TITLE: Large-scale Analysis of Branchpoint Usage across Species and Cell Lines

Supplementary Table Legends

Authors: Allison J Taggart[1,4], Chien-Ling Lin[1,4], Barsha Shrestha[1], Claire Heintzelman[1], Seongwon Kim[1] and William G Fairbrother*[1,2,3]

Affiliations:
1 Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence 02912
2 Center for Computational Molecular Biology, Brown University, Providence 02912
3 contributed equally
* corresponding author

Supplemental Table 1 – Table of all branchpoints found in human introns (hg19 annotation). Read count position is annotated "x,y,(z)" where x is the number of reads at the -2 position, y is the number of reads in the -1 position, and z is the number of reads at the bulged branchsite. In the case of a multi-nucleotide bulge, read count position is annotated "x,y,(z1,z2,z3*)", in which z1,z2, and z3 and positions in the bulge, and the * is the predicted branch site. Unbiased source is used to describe lariat reads that were discovered in samples that did not use probes. RNAseq source points to the RNAseq accession code where each branchpoint was discovered. The counts for each RNAseq source are described (a;b), in which a is the total number of lariat read counts from that sample and b is the number of unique lariat reads.

Supplemental Table 2 – Table of branchpoints found in human introns (hg19) using only ENCODE RNASeq samples. Same format as Supplemental Table 1.

Supplemental Table 3 – Table of branchpoints found in mouse introns (mm9). Same format as Supplemental Table 1.

Supplemental Table 4 – Table of branchpoints found in *S. pombe* introns (EF2). Same format as Supplemental Table 1.

Supplemental Table 5 – Table of branchpoints found in human introns (hg19), all sources included. Binary quality metrics are described in each column with either a 1 (satisfies quality metric) or a 0 (does not satisfy quality metric). Overall quality score string contains character if it satisfies the following metric:

- U: fits U2-premRNA alignment model
- A: branchpoint is an A nucleotide
- P: branchpoint is in expected position (10 to 60 nucleotides upstream of 3'ss)
- R: more than one unique validating read
- S: branchpoint validated by more than one RNAseq source
- M: mutation present in at least one validating read
- B: branchpoint derived from unbiased RNAseq source

Excel equations can be added to table to select subset of branchpoints that fulfill metrics of choice. Example equation is in final column, pulling branchpoints that fit a U2-premRNA alignment model, are in the expected position, and are found in multiple RNAseq sources.

Supplemental Table 6 – Table of branchpoints found in human introns (hg19), ENCODE sources only. Binary quality metrics can be selected to pull a specific subset of branchpoints. Same format as Supplemental Table 5.

Supplemental Table 7 – Table of branchpoints found in mouse introns (mm9). Binary quality metrics can be selected to pull a specific subset of branchpoints. Same format as Supplemental Table 5.

Supplemental Table 8 – Table of branchpoints found in *S. pombe* introns (EF2). Binary quality metrics can be selected to pull a specific subset of branchpoints. Same format as Supplemental Table 5.

Supplemental Table 9 – Table of full branchpoint sequences inserted into minigene reporter constructs for Figure 2E.

Supplemental Table 10 – Branchpoint counts and motif counts stratified by RNAseq sample.

Supplemental Table 11 - Table of primers used for branchpoint validations.