

Reactome Pathway Analysis

Guangchuang Yu

Jinan University, Guangzhou, China

May 22, 2012

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.

In *ReactomePA*, we plan to implement:

- pathway enrichment analysis
- gene set enrichment analysis
- functional subpathway (active or perturbed subpathway) detection
- methods for visualization.

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. We also implement a category net model for viusalization.

- Hypergeometric model

Hypergeometric model was implemented to assess whether the number of selected genes associated with reactome pathway is larger than expected.

- Category Net Plot

Category-gene network model was implemented to extract the complex relationships between genes and pathways. It provides a high-level model to understand the functionalities of genes.

- Case Study

Here, we used a vector of sample entrezgene ID, which was converted from an example list of genes from ProfCom <http://webclu.bio.wzw.tum.de/profcom/examples.php>.

```
> require(ReactomePA)
> data(sample)
> sample
```

```

[1] "10140" "1917" "1672" "9445" "5950" "307"
[7] "978" "7077" "6539" "3569" "1545" "1368"
[13] "6590" "3081" "3059" "5797" "5817" "5095"
[19] "4118" "6948" "1282" "1284" "2261" "4837"
[25] "1311" "1428" "3162" "5004" "728441" "2678"
[31] "5744" "3484" "2745" "2064" "3486" "9971"
[37] "8566" "7453" "11015" "10397" "4162" "7980"
[43] "486" "6696" "3929" "22795" "4286" "1410"
[49] "4653" "3915" "5358" "3912" "6781" "1474"
[55] "6648" "2719" "3306"

```

```

> x <- enrichPathway(gene=sample,pvalueCutoff=0.05, qvalueCutoff=0.05, readable=T)
> head(summary(x))

```

```

      pathwayID
216083      216083
422475      422475
381426      381426
1430728     1430728
381150      381150
1266738     1266738

216083
422475
381426 Homo sapiens: Regulation of Insulin-like Growth Factor (IGF) Activity by I
1430728
381150
1266738

      GeneRatio      BgRatio      pvalue      qvalue
216083      5/57      85/25529 1.327366e-06 0.0001075865
422475      5/57      296/25529 5.181103e-04 0.0176133918
381426      2/57      17/25529 6.519242e-04 0.0176133918
1430728      9/57     1136/25529 8.693219e-04 0.0176152070
381150      3/57      120/25529 2.463253e-03 0.0328676665
1266738      5/57     421/25529 2.466915e-03 0.0328676665

                                geneID Count
216083                                COL4A1/COL4A2/SPP1/LAMC1/LAMB1      5
422475                                COL4A1/COL4A2/ERBB2/LAMC1/LAMB1      5
381426                                IGFBP1/IGFBP3                    2
1430728 CDA/CYP1B1/HGD/PCCA/NNMT/HMOX1/GGT1/GLRX/PDXK              9
381150                                IGFBP1/IGFBP3/KDEL3              3
1266738                                COL4A1/COL4A2/ERBB2/LAMC1/LAMB1      5

```

```

> plot(x, showCategory=5)

```

- Compatible with *clusterProfiler* Bioconductor package *clusterProfiler* designed visualization for comparing biological themes among gene clusters (Yu et al., 2012). More details and parameters are described in the documentation (Rfunction?compareCluster). Figure 2 has been generated using the data, as in Yu et al. (2012).

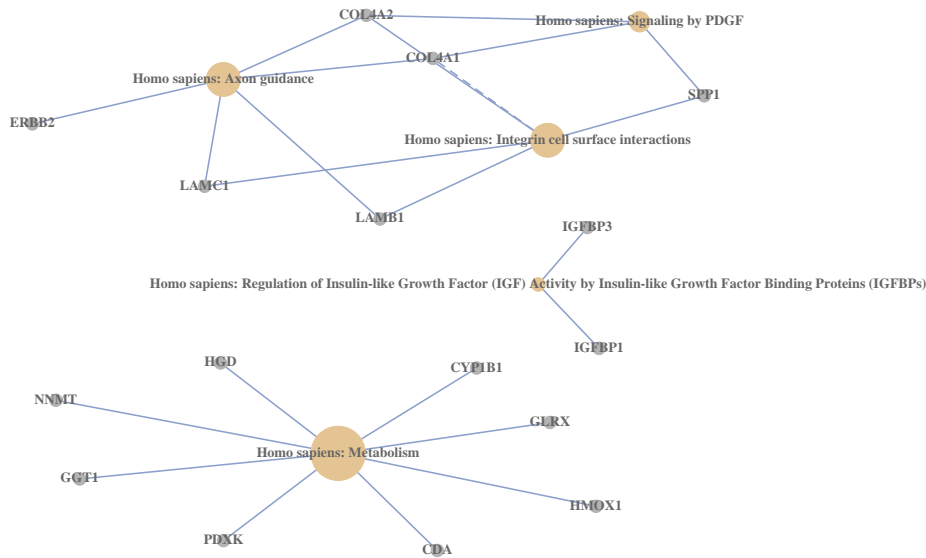


Figure 1: Visualization of Pathway enrichment analysis

3 Session Information

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

R version 2.15.0 (2012-03-30)

Platform: x86_64-unknown-linux-gnu (64-bit)

locale:

```
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8      LC_COLLATE=C
[5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=C                LC_NAME=C
[9] LC_ADDRESS=C             LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

attached base packages:

```
[1] stats      graphics  grDevices  utils      datasets
[6] methods    base
```

other attached packages:

```
[1] org.Hs.eg.db_2.7.1    ReactomePA_1.0.1
[3] AnnotationDbi_1.18.0 Biobase_2.16.0
[5] BiocGenerics_0.2.0   RSQLite_0.11.1
[7] DBI_0.2-5
```

loaded via a namespace (and not attached):

```
[1] IRanges_1.14.3      igraph0_0.5.5      plyr_1.7.1
[4] qvalue_1.30.0       reactome.db_1.0.40 stats4_2.15.0
```

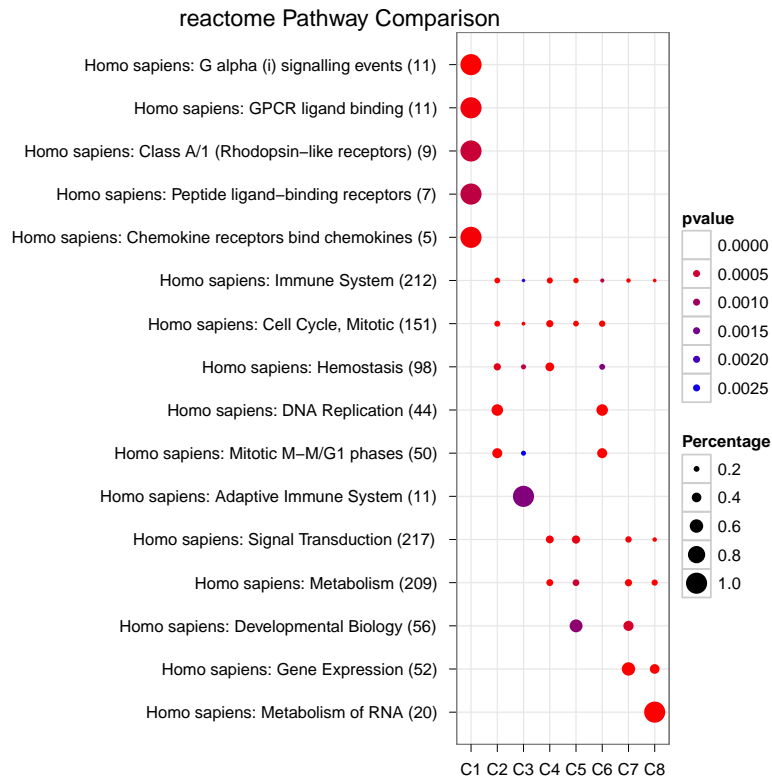


Figure 2: Example of working with clusterProfiler package

```
[7] tcltk_2.15.0      tools_2.15.0
```

References

Guangchuang Yu, Le-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:in press, 2012. ISSN 1536-2310.