

Reactome Pathway Analysis

Guangchuang Yu

April 30, 2014

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)

## Loading required package: org.Hs.eg.db

head(summary(x))
```

##	ID	Description	GeneRatio	BgRatio
## 1474244	1474244	Extracellular matrix organization	59/584	266/6960
## 69205	69205	G1/S-Specific Transcription	12/584	15/6960
## 69278	69278	Cell Cycle, Mitotic	83/584	489/6960
## 1640170	1640170	Cell Cycle	90/584	554/6960
## 1442490	1442490	Collagen degradation	22/584	60/6960
## 113510	113510	E2F mediated regulation of DNA replication	16/584	34/6960

```
##          pvalue p.adjust  qvalue
## 1474244 1.32e-12 3.46e-10 2.77e-10
## 69205   3.94e-11 5.18e-09 4.15e-09
## 69278   1.25e-10 1.10e-08 8.80e-09
## 1640170 2.08e-10 1.37e-08 1.09e-08
## 1442490 9.66e-10 5.08e-08 4.07e-08
## 113510  2.61e-09 1.14e-07 9.15e-08
##
## 1474244
## 69205
## 69278
## 1640170 CDC45/CDCA8/MCM10/CDC20/FOXM1/KIF23/CENPE/MYBL2/CCNB2/NDC80/NCAPH/RRM2/U
## 1442490
## 113510
##          Count
## 1474244     59
## 69205      12
## 69278      83
## 1640170     90
## 1442490     22
## 113510     16
```

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)
```

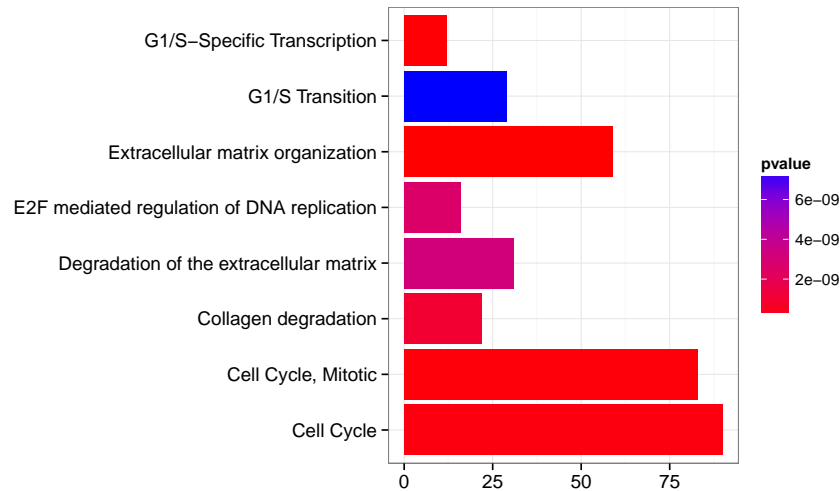


Figure 1: barplot of Reactome Pathway enrichment result.

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

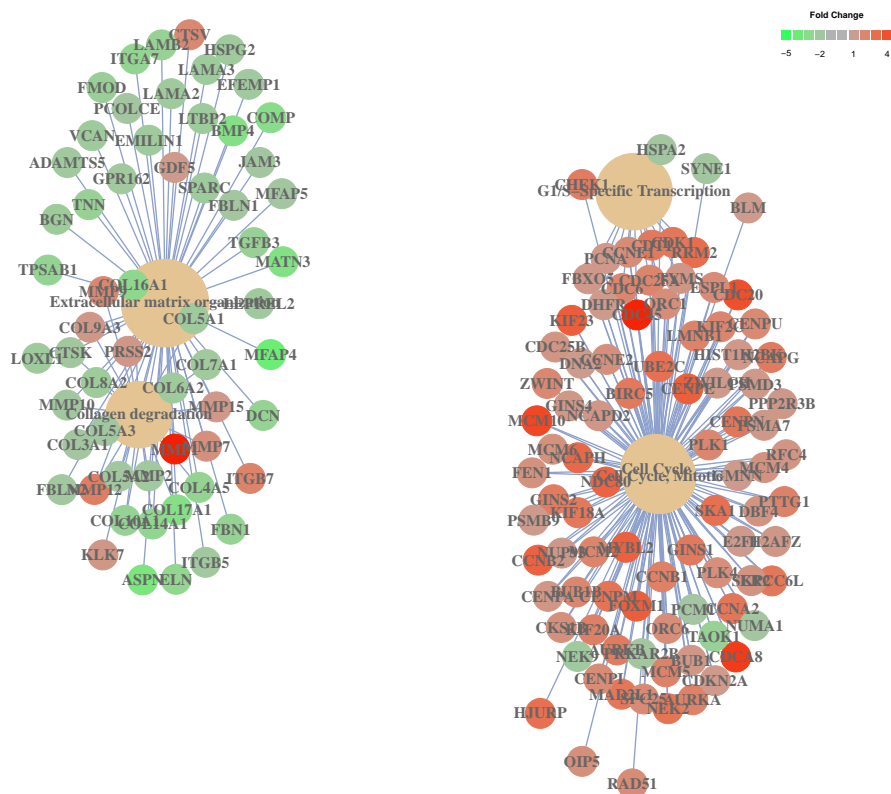


Figure 2: cnetplot of Reactome Pathway enrichment result.

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 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
##	1643685	Disease	919
##	162906	HIV Infection	191
##	1280218	Adaptive Immune System	520

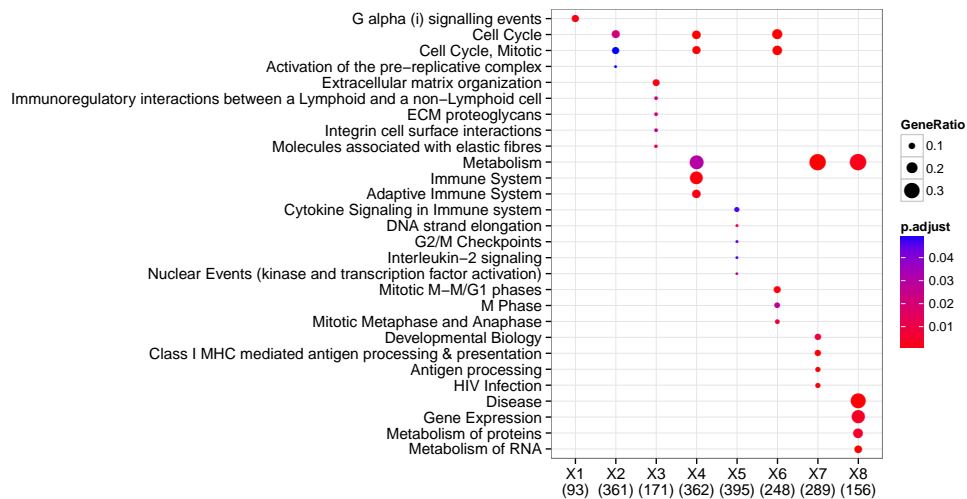


Figure 3: ReactomePA with clusterProfiler.

##	168256	168256		Immune System	951	
##	1280215	1280215		Cytokine Signaling in Immune system	254	
##	71291	71291		Metabolism of amino acids and derivatives	168	
##			enrichmentScore	pvalues	p.adjust	qvalues
##	1643685		0.216	0	0	0
##	162906		0.466	0	0	0
##	1280218		0.356	0	0	0
##	168256		0.316	0	0	0
##	1280215		0.347	0	0	0
##	71291		0.329	0	0	0

3.1 Visualize GSEA result

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

## [1] "1643685"

plot(y, geneSetID = topID)
```

4 Pathway Visualization

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

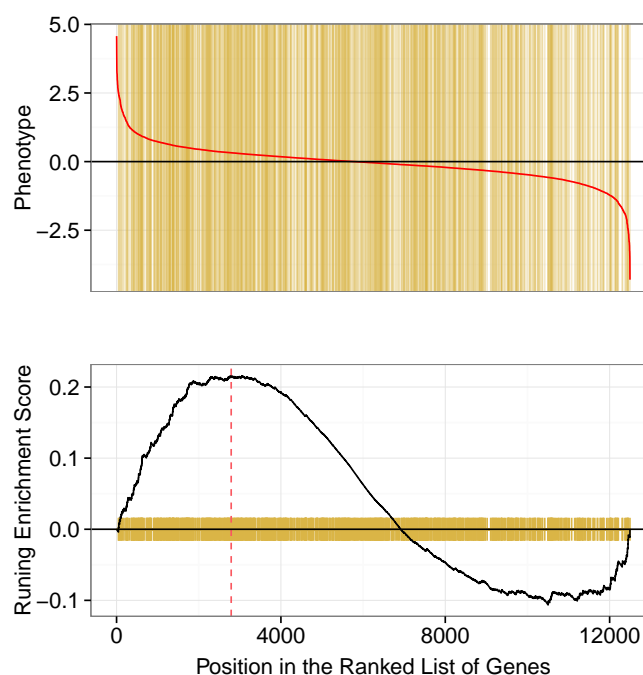


Figure 4: plotting gsea result

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

```
## Loading required package: graphite
```

5 Session Information

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

- R version 3.1.0 RC (2014-04-02 r65358), i386-w64-mingw32
- Locale: LC_COLLATE=C, LC_CTYPE=English_United States.1252, LC_MONETARY=English_United States.1252, LC_NUMERIC=C, LC_TIME=English_United States.1252
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, utils
- Other packages: AnnotationDbi 1.26.0, Biobase 2.24.0, BiocGenerics 0.10.0, DBI 0.2-7, DOSE 2.2.0, GenomInfoDb 1.0.2, RSQLite 0.11.4, ReactomePA 1.8.1, clusterProfiler 1.12.0, ggplot2 0.9.3.1, graph 1.42.0, graphite 1.10.0, knitr 1.5, org.Hs.eg.db 2.14.0
- Loaded via a namespace (and not attached): DO.db 2.8.0, GO.db 2.14.0, GOSemSim 1.22.0, IRanges 1.22.5, KEGG.db 2.14.0, MASS 7.3-32,

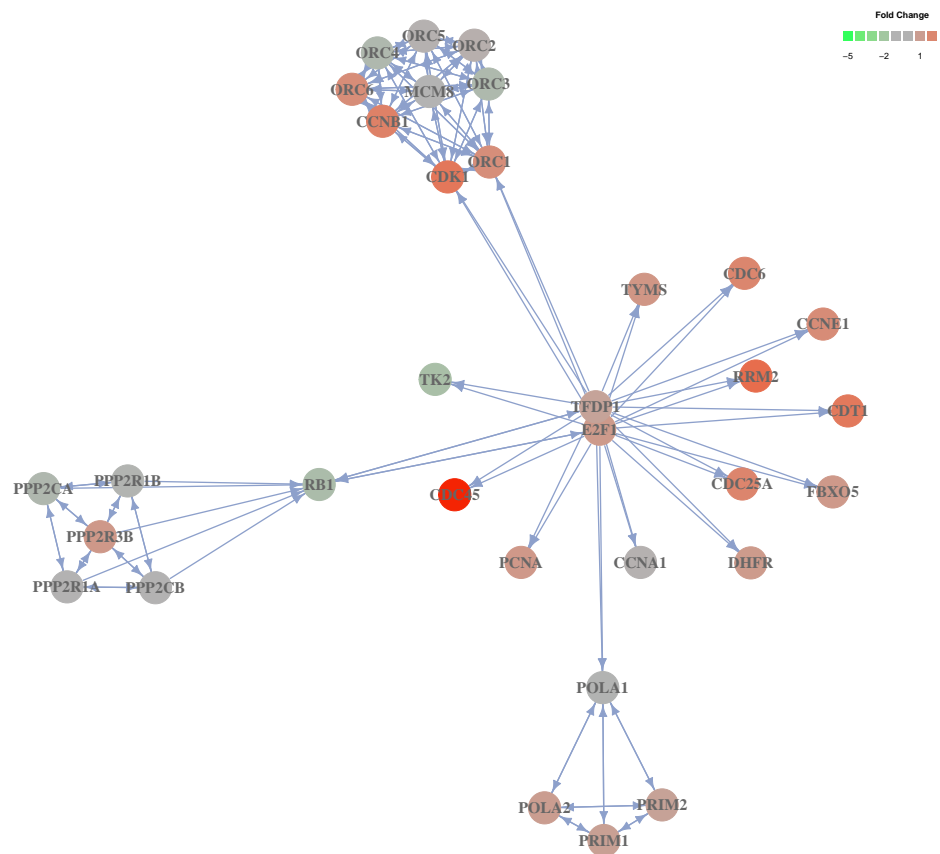


Figure 5: Reactome Pathway visualization.

Rcpp 0.11.1, codetools 0.2-8, colorspace 1.2-4, digest 0.6.4, evaluate 0.5.5, formatR 0.10, grid 3.1.0, gtable 0.1.2, highr 0.3, igraph 0.7.0, labeling 0.2, munsell 0.4.2, plyr 1.8.1, proto 0.3-10, qvalue 1.38.0, reactome.db 1.48.0, reshape2 1.4, scales 0.2.4, stats4 3.1.0, stringr 0.6.2, tcltk 3.1.0, tools 3.1.0

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.