

Reactome Pathway Analysis

Guangchuang Yu

Jinan University, Guangzhou, China

November 2, 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))

```

```

ID
216083 216083
186797 186797
422475 422475
1430728 1430728
381426 381426
1474244 1474244

```

```

216083
186797
422475
1430728

```

```

381426 Homo sapiens: Regulation of Insulin-like Growth Factor (IGF) Activity by I
1474244

```

```

GeneRatio BgRatio pvalue qvalue
216083 5/57 85/29492 6.569716e-07 5.878167e-05
186797 5/57 190/29492 3.359400e-05 1.502890e-03
422475 5/57 295/29492 2.648645e-04 7.656156e-03
1430728 10/57 1414/29492 3.422752e-04 7.656156e-03
381426 2/57 17/29492 4.899036e-04 8.766697e-03
1474244 3/57 91/29492 7.370647e-04 1.099132e-02

```

```

geneID
216083 COL4A1/COL4A2/SPP1/LAMC1/LAMB1
186797 COL4A1/COL4A2/FGFR3/ERBB2/SPP1
422475 COL4A1/COL4A2/ERBB2/LAMC1/LAMB1
1430728 CDA/CYP1B1/HGD/PCCA/NNMT/HMOX1/GGT1/GLRX/PDXK/GPC3
381426 IGFBP1/IGFBP3
1474244 TIMP2/COL4A1/COL4A2

```

```

Count
216083 5
186797 5
422475 5
1430728 10
381426 2
1474244 3

```

```

> plot(x, showCategory=5)

```

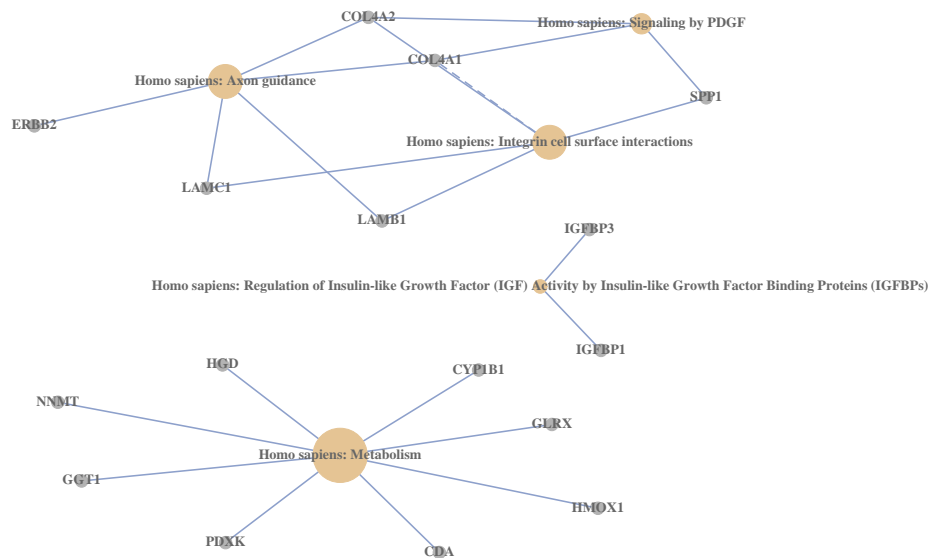


Figure 1: Visualization of Pathway enrichment analysis

- Compatibal 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).

3 Gene Set Enrichment Analysis

To be developed.

4 Session Information

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

```
R version 2.15.2 (2012-10-26)
Platform: i386-w64-mingw32/i386 (32-bit)

locale:
[1] LC_COLLATE=C
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

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

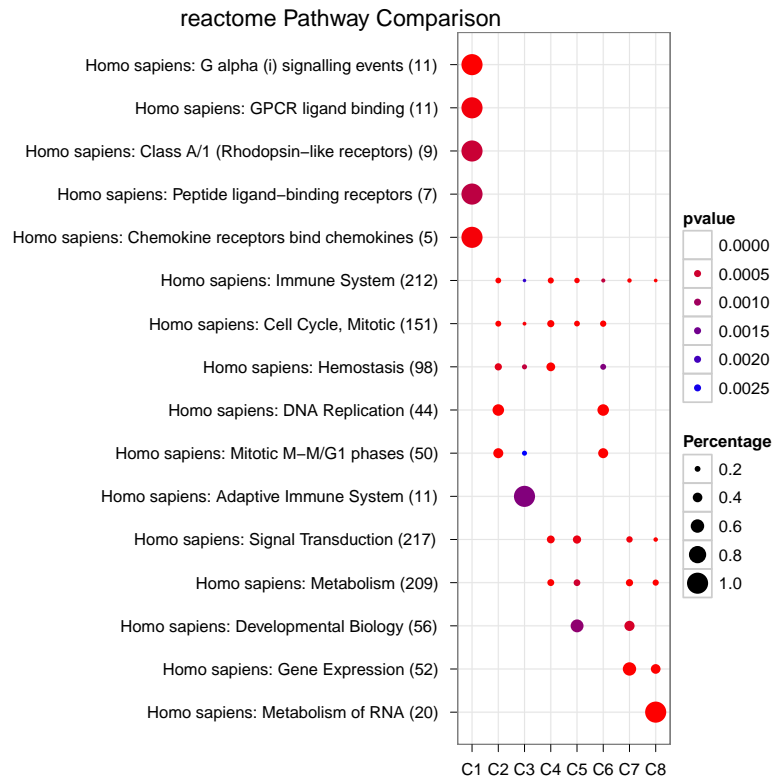


Figure 2: Example of working with clusterProfiler package

other attached packages:

```
[1] org.Hs.eg.db_2.8.0      ReactomePA_1.2.1
[3] DOSE_1.4.0              AnnotationDbi_1.20.2
[5] Biobase_2.18.0          BiocGenerics_0.4.0
[7] RSQLite_0.11.2          DBI_0.2-5
[9] ggplot2_0.9.2.1
```

loaded via a namespace (and not attached):

```
[1] DO.db_2.5.0             GO.db_2.8.0
[3] GOSemSim_1.16.1         IRanges_1.16.4
[5] MASS_7.3-22             RColorBrewer_1.0-5
[7] colorspace_1.2-0        dichromat_1.2-4
[9] digest_0.5.2            grid_2.15.2
[11] gtable_0.1.1            igraph_0.6-3
[13] labeling_0.1            memoise_0.1
[15] munsell_0.4             parallel_2.15.2
[17] plyr_1.7.1              proto_0.3-9.2
[19] qvalue_1.32.0           reactome.db_1.42.0
[21] reshape2_1.2.1          scales_0.2.2
[23] stats4_2.15.2           stringr_0.6.1
```

[25] tcltk_2.15.2 tools_2.15.2

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.