

Statistics for JSplicePlot

07/05/2010

1 Definitions

1.1 Constitutive

- Constitutive Base- A base that is found among all isoforms.
- Constitutive Region- A continuous interval of constitutive bases.

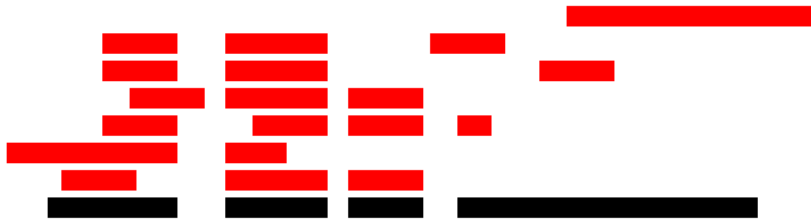
1.2 Compatible

- Compatible Read- A short read that is continuous in an exon

Compatible



Incompatible



2 Calculations

2.1 Short Reads Plot

2.2 RPKM of A Gene

$$\frac{\sum_{Interval\ A} (\# \text{ Body Reads in } A)}{\left(\frac{\sum_{Interval\ A} (A \text{ Length} - Read \text{ Length} + 1)}{1,000} \right)} \div \frac{Total \# \text{ Reads In } BAM}{1,000,000}$$

2.3 Height of Exon

Note- the Isoform heights are calculated using only compatible reads.

2.3.1 Method 1 - Average Coverage Per Exon

$$\frac{\sum_{BaseA} Coverage\ of\ A}{L}$$

L = Exon Length (Also equivalent to number of bases)

Coverage Of a base includes both junction and body reads

2.3.2 Method 2 - RPK

$$\frac{B}{\left(\frac{L}{1000}\right)}$$

B = Short Reads that Cross a Exon (Body Reads Only)

L = Exon Length (Also equivalent to number of bases)

2.3.3 Method 3 - RPKM

Number of Short Reads Over the Total Number of Possible Start Positions

Denominator Represents the Total Number of Possible Start Positions

- First or Last Exon

$$\frac{S}{\left(\frac{[L - 2(V - 1)]}{1000}\right)} \div \frac{Total\ \# \ Reads\ In\ BAM}{1,000,000}$$

- Inclusive Exon

$$\frac{S}{\left(\frac{[L - 2(V - 1)] + [(R - 1) - 2(V - 1)]}{1000}\right)} \div \frac{Total\ \# \ Reads\ In\ BAM}{1,000,000}$$

S = # Short Reads that Cross a Exon (Body and Junction)

R = Read Length

V = Overhang Length

L = Exon Length (Also equivalent to number of bases)

- Junction

$$\frac{J}{\left(\frac{(R - 1) - 2(V - 1)}{1000}\right)} \div \frac{Total\ \# \ Reads\ In\ BAM}{1,000,000}$$

J = Number of Junction Reads

R = Read Length

V = Overhang Length