Splice Graphs

Martin Morgan

2012-07-06 Fri

Contents

Work of Daniel Bindreither, Marc Carlson, Martin Morgan

1 Background: Gene models to splice graphs

Sammeth, 2010, Computation Biology 16: 1117-1140
Starting point (for us): known gene model of exons-within-transcripts-within-genes
Verticies represent gene start, exon donor / acceptor, gene ends, counting bin boundaries
Edges represent paths between verticies

- zero or more exons
- sometimes shared by two or more transcripts

Splice codes

- Summarize order of verticies
- Familiar (‘exon skip’; ‘alternative promoter’; etc) and exotic

Bubbles

- Represent alternative splicing events
- simple
- compound – three or more transcripts, three-way and higher splicing

2 Benefits

Edges contain >= 1 counting bins, so...
Classification of splice events
Genome-wide assessment of splicing events
Testing bubbles (alternative splicing), in addition to edges

3 Example (partial)

3.1 Implementation

SpliceGraph package, in development; also `spliceGraph` function in GenomicFeatures.
Transform exons-within-transcripts-within-genes GRangesList to edges-within-transcripts-within-genes.

- includes DESeq 'counting bins' operation (disjoin)

```
library("SpliceGraph")
library(TxDb.Dmelanogaster.UCSC.dm3.ensGene)
txdb <- TxDb.Dmelanogaster.UCSC.dm3.ensGene
spliced <- spliceGraphs(txdb)
save(spliced, file="2012-07-06-pasilla-spliceGraph.rda")
```

Count via `countSpliceGraphReads`
Normalization, etc: like DEXSeq
Statistical assessment

- edges: analogous to / using DEXSeq
- ??? bubbles ???

  ~ sample + edge + condition * $\sum_{bp}$ I(edge == bp_i)

3.2 Bubbles

event	Nr
retain 1 intron	1302
alternative promotor usage	892
alternative transcript end	824
2 alternative donors	734
skip 1 exon	695
2 alternative acceptors	658
	...
Total	11913

- FBgn0011224 annotated with 42 exons and 22 transcripts, has 63 bubbles

3.3 Edges vs. exons

90249 non-zero count exons, median 25 reads / exon 22981 edges, 32 reads / edge
Differential expression: stay tuned. . .

```
load("cD.exons-fb.Rda")
load("cD.sG-fb.Rda")
tbl <- read.table("phenoData.txt", header=TRUE)

exonCounts <- cD.ex[rowMeans(cD.ex) != 0,]
edgeCounts <- cD.sG[rowMeans(cD.sG) != 0,]

nrow(exonCounts)
mean(exonCounts)

nrow(edgeCounts)
mean(edgeCounts)
```

3.4 Differential representation (to come. . .)