Advanced Data Analysis Centre

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Table of Contents

NIPA 1
Code availability 1
Output 2
Gene Ontology (Enrichment using GOstats) 2
Reactome 4
Pathways (KEGG) 4
NIPA.report.txt: 5

NIPA

Code availability

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

Code to determine enriched Gene Ontology and Pathways using KEGG and Reactome 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

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, Pvalue, Odds Ratio, Expected Count, Count in your data, size of term in genome and description.

e.g.

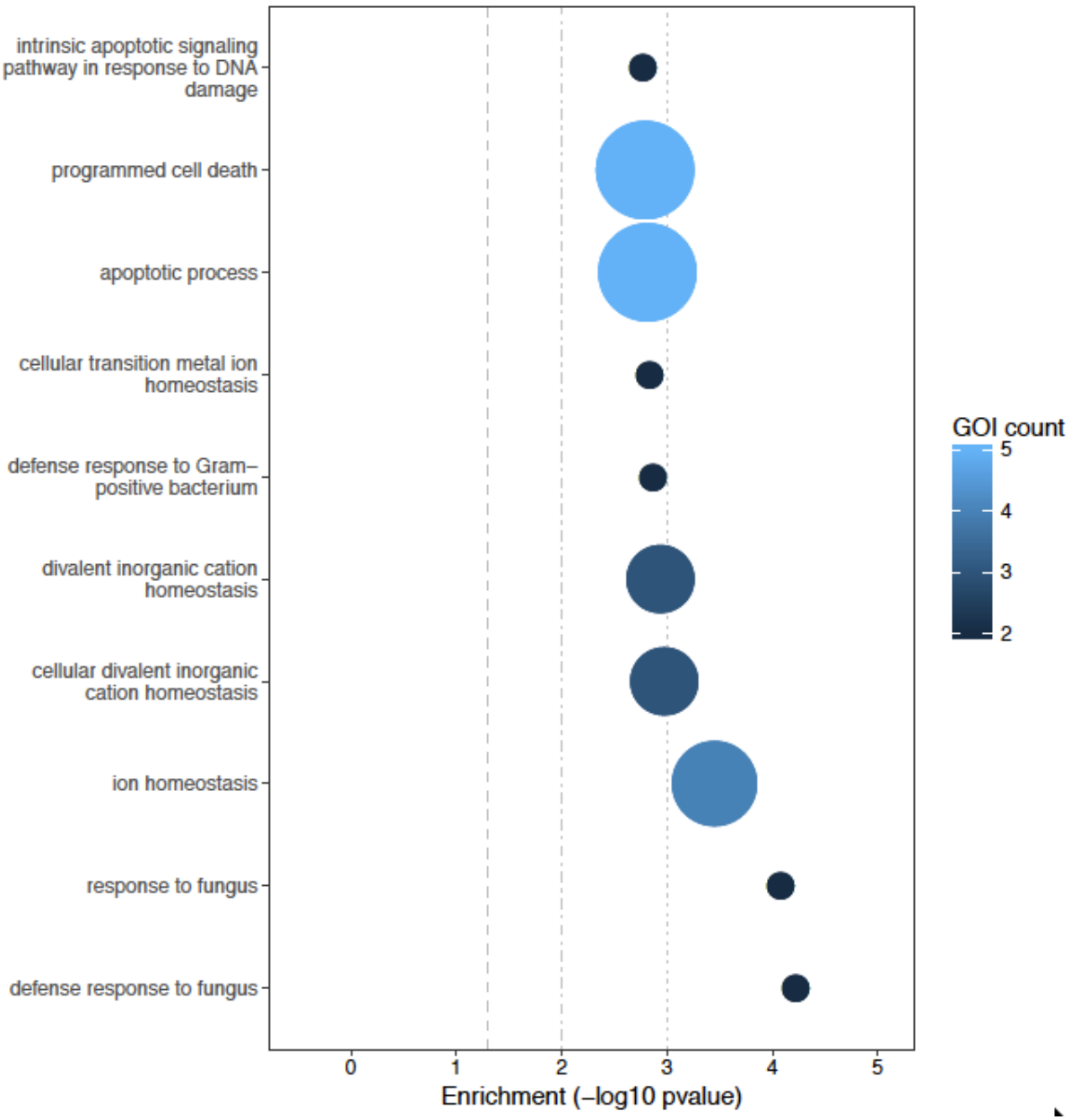
GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:2000826	0.0002825	110.115385	0.0260915	2	10	regulation of heart morphogenesis
GO:0048863	0.0003864	26.296296	0.1487215	3	57	stem cell differentiation
GO:0051241	0.0005481	9.850181	0.7357801	5	282	negative regulation of multicellular organismal process
GO:0001755	0.0007467	62.857143	0.0417464	2	16	neural crest cell migration
GO:0021766	0.0007467	62.857143	0.0417464	2	16	hippocampus development
GO:0007405	0.0009491	54.980769	0.0469647	2	18	neuroblast proliferation

Enriched terms are also shown in 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 and colour of circle.


intrinsic apoptotic signaling pathway in response to DNA damage
programmed cell death
apoptotic process
cellular transition metal ion homeostasis
defense response to Gram–positive bacterium
divalent inorganic cation homeostasis
cellular divalent inorganic cation homeostasis
ion homeostasis
response to fungus
defense response to fungus
0
1
2
3
4
5
Enrichment (–log10 pvalue)
GOI count
5
4
3
2

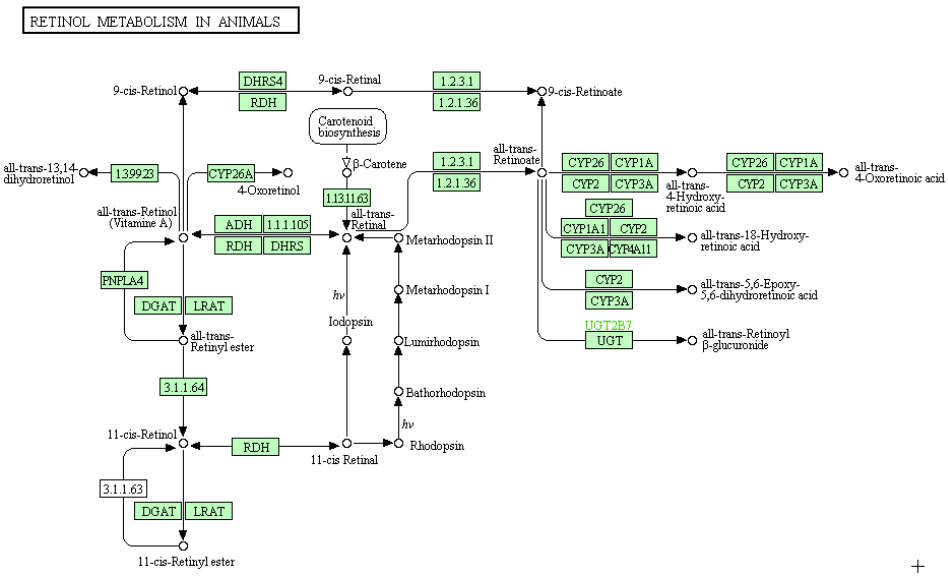
Reactome

If Human, mouse or rat data and the option is selected then enriched reactome terms are also plotted

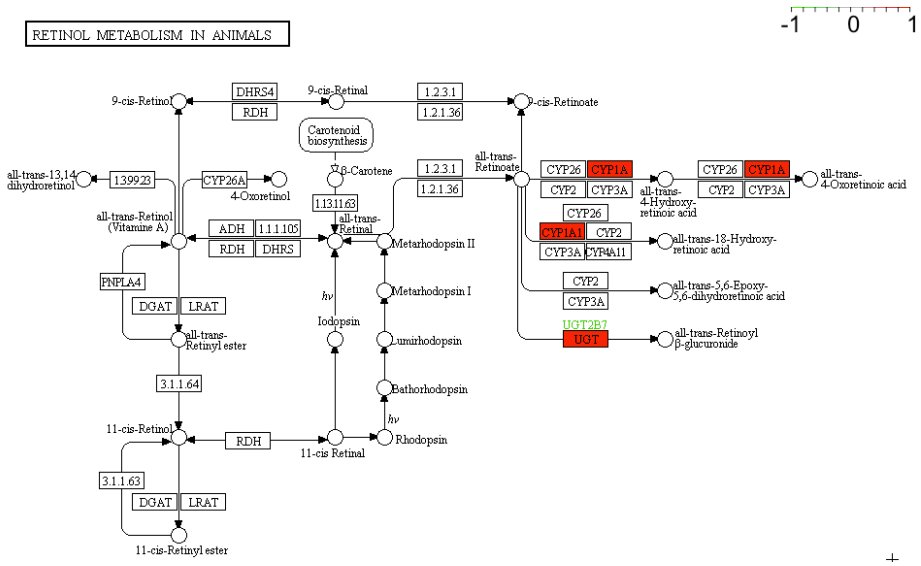
Pathways (KEGG)

(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

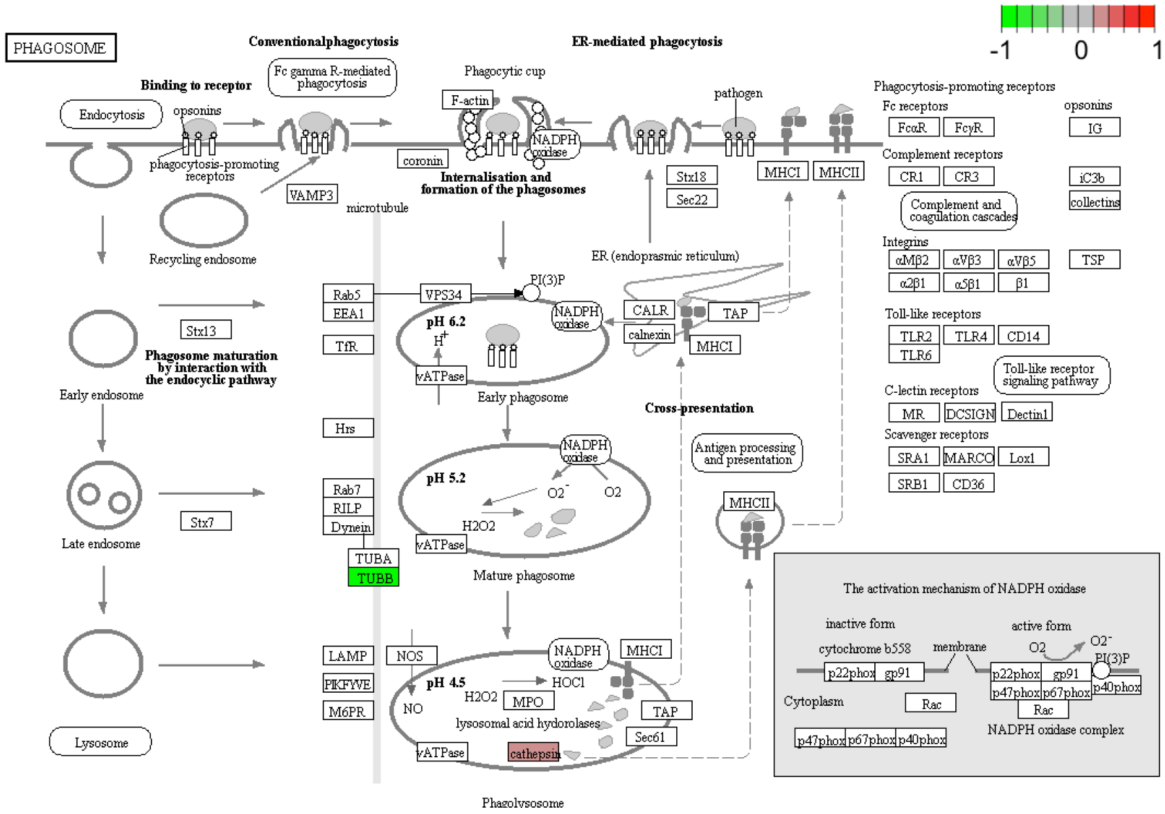
e.g. mmu03010.png: all entities present in test species shown in green



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.