Reactome Pathway Analysis

Guangchuang Yu

November 18, 2014

Contents

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
```

```
## Warning in grep(org, p):  input string 1 is invalid in this locale
```

```
head(summary(x))
```

```
##               ID                                  Description GeneRatio  BgRatio
## 1474244 1474244          Extracellular matrix organization    59/584 266/6958
## 69205     69205                G1/S-Specific Transcription    12/584  15/6958
## 69278     69278                        Cell Cycle, Mitotic    83/584 489/6958
## 1640170 1640170                              Cell Cycle     90/584 554/6958
## 1442490 1442490                      Collagen degradation    22/584  60/6958
## 113510   113510 E2F mediated regulation of DNA replication    16/584  34/6958
##            pvalue p.adjust   qvalue
## 1474244 1.33e-12 3.50e-10 2.80e-10
## 69205   3.95e-11 5.20e-09 4.16e-09
## 69278   1.27e-10 1.11e-08 8.92e-09
## 1640170 2.11e-10 1.39e-08 1.11e-08
## 1442490 9.71e-10 5.11e-08 4.09e-08
## 113510  2.62e-09 1.15e-07 9.19e-08
##
## 1474244
## 69205
## 69278                                                 CDC45/CDCA8/MCM10/CDC20/FOXM1/KIF
## 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 implement bar plot, enrichment map 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)
```

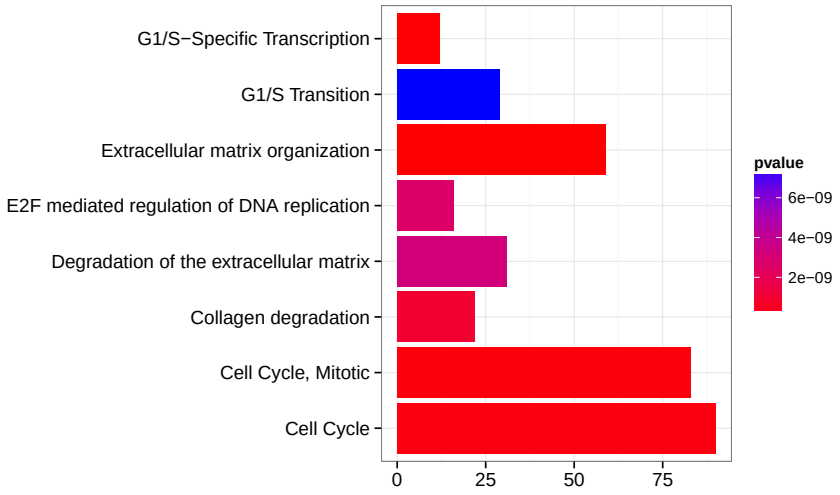

Figure 1: barplot of Reactome Pathway enrichment result.

Enrichment map can be viusalized by `enrichMap` :

```
enrichMap(x)
```

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")

## Warning in grep(org, p):  input string 1 is invalid in this locale

plot(res)
```

Figure 2: enrichment map of enrichment result

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

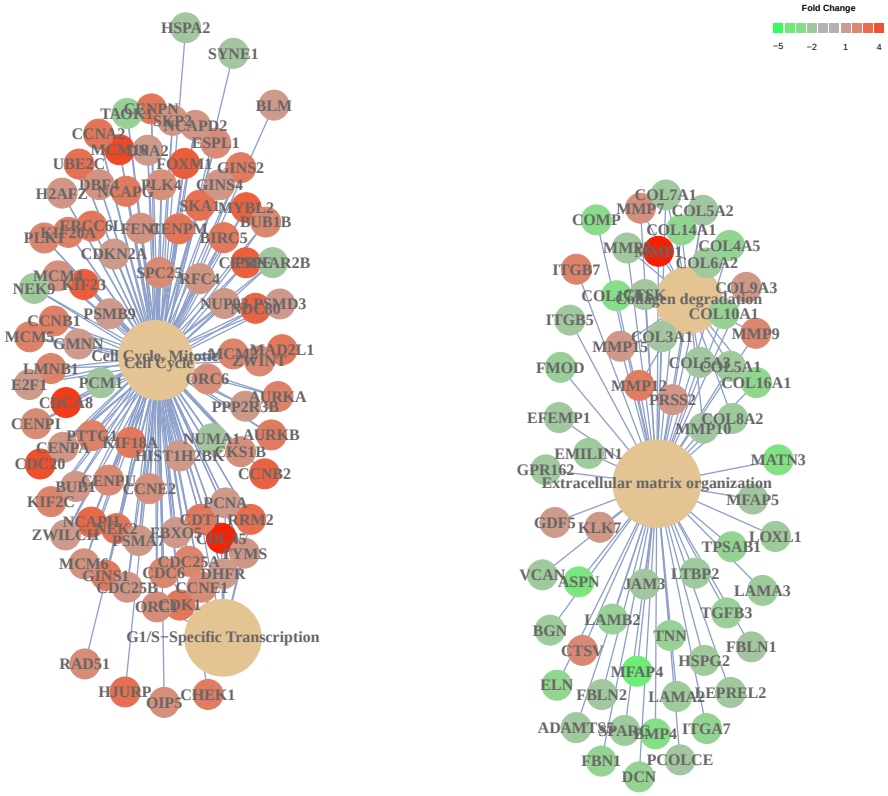

Figure 3: cnetplot of Reactome Pathway enrichment result.

```
##            ID                                  Description setSize enrichmentScore
## 68882   68882                             Mitotic Anaphase     158           0.653
## 68886   68886                                      M Phase     237           0.633
## 69278   69278                        Cell Cycle, Mitotic     393           0.622
## 71291   71291 Metabolism of amino acids and derivatives     168           0.329
## 109581 109581                                   Apoptosis     145           0.428
## 162906 162906                                HIV Infection     191           0.466
##           pvalue p.adjust qvalues
## 68882          0        0       0
## 68886          0        0       0
## 69278          0        0       0
## 71291          0        0       0
## 109581         0        0       0
## 162906         0        0       0
```

3.1 Visualize GSEA result

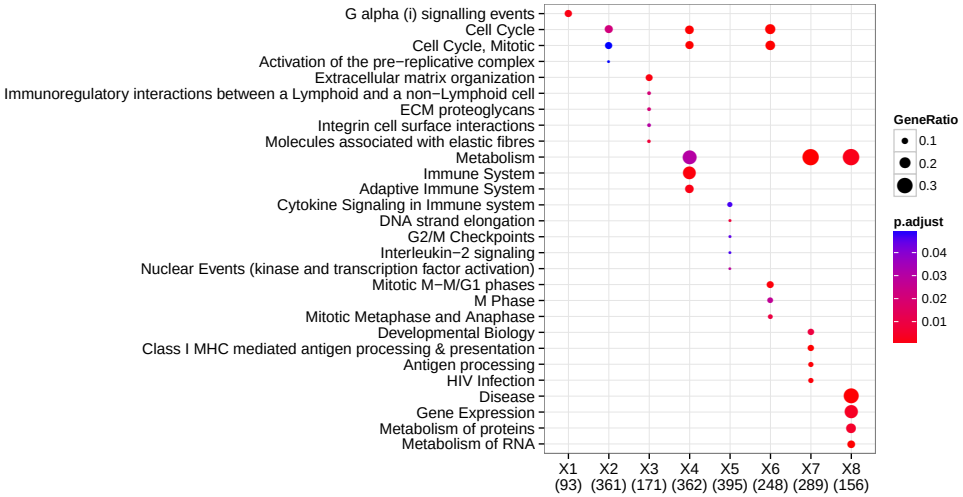

Figure 4: ReactomePA with clusterProfiler.

```
enrichMap(y)
```

```
gseaplot(y, geneSetID = "1280215")
```

4 Pathway Visualization

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

```
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 3.1.2 (2014-10-31), `x86_64-unknown-linux-gnu`
- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_US.UTF-8`, `LC_COLLATE=C`, `LC_MONETARY=en_US.UTF-8`, `LC_MESSAGES=en_US.UTF-8`, `LC_PAPER=en_US.UTF-8`, `LC_NAME=C`, `LC_ADDRESS=C`, `LC_TELEPHONE=C`, `LC_MEASUREMENT=en_US.UTF-8`, `LC_IDENTIFICATION=C`

Figure 5: enrichment map of gsea result

- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils
- Other packages: AnnotationDbi 1.28.1, Biobase 2.26.0, BiocGenerics 0.12.1, DBI 0.3.1, DOSE 2.4.0, GenomeInfoDb 1.2.3, IRanges 2.0.0, RSQLite 1.0.0, ReactomePA 1.10.1, S4Vectors 0.4.0, clusterProfiler 2.0.0, graph 1.44.0, graphite 1.12.0, knitr 1.8, org.Hs.eg.db 3.0.0
- Loaded via a namespace (and not attached): DO.db 2.8.0, GO.db 3.0.0, GOSemSim 1.24.1, KEGG.db 3.0.0, MASS 7.3-35, Rcpp 0.11.3, codetools 0.2-9, colorspace 1.2-4, digest 0.6.4, evaluate 0.5.5, formatR 1.0, ggplot2 1.0.0, grid 3.1.2, gtable 0.1.2, highr 0.4, igraph 0.7.1, labeling 0.3, munsell 0.4.2, plyr 1.8.1, proto 0.3-10, qvalue 1.40.0, reactome.db 1.50.0, reshape2 1.4, scales 0.2.4, stringr 0.6.2, tools 3.1.2

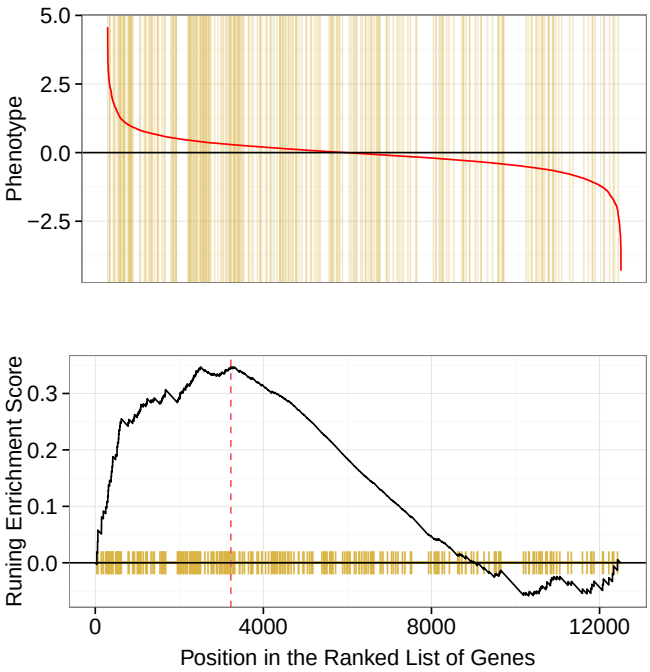

Figure 6: plotting gsea result

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.

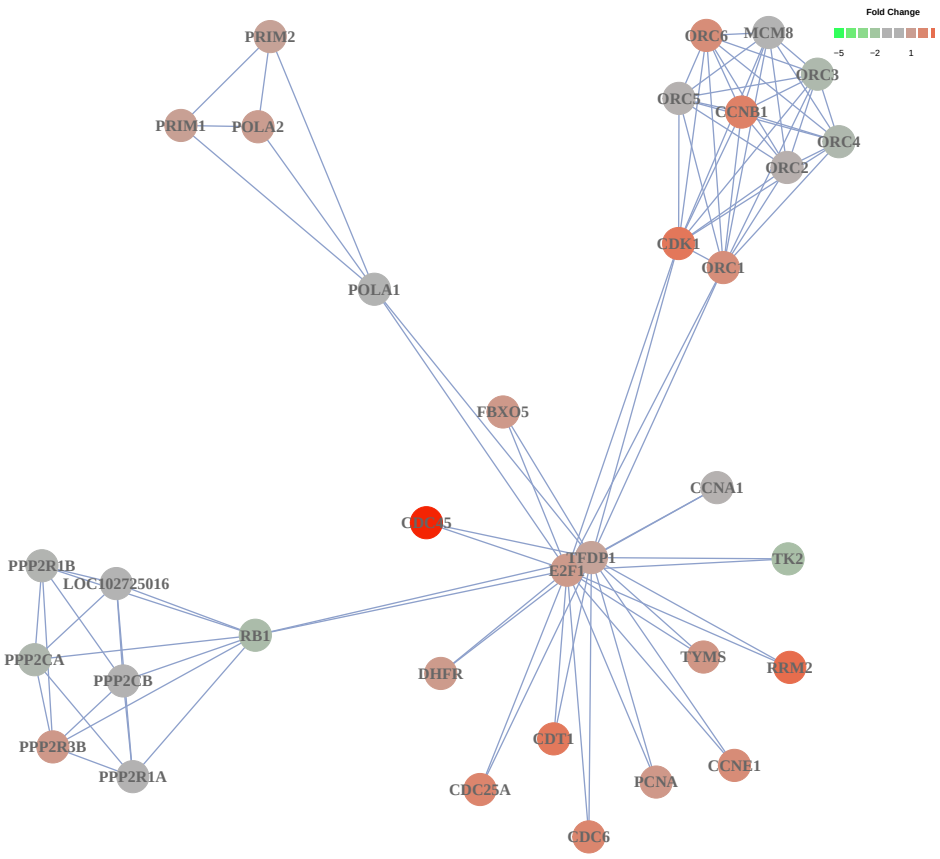

Figure 7: Reactome Pathway visualization.