

Table of Contents

[NIPA 1](#_Toc498946813)

[Code availability 1](#_Toc498946814)

[hypergeometric tests 1](#_Toc498946815)

[User input to change 2](#_Toc498946816)

[Required Packages 2](#_Toc498946817)

[User options 2](#_Toc498946818)

[Output 3](#_Toc498946819)

[Gene Ontology (Enrichment using GOstats) 3](#_Toc498946820)

[Output Tables 3](#_Toc498946821)

[Enriched GO figures 3](#_Toc498946822)

[GO Directed acyclic graph 4](#_Toc498946823)

[Pathways (KEGG) 6](#_Toc498946824)

[NIPA.report.txt: 7](#_Toc498946825)

# NIPA

## Code availability

https://github.com/ADAC-UoN/NIPA

Code to determine enriched Gene Ontology and Pathways using KEGG using hypergeometric tests.

### hypergeometric tests

using phyper

phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)

x, q vector of quantiles representing the number of white balls drawn

without replacement from an urn which contains both black and white

balls.

m the number of white balls in the urn.

n the number of black balls in the urn.

k the number of balls drawn from the urn.

if

pop size : 5260 # total number of entrez gene in all pathways

sample size : 131 # total goi

Number of items in the pop that are classified as successes : 1998 # entrez in a particular pathway

Number of items in the sample that are classified as successes : 62 # goi in a particular pathway

e.g pathway 100 genes 10 are in goi list of size 400 universe = 20,000

phyper(1,100,20000-100,400, lower.tail=FALSE) = 0.597 = probability of finding this many or greater goi in pathway

phyper(80,100,20000-100,400, lower.tail=FALSE) = 4.603708e-122 = probability of finding this many or greater goi in pathway

# User input to change

### Required Packages

See Bioconductor for instructions to install.

library(biomaRt)

library(pathview)

library(gage)

library(gageData)

library(ggplot2)

library(stringr)

library(dplyr)

library(RamiGO)

### User options

The code in lines 25-65 will need editing



Options

|  |  |
| --- | --- |
| goi.column | number of column with gene identifier in starts form 1. If a simple list, keep as = 1. |
| goi.header = "yes" | # "yes" or "no" if header on file |
| species = "mouse" | #currently one of "mouse", "human", "rat", "pig", "zebrafish, cow, fly, sheep", |
| outfile.prefix <- "ADAC.test" | prefix attached to output files. Change to user requirements. |
| keggFC = "yes" | yes or no. If yes will colour enriched KEGG pathways by FC data [specify column below] |
| keggFC.col = 5 | if keggFC = yes specify column of input table with FC values assumes tab delimited |
| id.type = "Entrez" | change to correct accepted identifier type |
| kegg.qval.cutoff = 0.05 | qvalue cutoff for KEGG enrichment analysis |
| GO.cutoff = 0.05 | qvalue cutoff for GO enrichment analysis |
| min.genes.cutoff = 2 | Miniumum number of genes in category to be reported |
| doGO = "yes" | # yes or no. Run hypergeometric test to find enriched GO terms in BP, MF and CC category |
| doKEGG = "yes" | # yes or no. Run hypergeometric test to find and plot enriched KEGG pathways and visualise using PathView |

# Output

## Gene Ontology (Enrichment using GOstats)

### Output Tables

GO enriched terms GO.BP.table, GO.CC.table, GO.MF.table for "Biological Process", "Cellular Component" and "Molecular Function" respectively. e.g. for Biological Process

Columns are

GO ID,

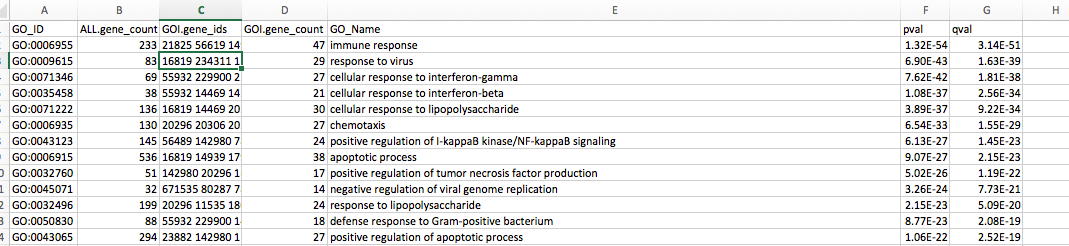
All\_gene\_count (number of genes in GO term)

GOI\_gene\_count (number of genes in GO term also in user input “genes of interest” list)

Pval (p value from hypergeometric test)

Qval (Benjamini-Hochberg corrected pvalue)

This is a tab-delimited text file that can open in excel or similar.



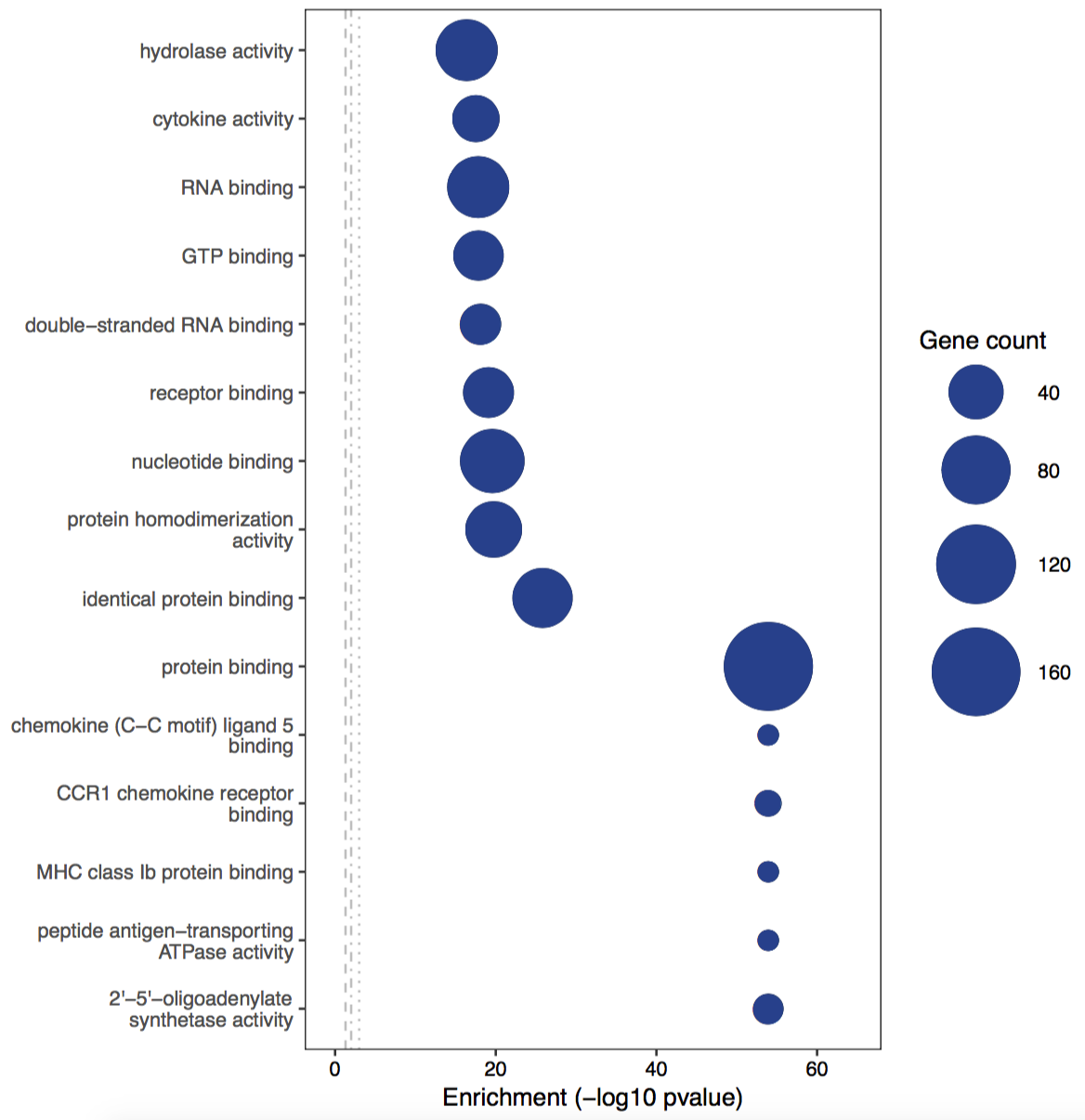
### Enriched GO figures

GO.BP.Significant.enrichment.plot.pdf, GO.MF.Significant.enrichment.plot.pdf and GO.CC.Significant.enrichment.plot.pdf

Example below

x axis = Enrichment (-log10 pvalue)

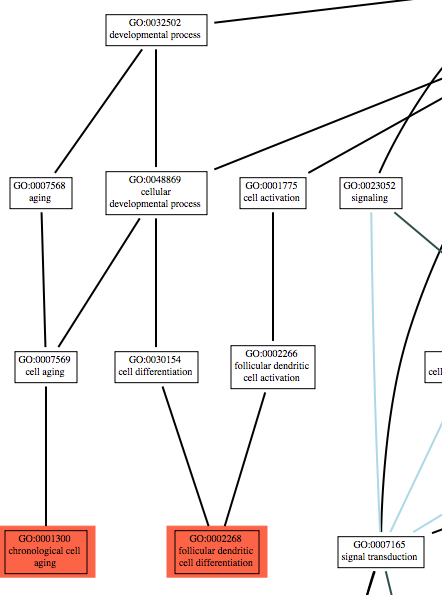
The number of genes in each term are shown by size circle (bigger circles reflect more GOI in the GO terms).



### GO Directed acyclic graph

The relationships of enriched GO terms are shown in the directed acyclic graph in svg format e.g. GO.BP.top.DAG.svg.

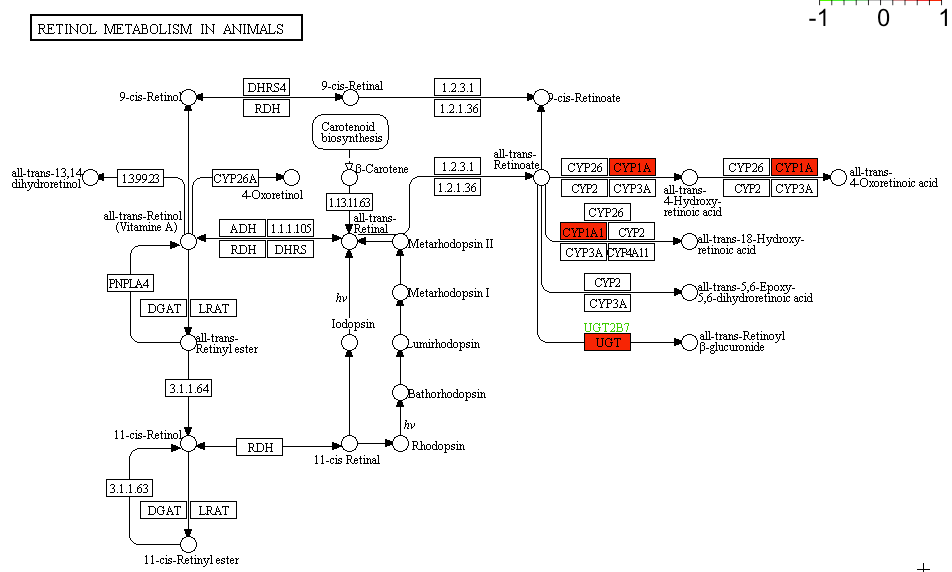
Figures are generated using getAmigoTree see https://www.rdocumentation.org/packages/RamiGO/versions/1.18.0/topics/getAmigoTree



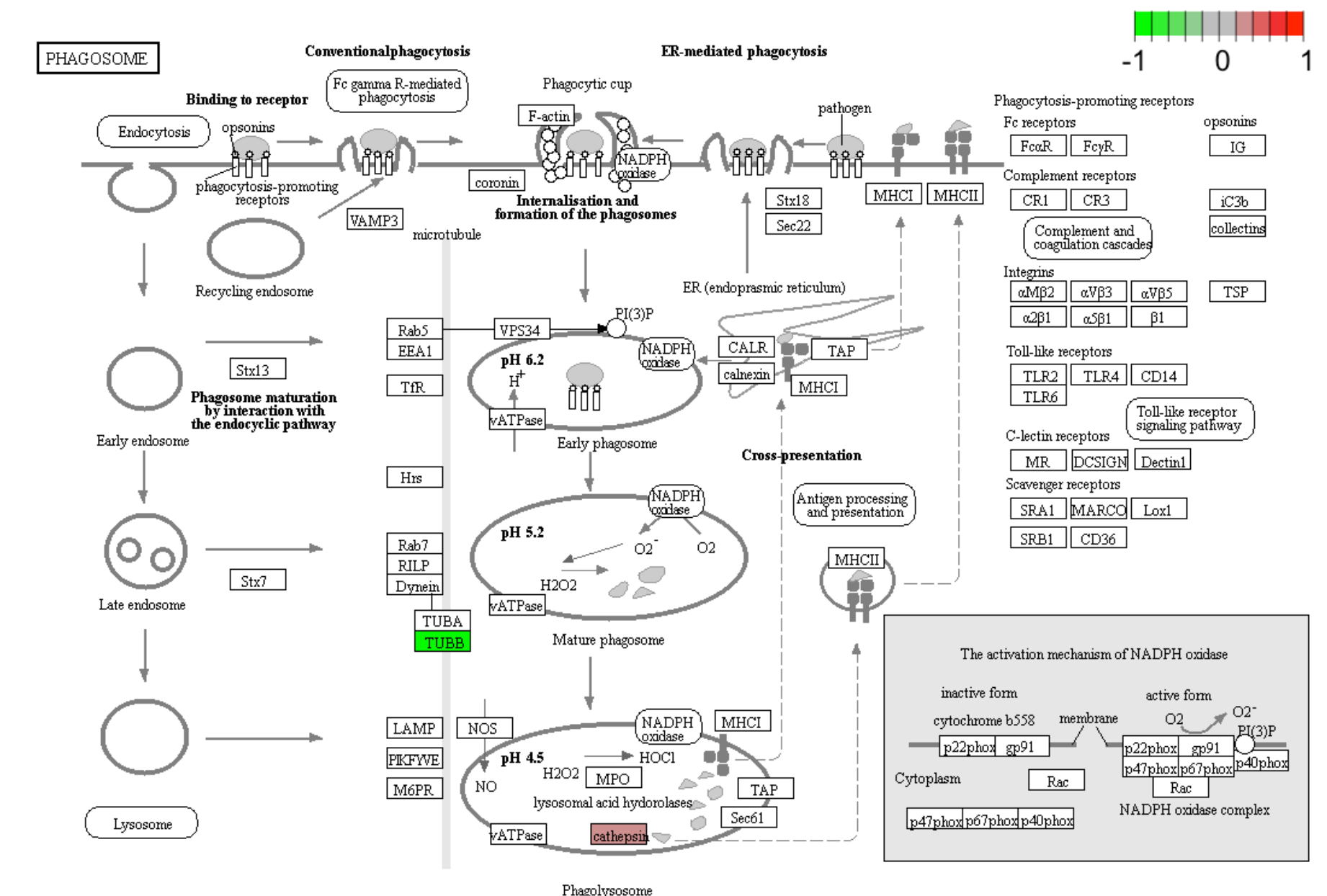
## Pathways (KEGG)

(KEGG.enrichment.analysis.results.table) pathview output e.g mmu03010.pathview.png: red shows entities in user input list which are present in enriched pathway

If no fold change values are given to NIPA (option keggFC = no)- Those genes enriched in your dataset are shown in red



If fold change values are given to NIPA (option keggFC = yes)- Fold change values for those genes enriched in your dataset are shown in in colour scale of green-red [-ve to +ve fold change].



## NIPA.report.txt:

Any errors will appear here.