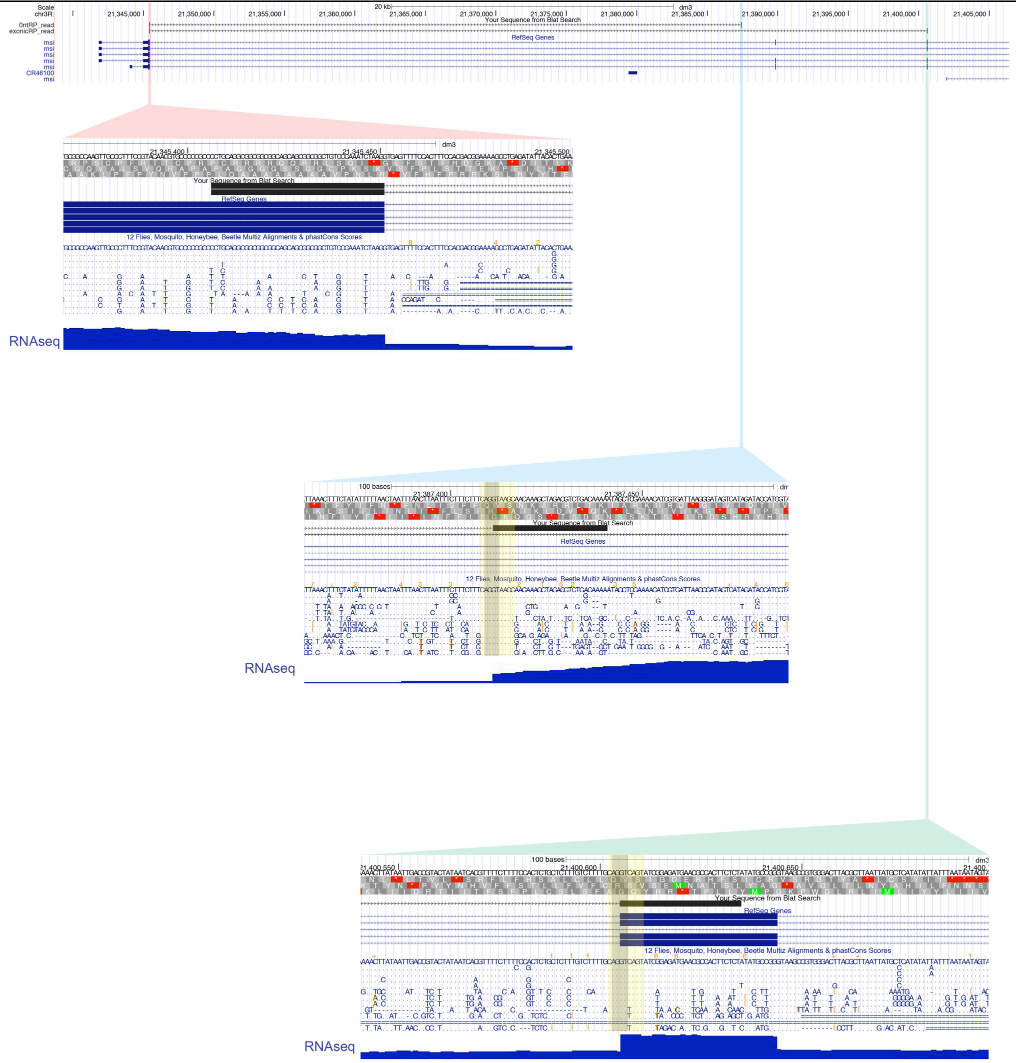


	mean length (nt)
Duff et al (187 sites)	~56000
This study (117 sites)	~53500

Supplementary Figure 6

Novel ratchet points share sequence, structural, and evolutionary properties of known ratchet points.

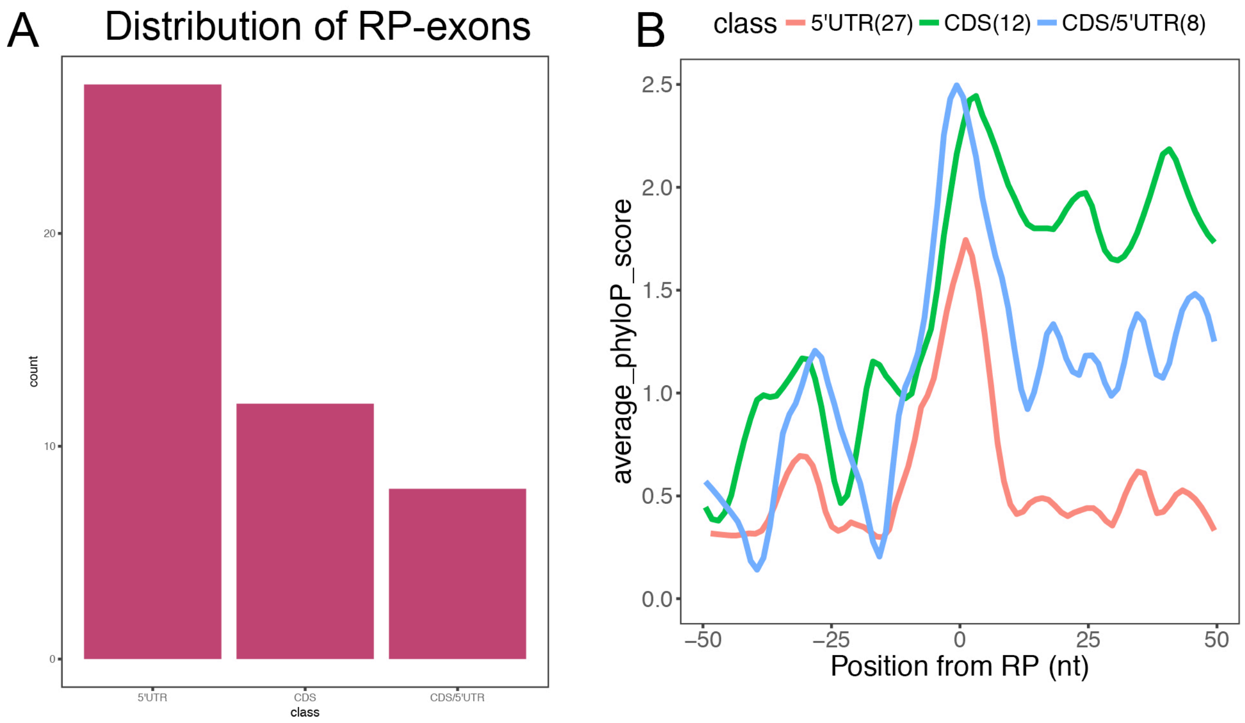
(A) Comparison of average phyloP scores for “0-nt” RPs. New RPs were grouped into two categories based on whether they had sawtooth patterns in RNA-seq data or not. (B) Average length of host introns for RPs found in Duff et all and this study. (C) Sequence logos for categorized RPs.



Supplementary Figure 7

Example of intronic RP and RP-exon annotation in the *msi* gene.

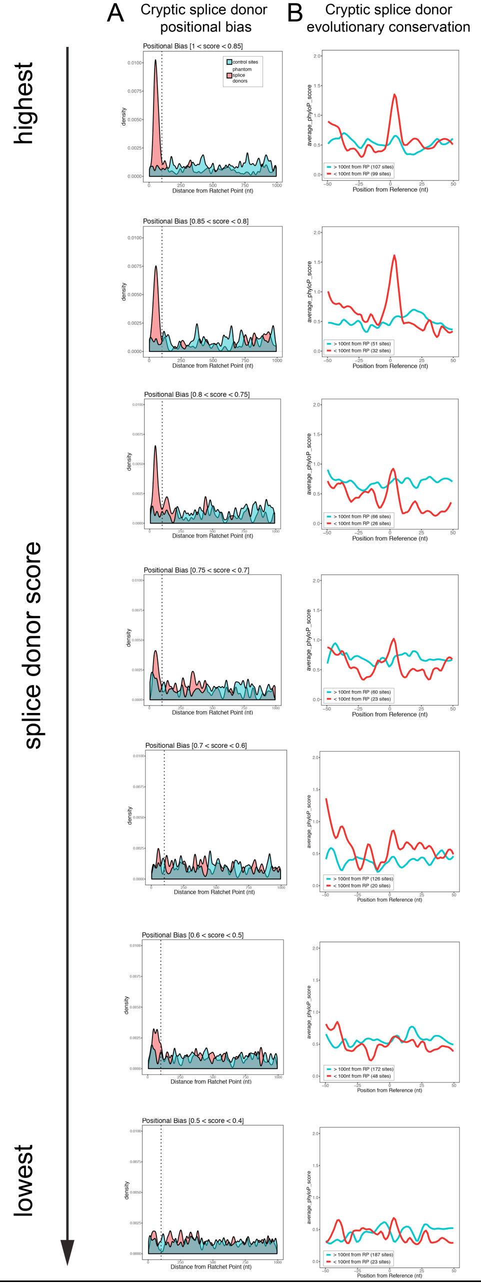
The BLAT tool in UCSC genome browser was used to map RNA-seq reads to *musashi (msi).* 5’ ends of reads map to a *msi* 5' exon (zoomed in shot highlighted in red). 3’ ends of one read maps to an intronic RP (blue highlight) and a zoomed in nucleotide-level screenshot is included in blue. 3’ ends of the other read maps to an RP-exon (green highlight) and a zoomed in nucleotide-level screenshot is included in green. Note the RNA-seq coverage in screenshots and that RP-exons have distinct exon coverage, whereas intronic RPs have sawtooth coverage pattern. The core AGGT splice acceptor-donor pairs are marked in gray, while the larger splice consensus motifs are highlighted in yellow.



Supplementary Figure 8

Conservation and coding properties of RP-exons by subcategory.

(A) Distribution of RP-exons according to location in gene models. (B) RP-exons were divided according to their location in 5'UTR, CDS, and 5'UTR/CDS (ones that contained alternate 5'UTR/start sites). The fully coding RP-exons have a high level of evolutionary conservation, and the set with partial coding potential exhibit an intermediate level of conservation.



C # unique RPs with at least one cryptic splice donor in the score category. Categories contain non-overlapping sets.

score	counts
>=0.85	98
0.8-0.85	30
0.75-0.8	24
0.7-0.75	19
0.6-0.70	11
0.5-0.60	24
0.4-0.50	9
others	89
total	304

Supplementary Figure 9

Positional bias and conservation of cryptic donors stratified by splice scores.

(A) Splice donors found downstream of RPs (cryptic splice donors) or 1000 control AGGTs sites were grouped based on NNSPLICE splice site strength. Plotted is the positional bias of splice donor site position relative to ratchet points. Substantial enrichment is observed in the ~40-80 window downstream of ratchet points, but not control AGGT sequences, at NNSPLICE scores down to 0.5-0.6 (B) Average phyloP scores of splice donor sites downstream of ratchet points (RPs) segregated into those that are <100nt from RPs and >100nt away from RPs. Clear local conservation is observed amongst groups of cryptic donors scored down to ~0.6. (C) Table showing the number of non-overlapping RPs with cryptic splice sites grouped by splice site score.