PathwaySplice testing for RNA-seq datasets

Aimin Yan

2017-08-07

1 Introduction

This document gives an introduction to the use of the PathwaySplice R Bioconductor package. This package provides methods for (1) performing pathway analysis that ex-plicitly adjusts for the number of exons or junctions associated with each gene; (2) visualizing selection bias due to different number of exons or junctions for each gene and formally tests for presence of bias using logistic regression; (3) supporting gene sets based on the Gene Ontology terms, as well as more broadly defined gene sets (e.g. MSigDB) or user defined gene sets; (4) identifing the significant genes driving pathway significance and (5) organizing significant pathways with an enrichment map, where pathways with large number of overlapping genes are grouped together in a network graph.

```
Once installed, the PathwaySplice package can be easily loaded into R using:

    library(PathwaySplcie)
```

2 Use PathwaySplice

PathwaySplice uses the output from DEXSeq(Anders, Reyes, and Huber 2012) or JunctionSeq(Hartley and Mullikin 2016) as input basis. Assume users have gotten the differnetial exon usage or differential splicing junctions usage results from DEXSeq or JunctionSeq. User needs to convert these resutls into gene based table, then use this gene based table to perform analysis using the following commands:

```
featureBasedData <- PathwaySplice:::makeFeatureTable(res)
gene.based.table <- makeGeneTable(featureBasedDat)
```

For a showcase, PathwaySplice supply a feature based data with this packages, users can load this data using commands:

```
data("featureBasedData")
gene.based.table <- makeGeneTable(featureBasedData)
```
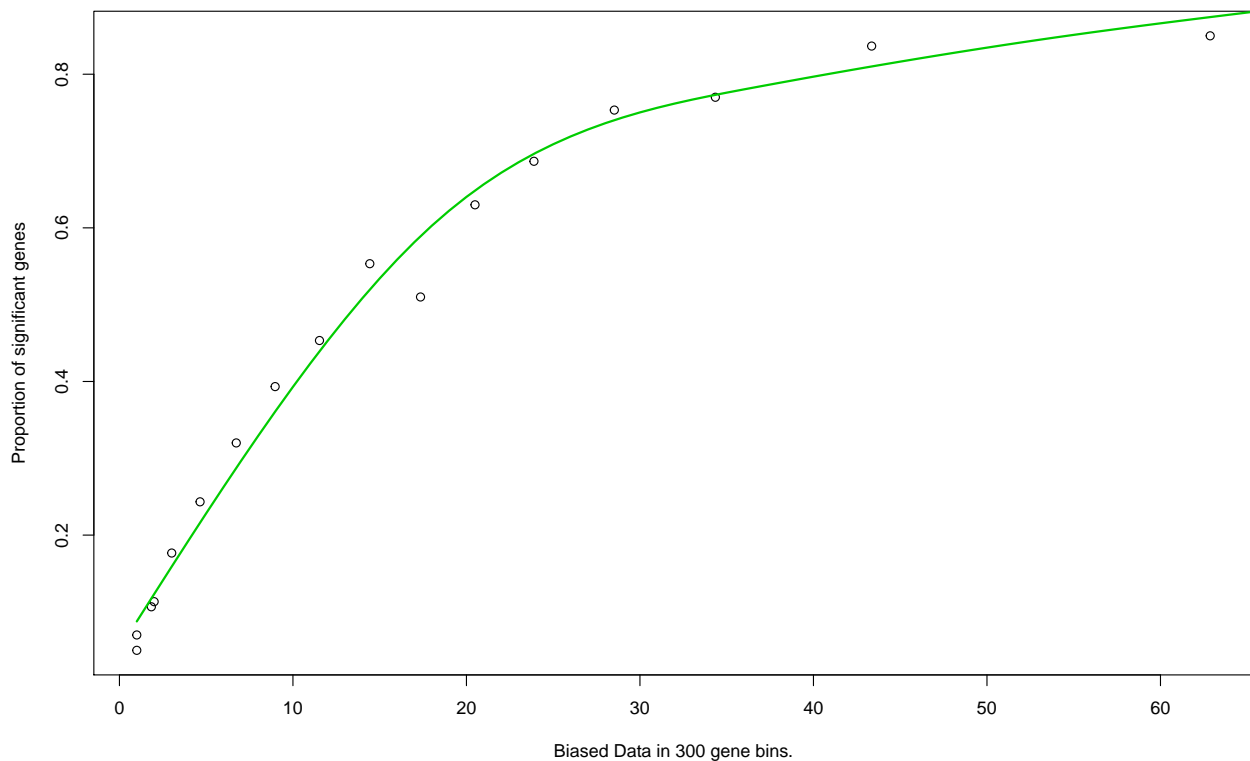
After this step, use can identify bias factor by the function lrTestBias

```
lrTestBias(gene.based.table,boxplot.width=0.3)
```

P–value from logistic regression <1e–04
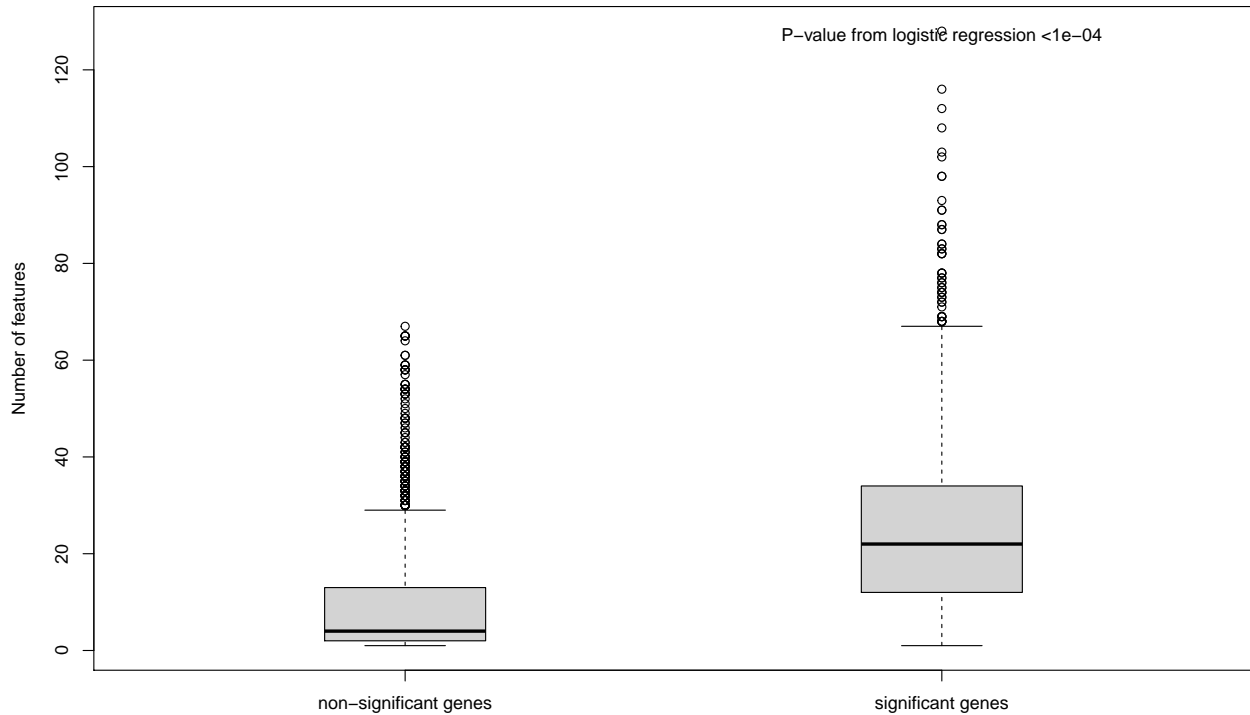
To perform analysis by adjusting number of feature, users can perform the following step:

```
res <- runPathwaySplice(gene.based.table,genome='hg19',id='ensGene',
                        test.cats=c('GO:BP'),
                        go.size.limit=c(5,30),method='Wallenius')
```


User can use the resuls from runPathwaySplice to build up enrichment network using the following commands:

```
output.file.dir <- file.path(tempdir(),"OutputEnmap")
enmap <- enrichmentMap(res,n=3,output.file.dir
                       =output.file.dir,similarity.threshold=0)
```
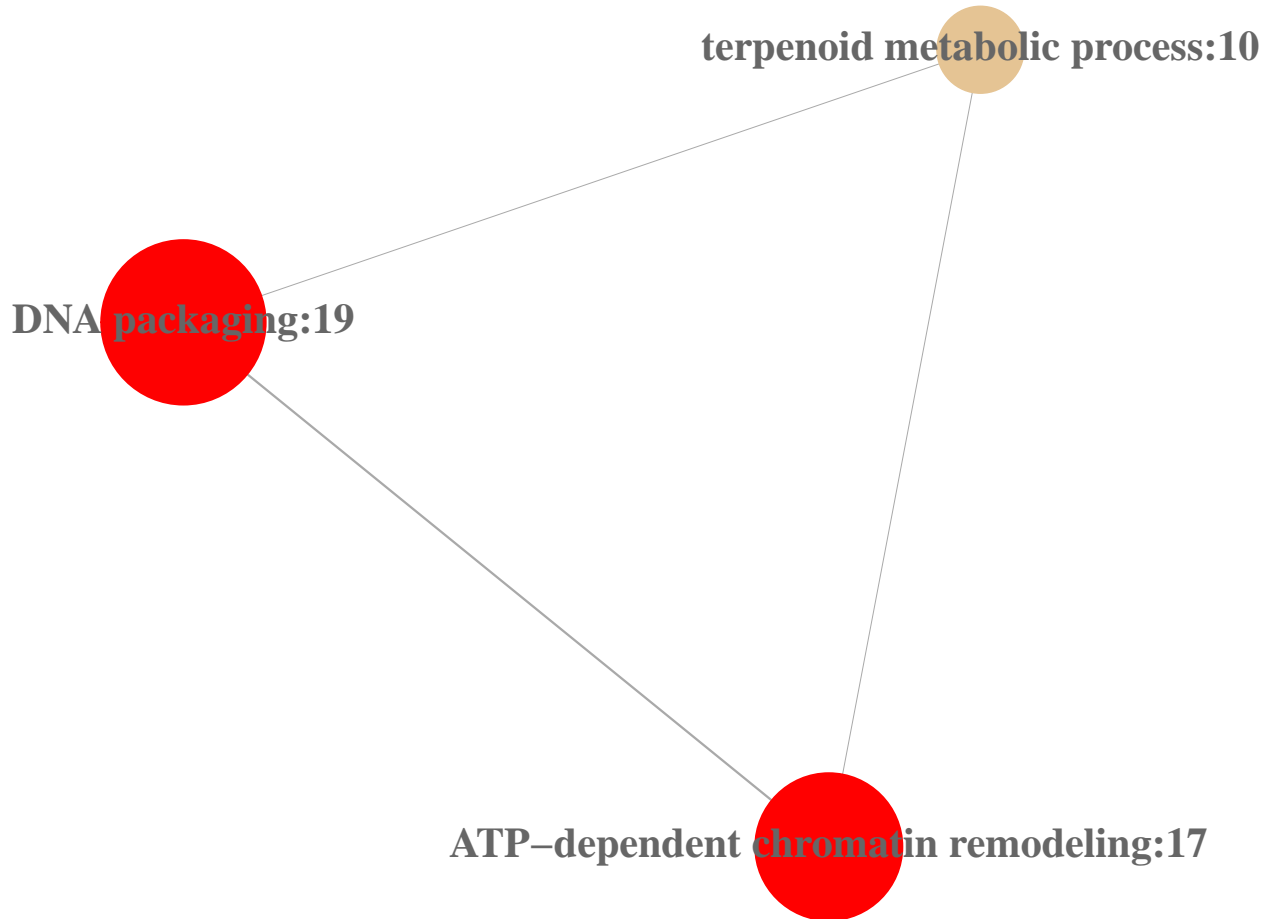

2.1 Conclusion

In summary, our work identified the potential bias in RNA-Seq data analysis when using differential exons or/and splicing junction usage to represent differential expressed gene, and using these differential expressed genes to perform subsequent gene set and pathway analysis. We futher sugguest a bias correction approach that could provide a more accutate gene set and pathway analsis. We implemented this method into an R package, and we believe that this R packcage could help other researchers when they perform similar type of RNA-Seq data analysis.

Reference

Anders, Simon, Alejandro Reyes, and Wolfgang Huber. 2012. "Detecting differential usage of exons from RNA-seq data." *Genome Research* 22 (10): 2008–17. doi:10.1101/gr.133744.111.

Hartley, Stephen W., and James C. Mullikin. 2016. "Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq." *Nucleic Acids Research*, June. Oxford University Press, gkw501. doi:10.1093/nar/gkw501.