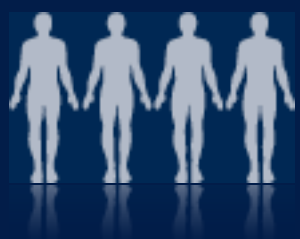
Wellcome Trust Centre for Human Genetics
Nuffield Department of Clinical Medicine

UNIVERSITY OF OXFORD

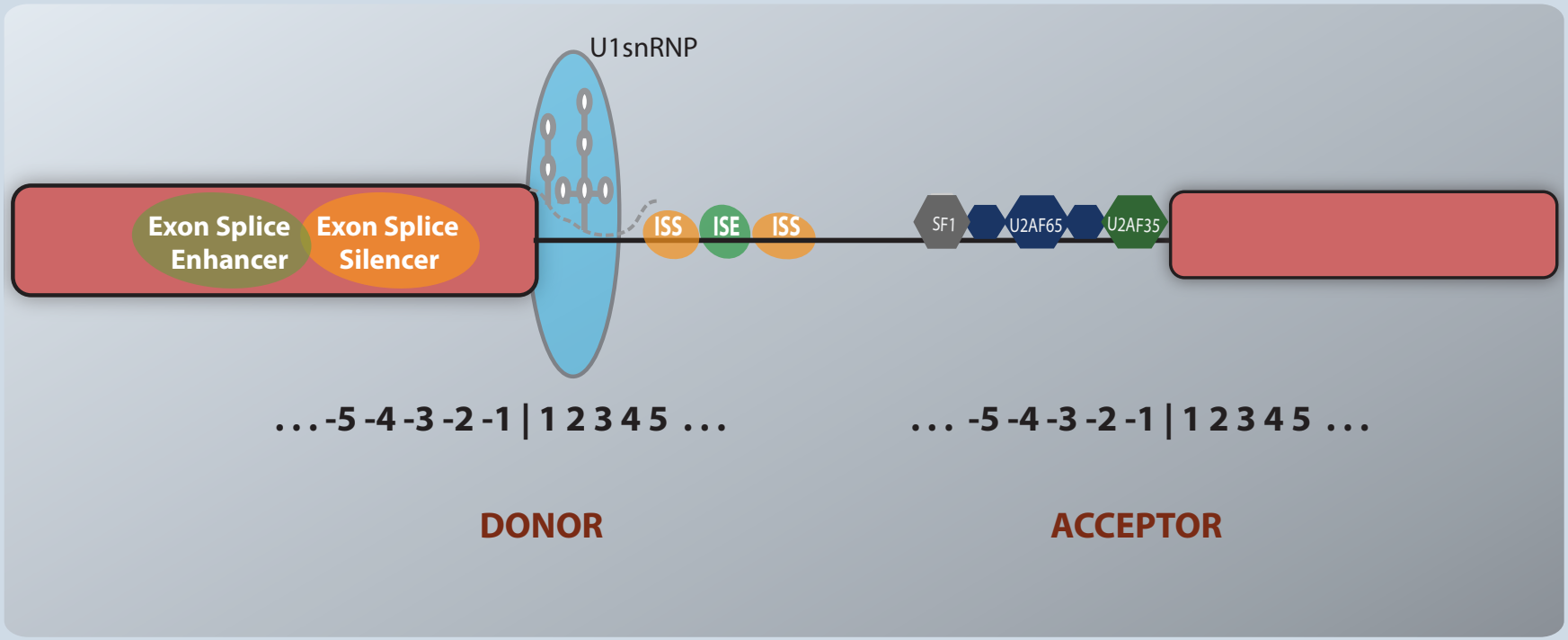
PAN-Transcriptome analysis of variants disrupting splicing

Manuel Rivas



September 27, 2012

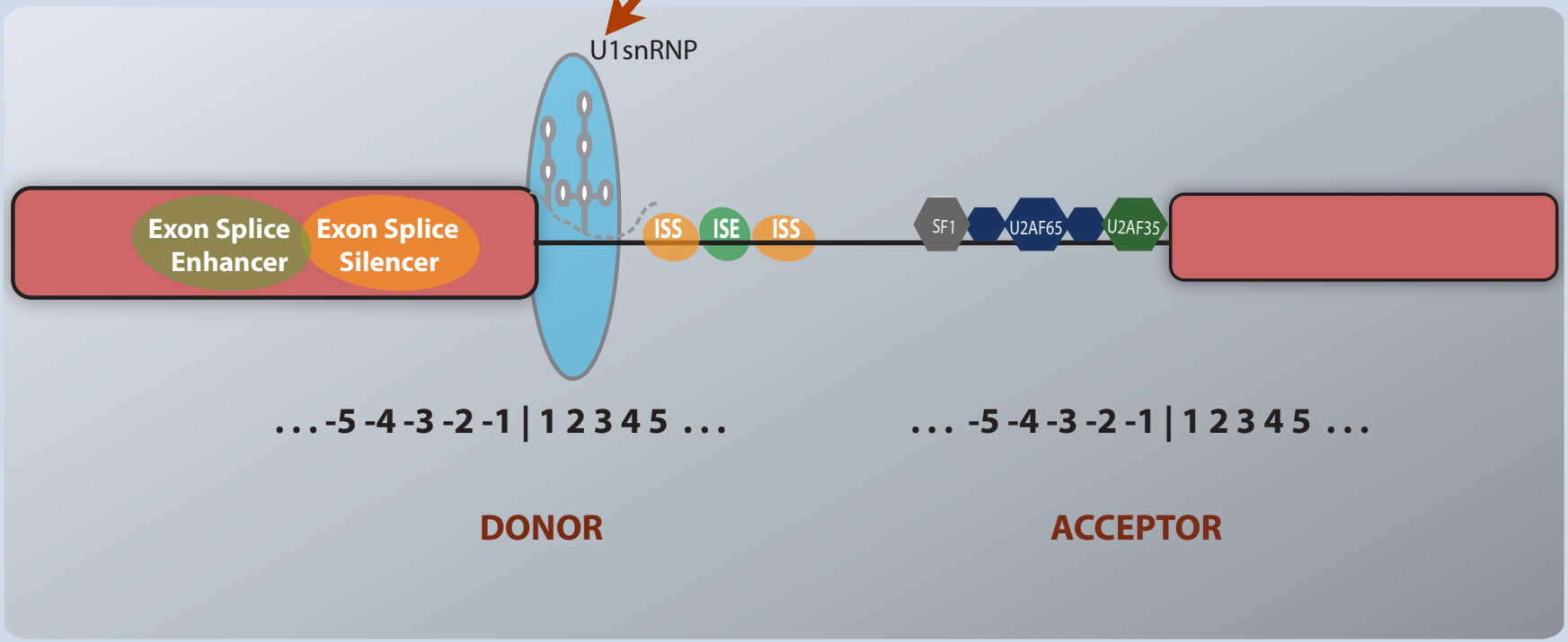
Motivation



We have a pretty decent idea of what is involved in splicing machinery.

Motivation

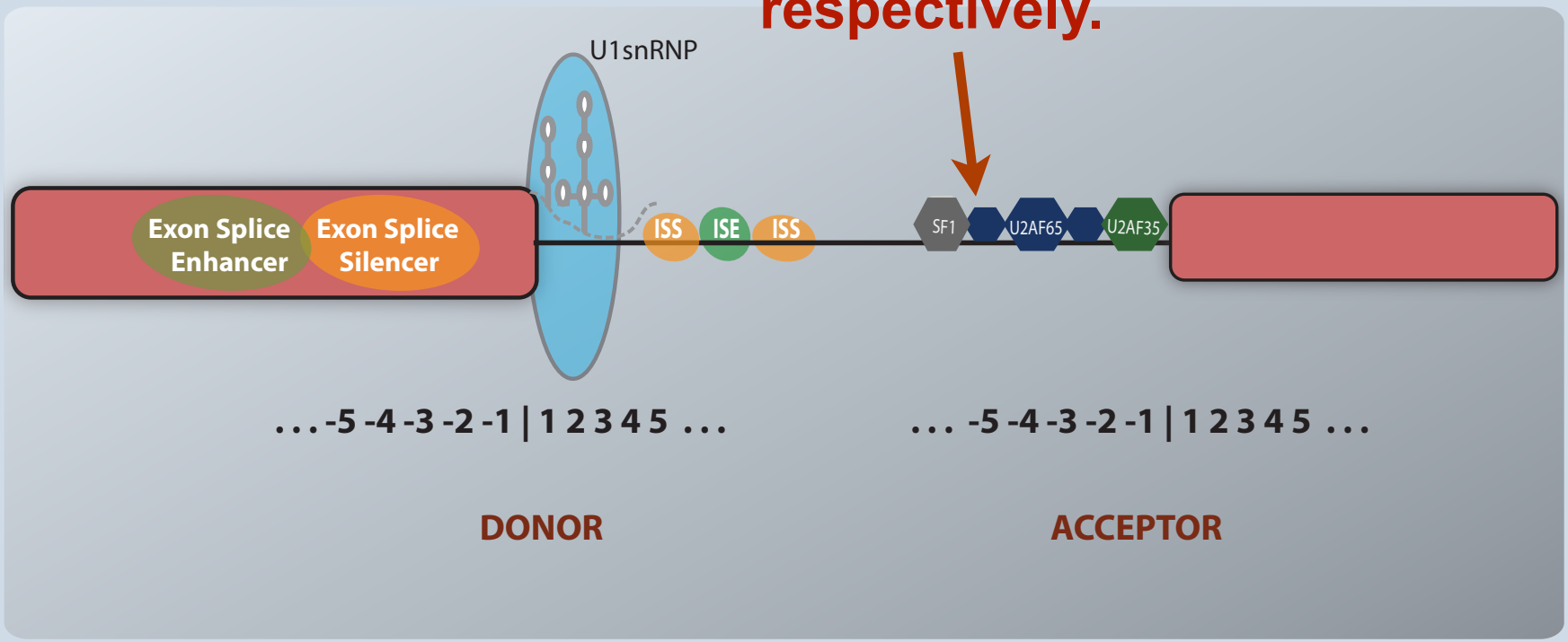
1. Recruitment of U1snRNP to the 5' Splice Site Junction.



Adapted from K Yoshida *et al*. *Nature* **000**, 1-6 (2011) doi:10.1038/nature10496

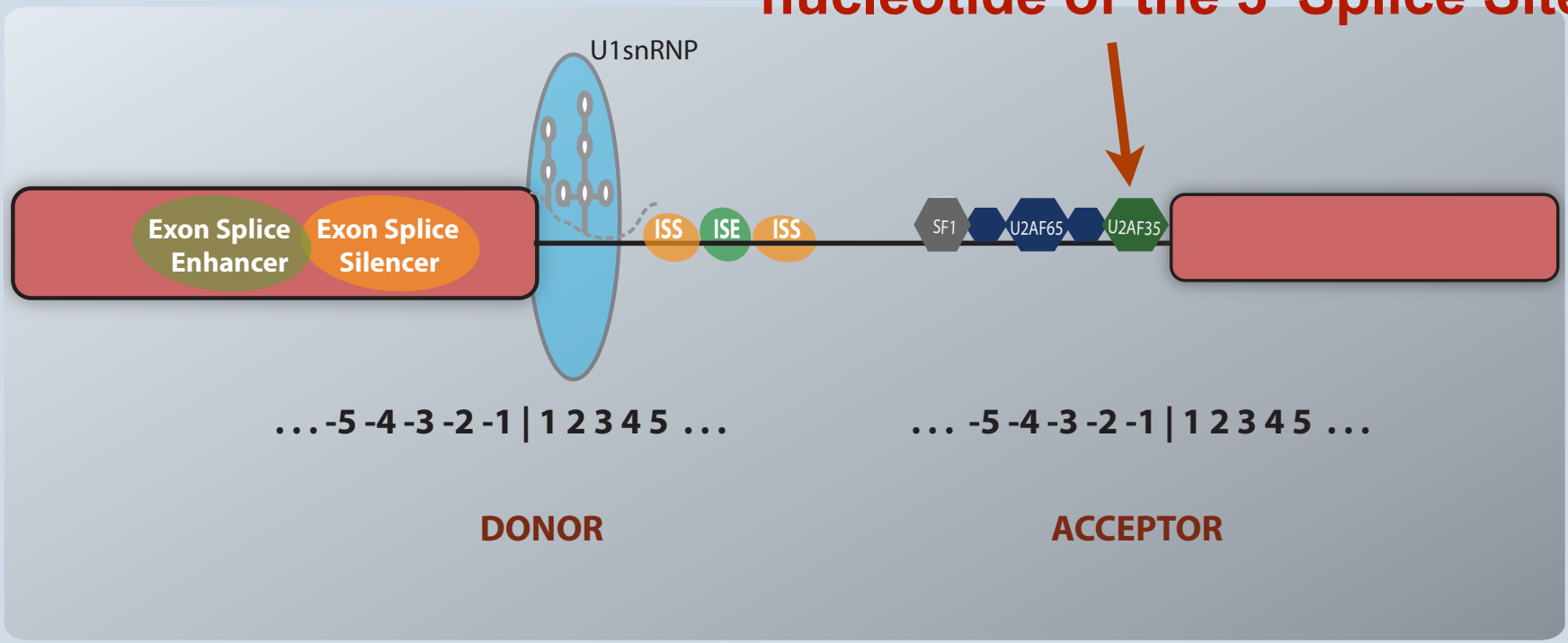
Motivation

2. SF1 and U2AF65 bind the branch sequence and polypyramidine tract respectively.



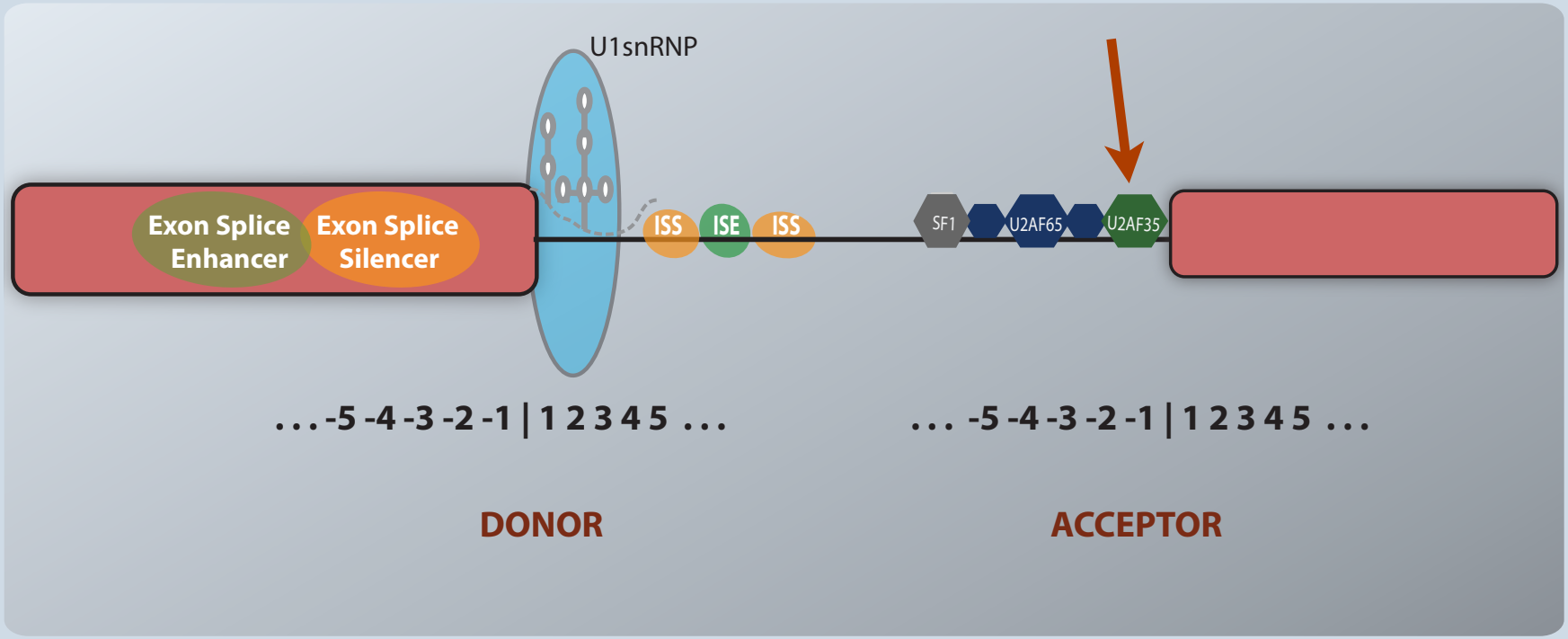
Motivation

3. U2AF35 binds to AG nucleotide of the 3' Splice Site.

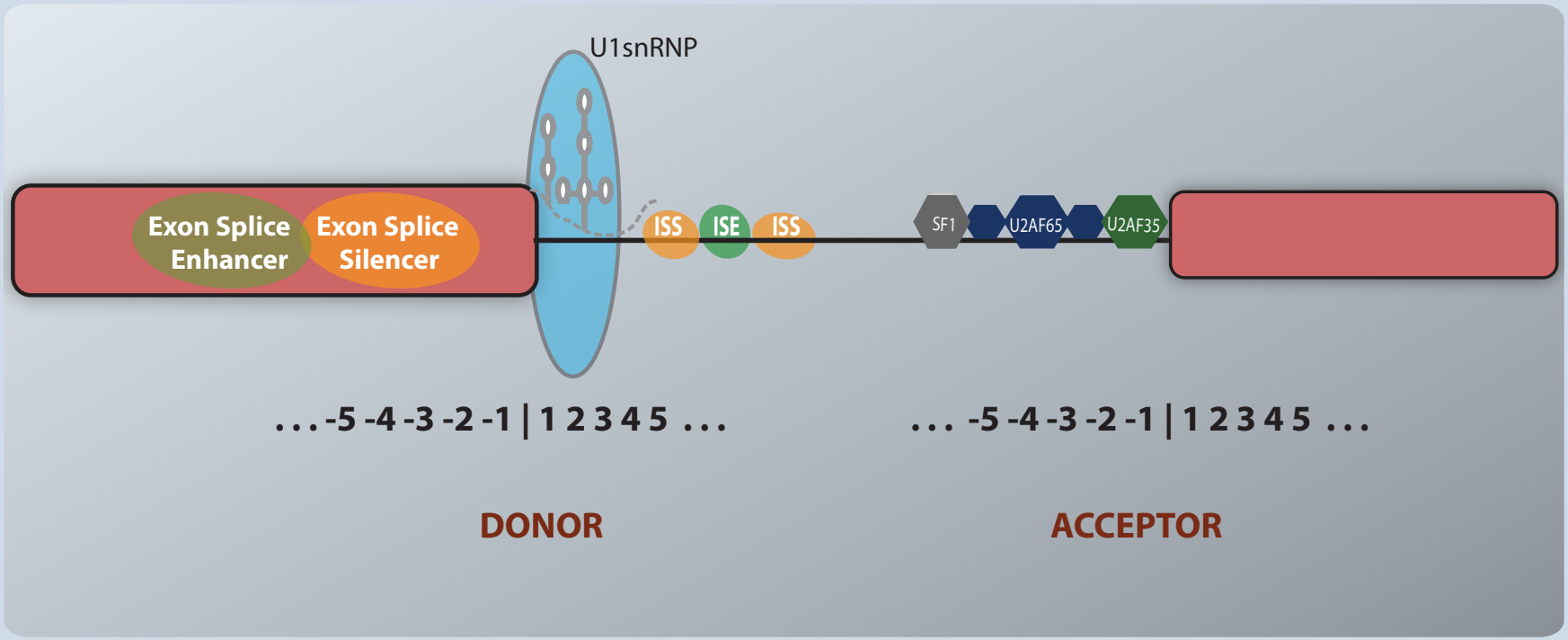


Motivation

4. Recruitment of U2snRNP together with SF3A1 and SF3B1 to generate splicing complex A.



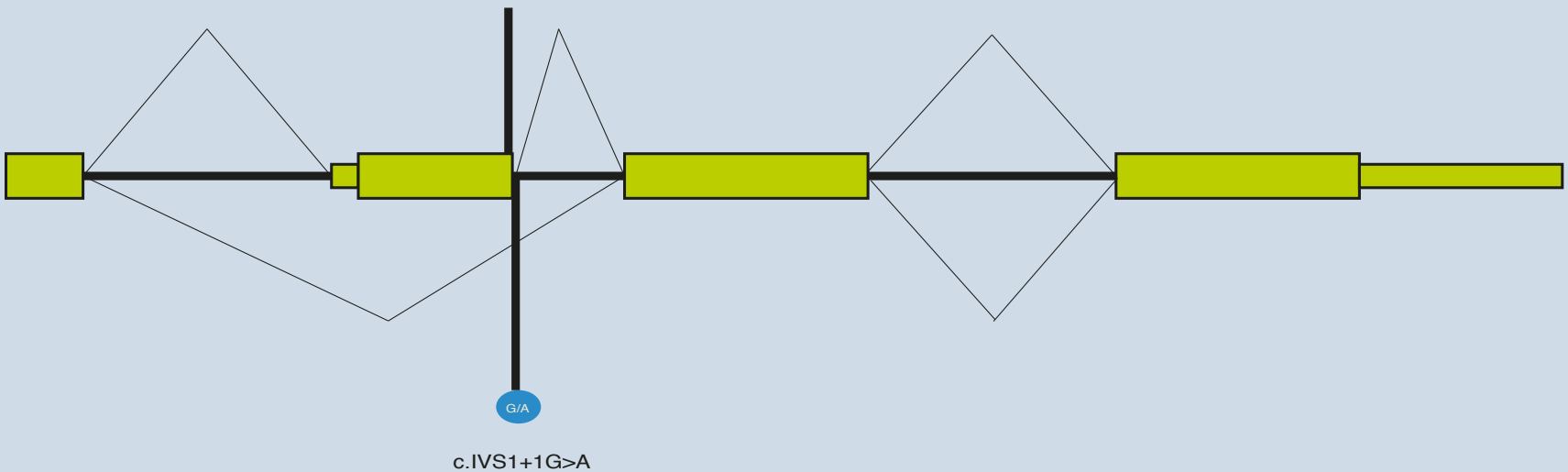
Motivation



How crucial disruption of a sequence near a splice junction is somewhat unknown.

Motivation

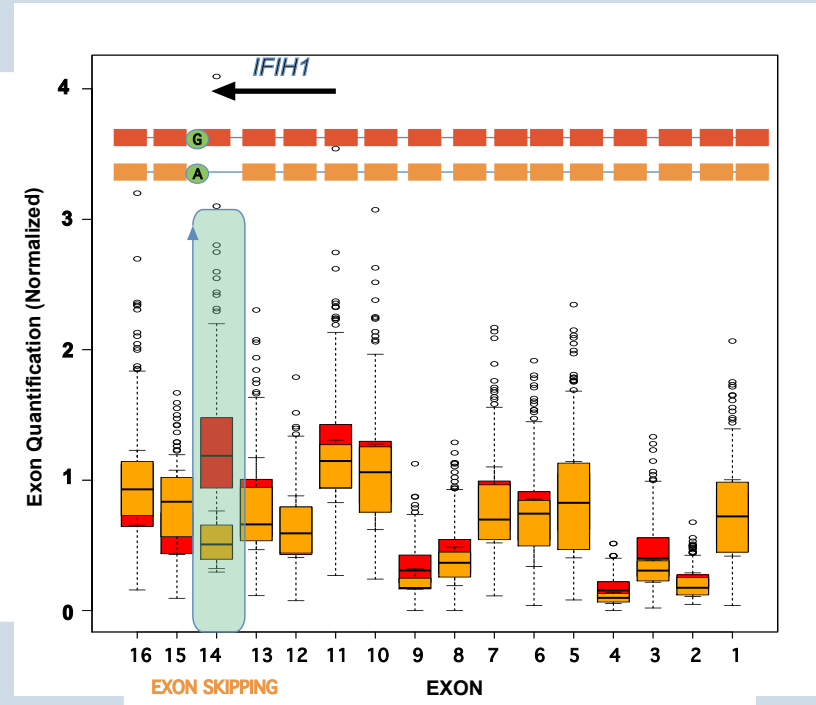
Example of splice disruption at the conserved/canonical 5' 'GT' site.



Motivation

Empirical example based on IFIH1 protective variant associated to Type I Diabetes.

RNASeq data implies that this variant would most likely have a dominant negative function as escapes nonsense medidated decay.



Motivation

- Current strategy is annotate variants as disrupting splicing if near splice junction and considered "LoF" variant
- Fortunately, we have empirical data to be able to evaluate how often these variants lead to disruption of variants
- Challenge is we do not have a complete catalog of all splice disrupting variants across all genes, so difficult to tell you whether or not your variant of interest leads to alternative splicing.
- However, we can tell you based on variants across the genome near splice junction what the proportion of times these variants impact splicing

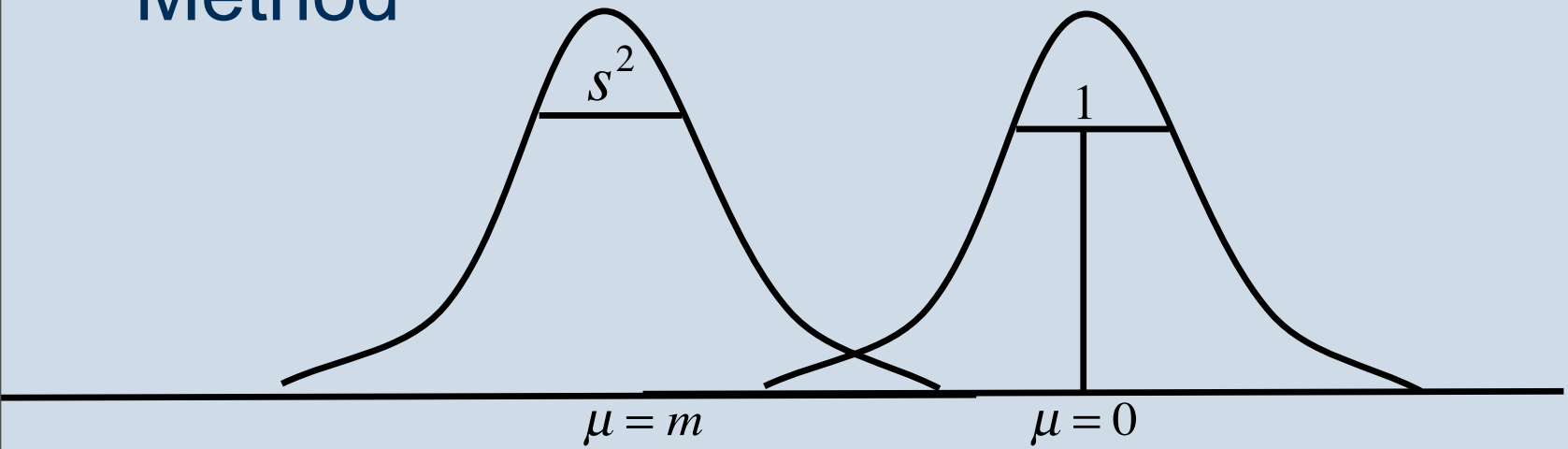
Pan-Transcriptome Analysis

Dataset Required

Normalized splice-junction quantification

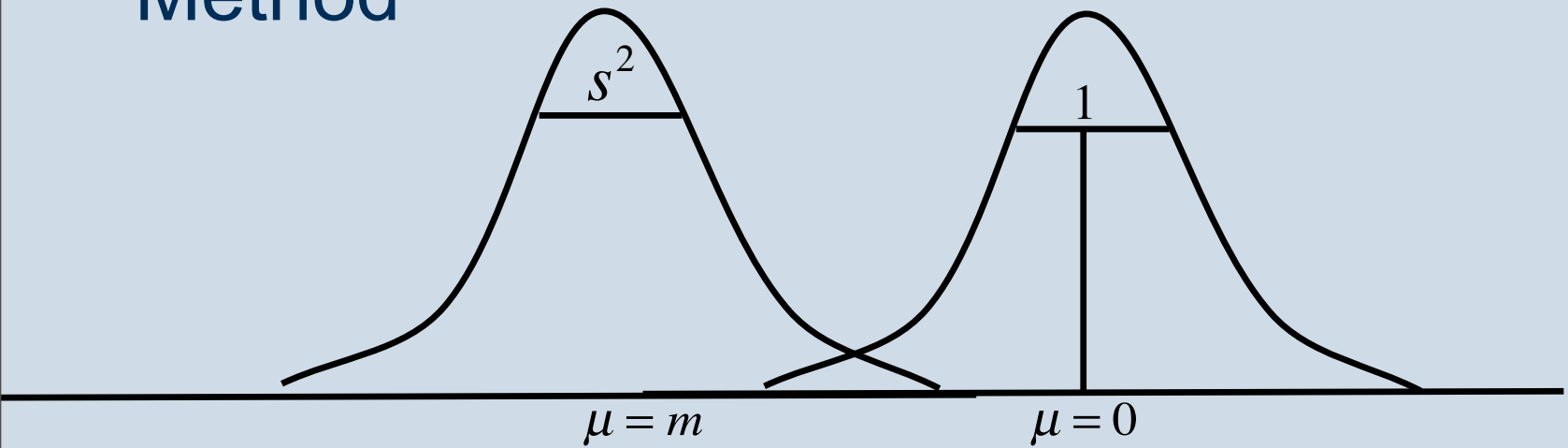
Variant Genotype data and relative distance of variant to splice junction

Method



Idea is that for splice disrupting variants the splice junction phenotype of the individual with the splice variant will either come from N(0,1) or from some other distribution N(m,s^2).

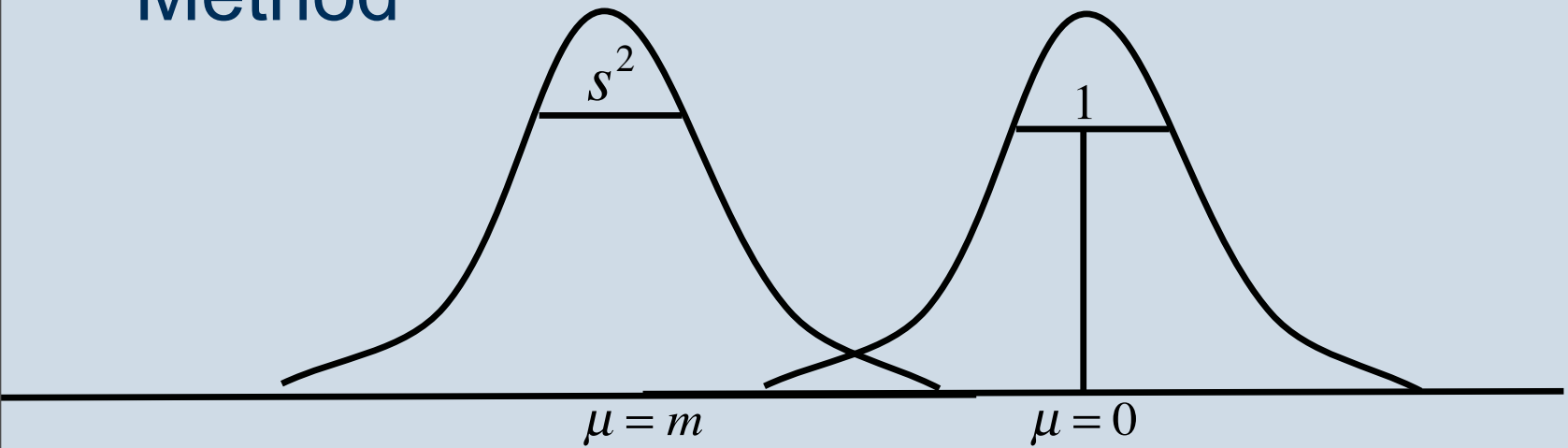
Method



Idea is that for splice disrupting variants the splice junction phenotype of the individual with the splice variant will either come from N(0,1) or from some other distribution N(m,s^2).

Challenge is that not all candidate splice disrupting variants will actually impact splicing

Method

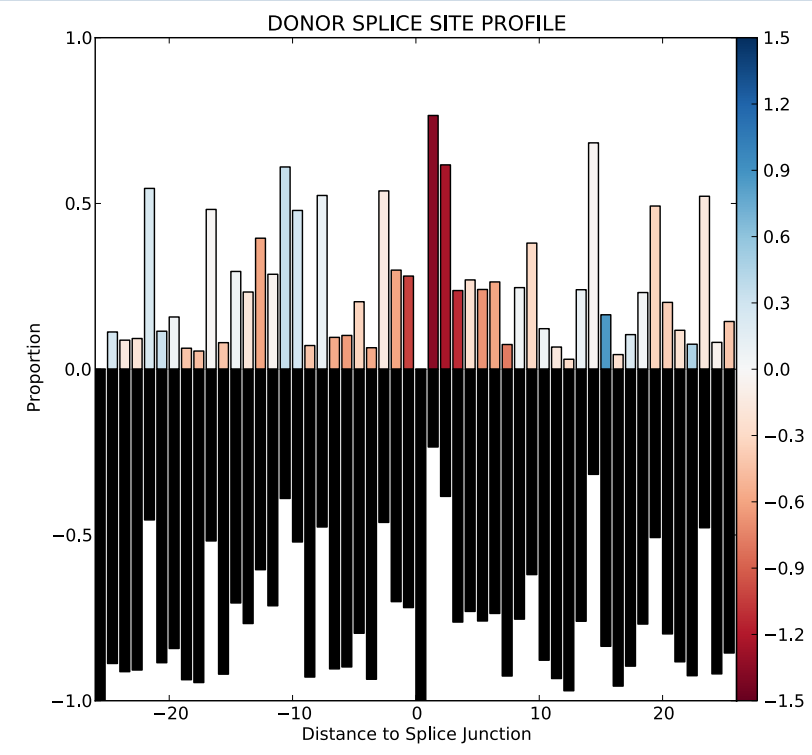


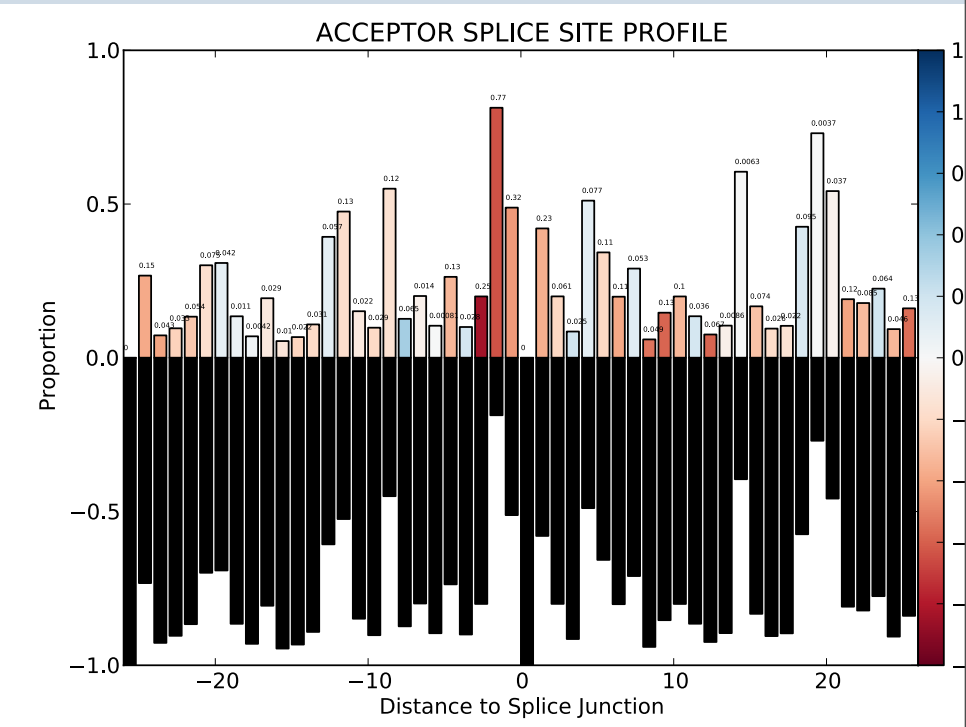
Idea is that for splice disrupting variants the splice junction phenotype of the individual with the splice variant will either come from N(0,1) or from some other distribution N(m,s^2).

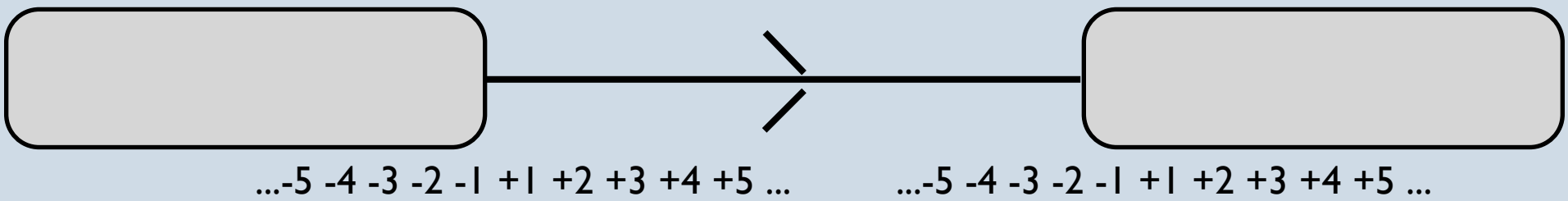
Challenge is that not all candidate splice disrupting variants will actually impact splicing

Propose to estimate the proportion of times these splice variant types come from N(0,1) or N(m,s^2) using **Gibbs Sampling**.

Application to Data



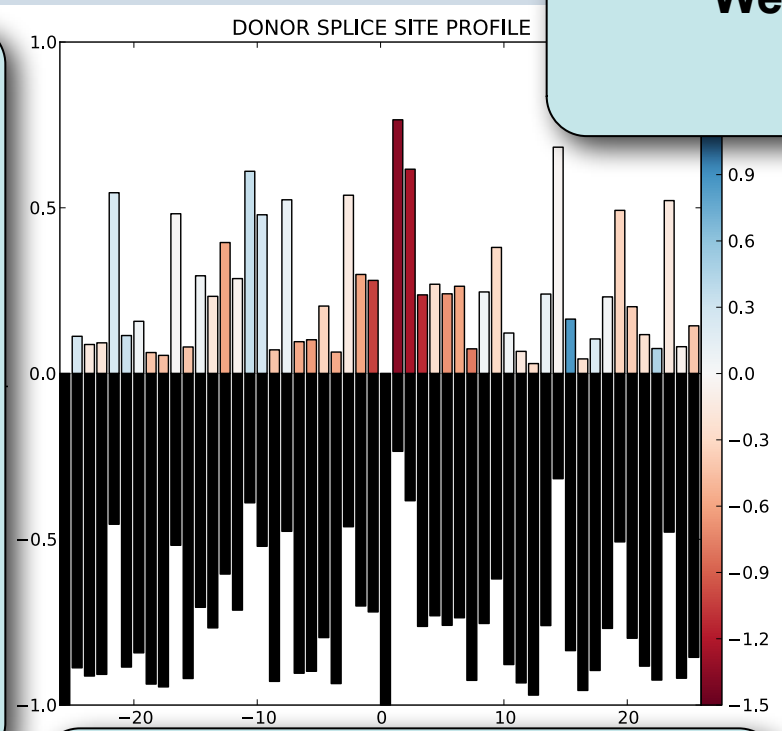


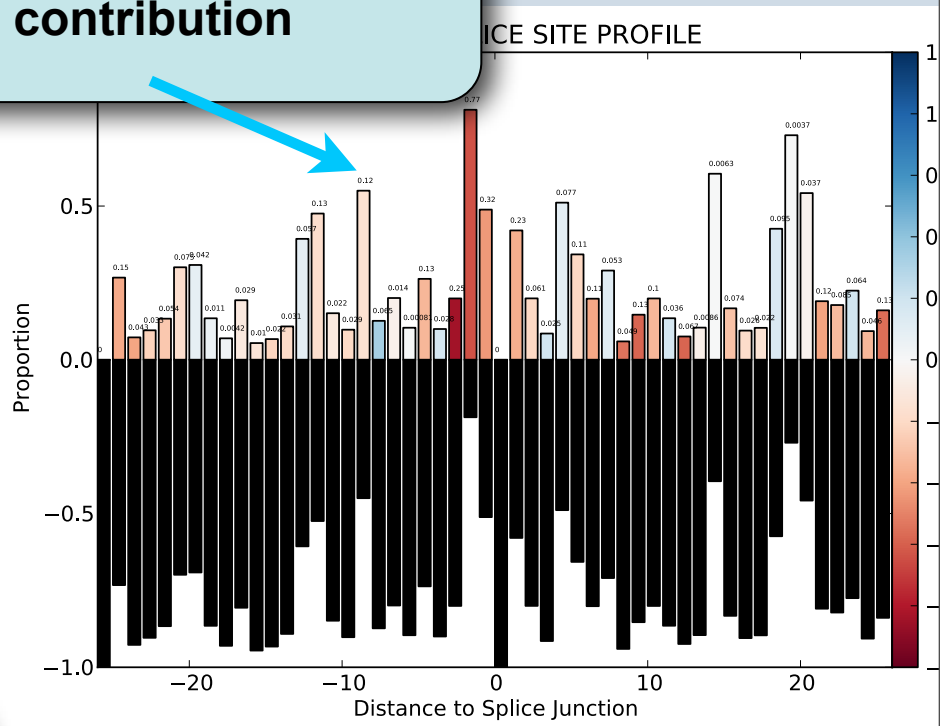


Application to Data

Weight of component contribution

Proportion N(01,) N(m,s^2)





Distance to Splice Junction

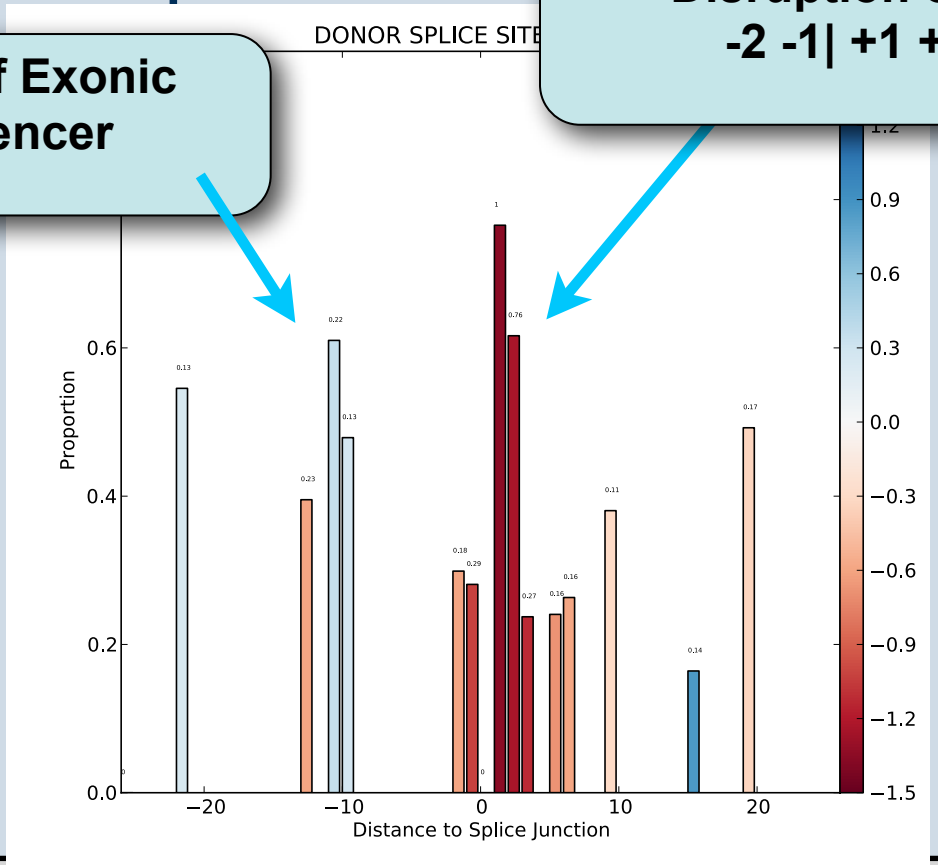
...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ...

...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ...

Significant components

Disruption of Exonic Splice Silencer

Disruption of Conserved -2 -1| +1 +2 +3 +4 +5



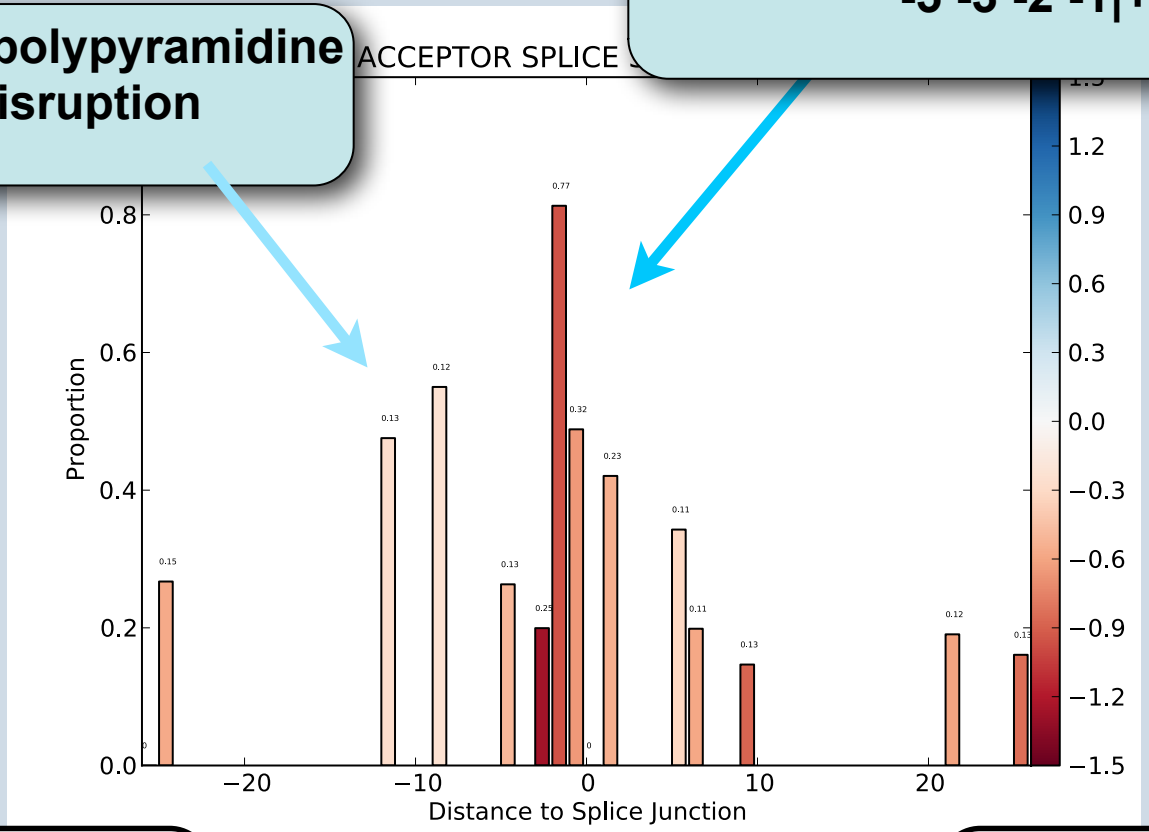
...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ...

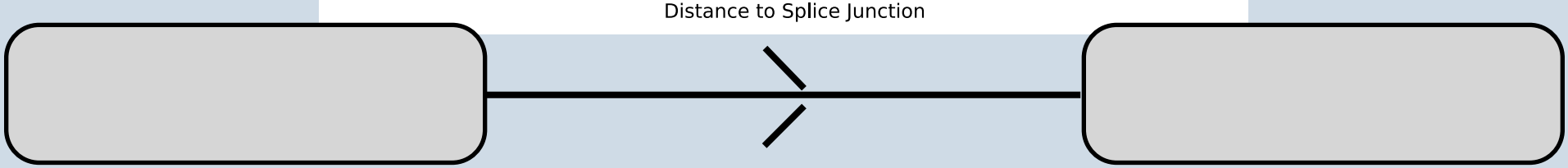
...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ...

Significant components

Disruption of Conserved -5 -3 -2 -1|+1

Evidence of polypyramidine tract disruption





...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ... ...-5 -4 -3 -2 -1 +1 +2 +3 +4 +5 ...

Ongoing Analysis

TL is updating splice junction quantifications normalized for overall gene expression levels (should give cleaner dataset).