Using Reporting Tools in an Analysis of RNA-seq Data

Jessica L. Larson and Christina Chaivorapol

May 1, 2024

Contents

1 Introduction
2 Differential expression analysis with edgeR
3 Differential expression analysis with DESeq2
4 GO analysis using GOstats
5 PFAM analysis
6 Putting it all together
7 References
1 Introduction

The ReportingTools package can be used with differential gene expression results from RNA-sequencing analysis. In this vignette we show how to publish output from an edgeR, Gene Ontology (GO) and/or Protein family (PFAM) analysis. In the final section we publish all our pages onto one, creating a comprehensive output page.

2 Differential expression analysis with edgeR

In this section we demonstrate how to use the ReportingTools package to generate a table of differentially expressed genes as determined by the edgeR software. We begin by loading our library and data set. The mockRnaSeqData contains random RNA-seq output for random mouse genes.

```r
> library(ReportingTools)
> data(mockRnaSeqData)

Next, we run edgeR to find differentially expressed genes.

```r
> library(edgeR)
> conditions <- c(rep("case",3), rep("control", 3))
> d <- DGEList(counts = mockRnaSeqData, group = conditions)
> d <- calcNormFactors(d)
> d <- estimateCommonDisp(d)
> ## Get an edgeR object
> edgeR.de <- exactTest(d)
```

Now the results can be written to a report using the DGEExact object.

```r
> library(lattice)
> rep.theme <- reporting.theme()
> ## Change symbol colors in plots
> rep.theme$superpose.symbol$col <- c("blue", "red")
> rep.theme$superpose.symbol$fill <- c("blue", "red")
> lattice.options(default.theme = rep.theme)
> ## Publish a report of the top 10 genes with p-values < 0.05 and log-fold change > 2
> ## In this case, the plots contain the counts from mockRnaSeqData, which are not normalized.
> ## The publish function does not normalize counts for the countTable argument to allow for
> ## flexibility in plotting various units (e.g. RPKM instead of counts).
> deReport <- HTMLReport(shortName = 'RNAseq_analysis_with_edgeR',
+ title = 'RNA-seq analysis of differential expression using edgeR',
+ reportDirectory = './reports')
> publish(edgeR.de, deReport, countTable=mockRnaSeqData,
+ conditions=conditions, annotation.db = 'org.Mm.eg',
+ pvalueCutoff = .05, lfc = 2, n = 10, name="edgeR")
> finish(deReport)
```

If you would like to plot normalized counts, run the following commands instead:

```r
## mockRnaSeqData.norm <- d$pseudo.counts
## publish(edgeR.de, deReport, mockRnaSeqData.norm,
## conditions, annotation.db = 'org.Mm.eg',
## pvalueCutoff = .05, lfc = 2, n = 10)
## finish(deReport)
```
We can also output results of the LRT test from edgeR.

\[
\begin{align*}
&\text{d <- DGEList(counts = mockRnaSeqData, group = conditions)} \\
&\text{d <- calcNormFactors(d)} \\
&\text{design <- model.matrix(~conditions)} \\
&\text{d <- estimateGLMCommonDisp(d, design)} \\
&\text{d <- estimateGLMTrendedDisp(d, design)} \\
&\text{d <- estimateGLMTagwiseDisp(d, design)} \\
&\text{fit <- glmFit(d, design)} \\
&\text{edgeR.lrt <- glmLRT(fit, coef=2)} \\
&\text{deReport2 <- HTMLReport(shortName = 'RNAseq_analysis_with_edgeR_2',}
&\text{ title = 'RNA-seq analysis of differential expression using edgeR (LRT)'}, \\
&\text{ reportDirectory = "./reports")} \\
&\text{publish(edgeR.lrt, deReport2, countTable=mockRnaSeqData,}
&\text{ conditions=conditions, annotation.db = 'org.Mm.eg',}
&\text{ pvalueCutoff = .05, lfc = 2, n = 10, name="edgeRlrt")} \\
&\text{finish(deReport2)}
\end{align*}
\]

3 Differential expression analysis with DESeq2

In this section we demonstrate how to use the ReportingTools package to generate a table of differentially expressed genes as determined by the DESeq2 packages.

We start by running DESeq2 to find differentially expressed genes.

\[
\begin{align*}
&\text{library(DESeq2)} \\
&\text{conditions <- c(rep("case",3), rep("control", 3))} \\
&\text{mockRna.dse <- DESeqDataSetFromMatrix(countData = mockRnaSeqData,}
&\text{ colData = as.data.frame(conditions), design = ~ conditions)} \\
&\text{colData(mockRna.dse)$conditions <- factor(colData(mockRna.dse)$conditions, levels=c("control", "case"))} \\
&\text{## Get a DESeqDataSet object} \\
&\text{mockRna.dse <- DESeq(mockRna.dse)}
\end{align*}
\]

Now the results can be written to a report using the DESeqDataSet object.
**RNA-seq analysis of differential expression using DESeq2**

<table>
<thead>
<tr>
<th>EntrezID</th>
<th>Symbol</th>
<th>GeneName</th>
<th>Image</th>
<th>logFC</th>
<th>p-Value</th>
<th>Adjusted p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>665972</td>
<td>Um7871</td>
<td>predicted gene 7871</td>
<td>-</td>
<td>5.77</td>
<td>3.15e-22</td>
<td>5.77e-18</td>
</tr>
<tr>
<td>111941</td>
<td>lap5rc10</td>
<td>intracisternal A-type particles, IS region, SINE repeat 0-10</td>
<td>-</td>
<td>-4.96</td>
<td>8.26e-16</td>
<td>7.58e-12</td>
</tr>
<tr>
<td>100040696</td>
<td>Gm2912</td>
<td>predicted gene 2912</td>
<td>-</td>
<td>3.31</td>
<td>8.13e-12</td>
<td>4.97e-08</td>
</tr>
<tr>
<td>109594</td>
<td>Lmo1</td>
<td>LIM domain only 1</td>
<td>-</td>
<td>-4.41</td>
<td>1.47e-11</td>
<td>6.74e-08</td>
</tr>
<tr>
<td>85078</td>
<td>Dnm114</td>
<td>DNA segment, Chr 6, Massachusetts Institute of Technology 14</td>
<td>-</td>
<td>4.14</td>
<td>5.50e-10</td>
<td>2.02e-06</td>
</tr>
<tr>
<td>100035505</td>
<td>Kdrl</td>
<td>kidney weight 1</td>
<td>-</td>
<td>-4.32</td>
<td>7.59e-10</td>
<td>2.32e-06</td>
</tr>
</tbody>
</table>

Figure 2: Resulting page created with DESeqDataSet object from DESeq2 analysis

```r
> des2Report <- HTMLReport(shortName = 'RNAseq_analysis_with_DESeq2',
+ title = 'RNA-seq analysis of differential expression using DESeq2',
+ reportDirectory = './reports')
> publish(mockRna.dse,des2Report, pvalueCutoff=0.05,
+ annotation.db="org.Mm.eg.db", factor = colData(mockRna.dse)$conditions,
+ reportDir='./reports')
> finish(des2Report)
```

### 4 GO analysis using GOstats

This section will demonstrate how to use ReportingTools to write a table of GO analysis results to an html file. First, we select our genes of interest, and then run the hyperGTest.

```r
> library(GOstats)
> library(org.Mm.eg.db)
> tt <- topTags(edgeR.de, n = 1000, adjust.method = 'BH', sort.by = 'p.value')
> selectedIDs <- rownames(tt$table)
> universeIDs <- rownames(mockRnaSeqData)
> goParams <- new("GOHyperGParams",
+ geneIds = selectedIDs,
+ universeGeneIds = universeIDs,
+ annotation ="org.Mm.eg",
+ ontology = "MF",
+ pvalueCutoff = 0.01,
+ conditional = TRUE,
+)```
goResults <- hyperGTest(goParams)

With these results, we can then make the GO report.

> goReport <- HTMLReport(shortName = 'go_analysis_rnaseq',
+   title = "GO analysis of mockRnaSeqData",
+   reportDirectory = "./reports")
> publish(goResults, goReport, selectedIDs=selectedIDs, annotation.db="org.Mm.eg",
+   pvalueCutoff= 0.05)
> finish(goReport)

## 5 PFAM analysis

In this section, we show how to use ReportingTools to write a table of PFAM analysis results to an html file.

First, we run the hyperGTest using our genes of interest from the previous section.

> library(Category)
> params <- new("PFAMHyperGParams",
+   geneIds= selectedIDs,
+   universeGeneIds=universeIDs,
+   annotation="org.Mm.eg",
+   pvalueCutoff= 0.01,
+   testDirection="over")
> PFAMResults <- hyperGTest(params)

Then we make the PFAM report.

> PFAMReport <- HTMLReport(shortName = 'pfam_analysis_rnaseq',
+   title = "PFAM analysis of mockRnaSeqData",
+   reportDirectory = "./reports")
> publish(PFAMResults, PFAMReport, selectedIDs=selectedIDs, annotation.db="org.Mm.eg",categorySize=5)
> finish(PFAMReport)

## 6 Putting it all together

Here, we make an index page that puts all three analyses together for easy navigation.

> indexPage <- HTMLReport(shortName = "indexRNASeq",
+   title = "Analysis of mockRnaSeqData",
+   reportDirectory = "./reports")
> publish(Link(list(deReport,des2Report, goReport, PFAMReport), report = indexPage),
+   indexPage)
> finish(indexPage)

## 7 References

**PFAM analysis of mockRnaSeqData**

<table>
<thead>
<tr>
<th>PFAM ID</th>
<th>PFAM Term</th>
<th>PFAM Size</th>
<th>Image</th>
<th>Overlap</th>
<th>Odds Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF00413</td>
<td>Maltolix</td>
<td>8</td>
<td></td>
<td>4</td>
<td>15.40</td>
<td>0.000653</td>
</tr>
<tr>
<td>PF00057</td>
<td>Low-density lipoprotein receptor domain A</td>
<td>15</td>
<td></td>
<td>5</td>
<td>8.21</td>
<td>0.001190</td>
</tr>
</tbody>
</table>

Figure 3: Resulting page created by `publish` for PFAMResults

**Analysis of mockRnaSeqData**

RNA-seq analysis of differential expression using edgeR  
RNA-seq analysis of differential expression using DESeq2  
GO analysis of mockRnaSeqData  
PFAM analysis of mockRnaSeqData

Figure 4: Resulting page created by calling `publish` on all our analysis pages