

Reactome Pathway Analysis

Guangchuang Yu

July 10, 2013

Contents

1	Introduction	1
2	Pathway Enrichment Analysis	1
2.1	Visualize enrichment result	2
2.2	Comparing enriched reactome pathways among gene clusters with clusterProfiler	3
3	Gene Set Enrichment Analysis	3
3.1	Visualize GSEA result	5
4	Pathway Visualization	5
5	Session Information	6

1 Introduction

This package is designed for reactome pathway-based analysis. Reactome is an open-source, open access, manually curated and peer-reviewed pathway database.

2 Pathway Enrichment Analysis

Enrichment analysis is a widely used approach to identify biological themes. Here, we implement hypergeometric model to assess whether the number of selected genes associated with reactome pathway is larger than expected. The p values were calculated based the hypergeometric model [1],

```
require(DOSE)
data(geneList)
de <- names(geneList)[abs(geneList) > 1]
head(de)
```

```
## [1] "4312" "8318" "10874" "55143" "55388" "991"

require(ReactomePA)
x <- enrichPathway(gene = de, pvalueCutoff = 0.05,
  readable = T)
head(summary(x))

##           ID                Description GeneRatio  BgRatio
## 69278      69278          Cell Cycle, Mitotic    79/558 394/6438
## 1474244 1474244 Extracellular matrix organization  42/558 155/6438
## 1640170 1640170                Cell Cycle      86/558 486/6438
## 69205     69205      G1/S-Specific Transcription  12/558  15/6438
## 1442490 1442490      Degradation of collagen     22/558  60/6438
## 453277   453277      Mitotic M-M/G1 phases     52/558 255/6438
##           pvalue p.adjust  qvalue
## 69278  2.552e-13 5.666e-11 4.379e-11
## 1474244 7.850e-12 8.713e-10 6.734e-10
## 1640170 2.741e-11 2.029e-09 1.568e-09
## 69205   5.741e-11 3.186e-09 2.462e-09
## 1442490 1.755e-09 7.790e-08 6.021e-08
## 453277  2.387e-09 8.831e-08 6.825e-08
##
## 69278                                CDC45/CDCA8/MCM10/CDC20/KIF
## 1474244
## 1640170 CDC45/CDCA8/MCM10/CDC20/KIF23/CENPE/MYBL2/CCNB2/NDC80/NCAPH/RRM2/UBE2C/H
## 69205
## 1442490
## 453277
##           Count
## 69278         79
## 1474244        42
## 1640170        86
## 69205         12
## 1442490        22
## 453277        52
```

2.1 Visualize enrichment result

We also implement a bar plot and category-gene-network for visualization. It is very common to visualize the enrichment result in bar or pie chart. We believe the pie chart is misleading and only provide bar chart.

```
barplot(x, showCategory = 8)
```

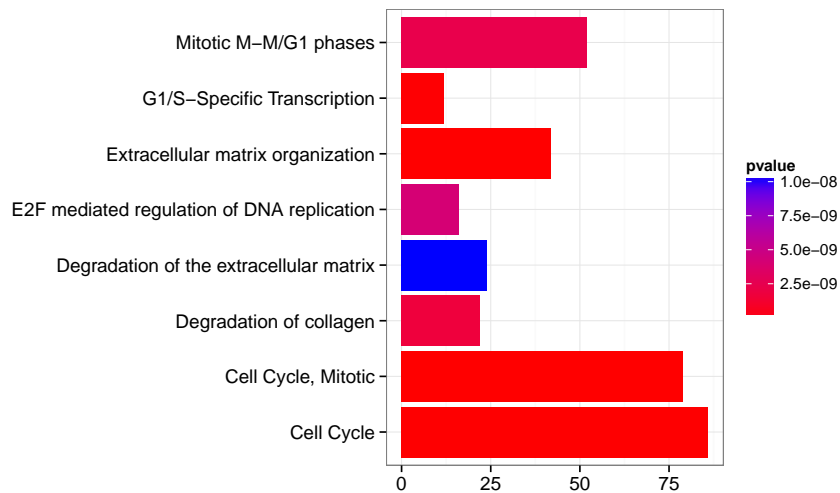


Figure 1: barplot of Reactome Pathway enrichment result.

In order to consider the potentially biological complexities in which a gene may belong to multiple annotation categories, we developed `cnetplot` function to extract the complex association between genes and diseases.

```
cnetplot(x, categorySize = "pvalue", foldChange = geneList)
```

2.2 Comparing enriched reactome pathways among gene clusters with clusterProfiler

We have developed an R package *clusterProfiler* [2] for comparing biological themes among gene clusters. *ReactomePA* works fine with *clusterProfiler* and can compare biological themes at reactome pathway perspective.

```
require(clusterProfiler)
data(gcSample)
res <- compareCluster(gcSample, fun = "enrichPathway")
plot(res)
```

3 Gene Set Enrichment Analysis

A common approach in analyzing gene expression profiles was identifying differential expressed genes that are deemed interesting. The `enrichPathway` function we demonstrated previously were based on these differential expressed genes. This approach will find genes where the difference is large, but it will not detect a situation where the difference is small, but evidenced in coordinated way in a

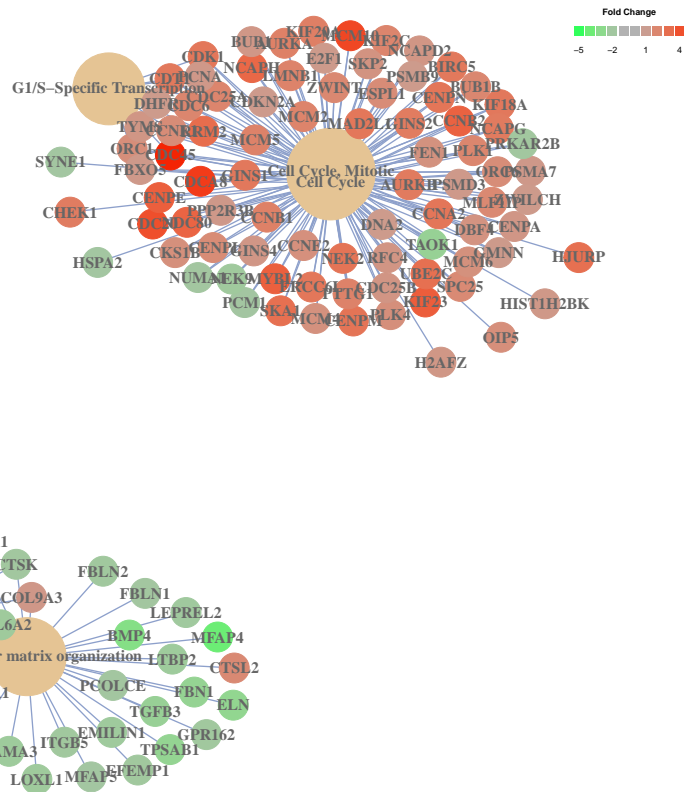


Figure 2: cnetplot of Reactome Pathway enrichment result.

set of related genes. Gene Set Enrichment Analysis (GSEA) directly addressed this limitation. All genes can be used in GSEA; GSEA aggregates the per gene statistics across genes within a gene set, therefore making it possible to detect situations where all genes in a predefined set change in a small but coordinated way.

```

y <- gseAnalyzer(geneList, nPerm = 100, minGSSize = 120,
  pvalueCutoff = 0.05, pAdjustMethod = "BH", verbose = FALSE)
res <- summary(y)
head(res)

##          ID                                     Description setSize
## 162906  162906                                     HIV Infection      180
## 1280218 1280218                               Adaptive Immune System  520
## 168256   168256                                     Immune System      910
## 1280215 1280215           Cytokine Signaling in Immune system  252
## 71291    71291  Metabolism of amino acids and derivatives      165
## 373076   373076  Class A/1 (Rhodopsin-like receptors)          250

```

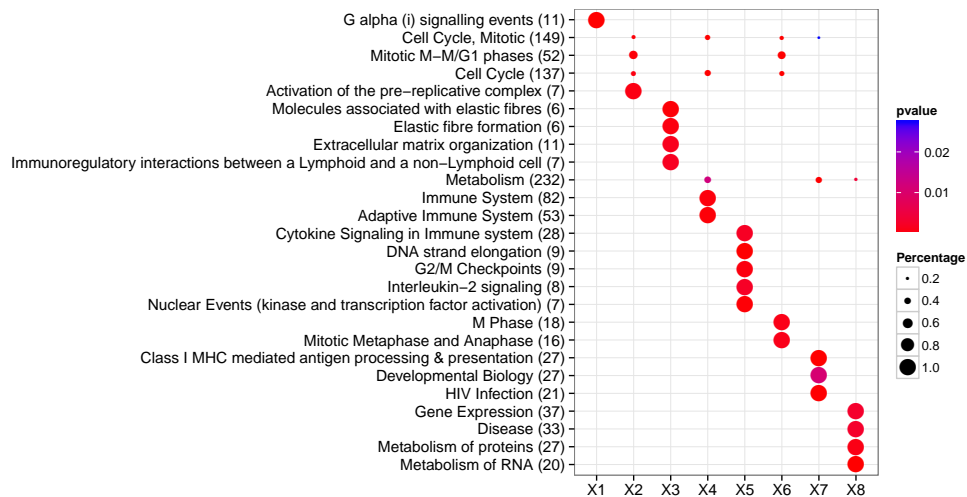


Figure 3: ReactomePA with clusterProfiler.

##	enrichmentScore	pvalues	p.adjust	qvalues
## 162906	0.4989	0	0	0
## 1280218	0.3556	0	0	0
## 168256	0.3204	0	0	0
## 1280215	0.3489	0	0	0
## 71291	0.3314	0	0	0
## 373076	0.2853	0	0	0

3.1 Visualize GSEA result

```
topID <- res[1, 1]
topID

## [1] "162906"

plot(y, geneSetID = topID)
```

4 Pathway Visualization

In *ReactomePA*, we also implemented `viewPathway` to visualized the pathway.

```
viewPathway("E2F mediated regulation of DNA replication",
  readable = TRUE, foldChange = geneList)
```

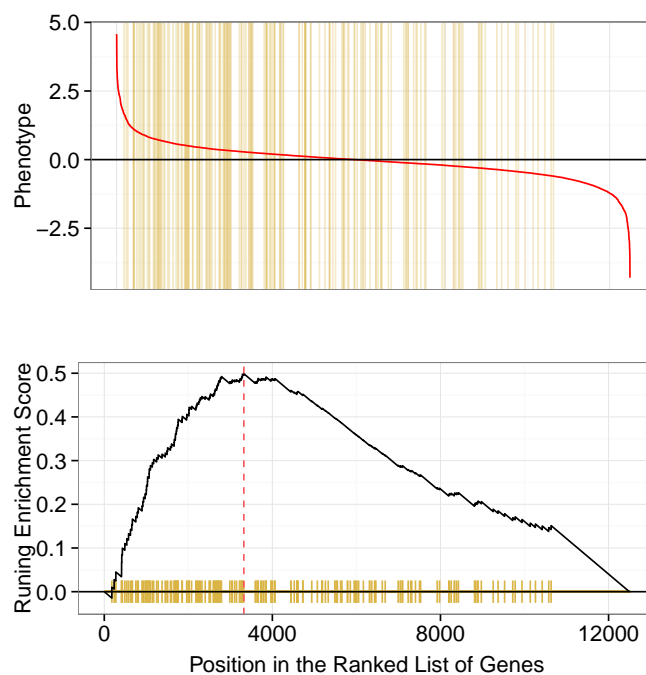


Figure 4: plotting gsea result

5 Session Information

The version number of R and packages loaded for generating the vignette were:

- R version 3.0.1 (2013-05-16), x86_64-apple-darwin10.8.0
- Locale: C/UTF-8/C/C/C/C
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, utils
- Other packages: AnnotationDbi 1.23.16, Biobase 2.20.1, BiocGenerics 0.6.0, DBI 0.2-7, DOSE 1.99.4, RSQLite 0.11.4, ReactomePA 1.5.1, clusterProfiler 1.9.4, ggplot2 0.9.3.1, graph 1.38.2, graphite 1.6.0, knitr 1.2, org.Hs.eg.db 2.9.0
- Loaded via a namespace (and not attached): DO.db 2.6.0, GO.db 2.9.0, GOSemSim 1.19.3, IRanges 1.18.1, KEGG.db 2.9.1, MASS 7.3-27, RColorBrewer 1.0-5, codetools 0.2-8, colorspace 1.2-2, dichromat 2.0-0, digest 0.6.3, evaluate 0.4.4, formatR 0.8, grid 3.0.1, gtable 0.1.2, igraph 0.6.5-2, labeling 0.2, munsell 0.4, plyr 1.8, proto 0.3-10, qvalue 1.34.0, reactome.db 1.44.0, reshape2 1.2.2, scales 0.2.3, stats4 3.0.1, stringr 0.6.2, tcltk 3.0.1, tools 3.0.1

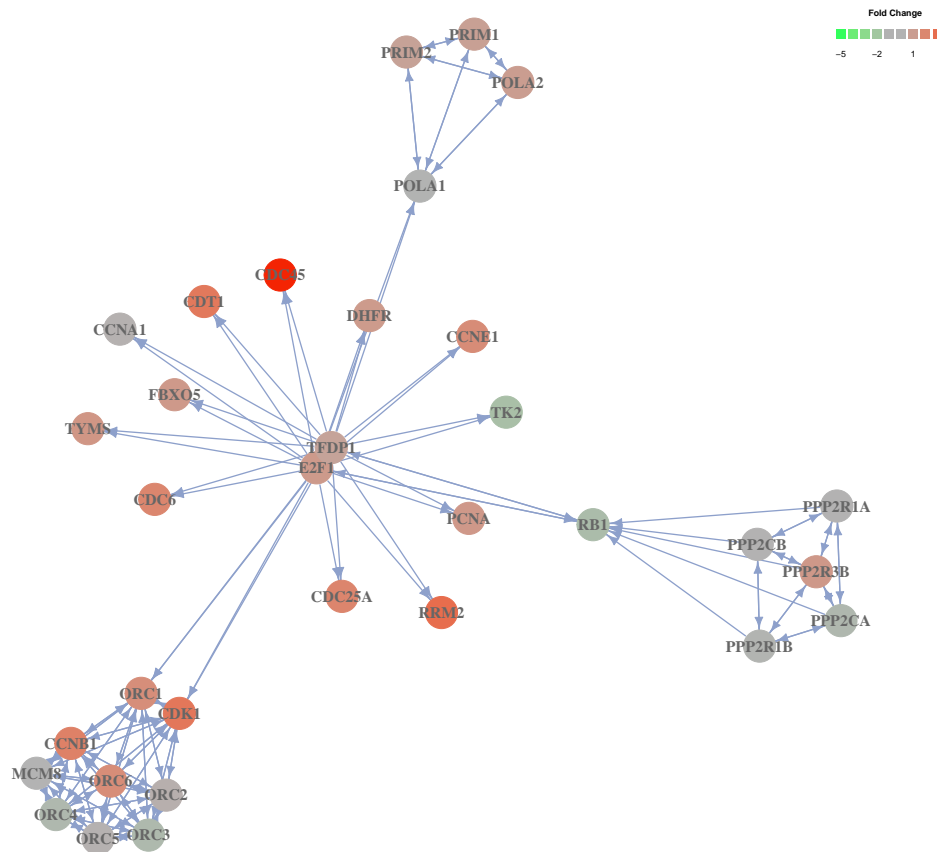


Figure 5: Reactome Pathway visualization.

References

- [1] Elizabeth I Boyle, Shuai Weng, Jeremy Gollub, Heng Jin, David Botstein, J Michael Cherry, and Gavin Sherlock. GO::TermFinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics (Oxford, England)*, 20(18):3710–3715, December 2004. PMID: 15297299.
- [2] Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, May 2012.