

# Exploring and Presenting Results

Simon Andrews, Laura Biggins

[simon.andrews@babraham.ac.uk](mailto:simon.andrews@babraham.ac.uk)  
[laura.biggins@babraham.ac.uk](mailto:laura.biggins@babraham.ac.uk)

v2020-10



# Functional enrichment results

- Gene set information
  - Gene set name
  - Gene set source
  - Gene set description
- Count information
  - Hit genes in category
  - Hit genes outside category
  - Background genes in category
  - Background genes outside category
- Statistical information
  - Raw p-value
  - Corrected p-value
  - Enrichment value

# Functional enrichment results

- Gene set information
  - Gene set name
  - Gene set source
  - Gene set description
- Count information
  - Hit genes in category
  - Hit genes outside category
  - Background genes in category
  - Background genes outside category
- Statistical information
  - Raw p-value
  - Corrected p-value
  - Enrichment value

# Presenting results

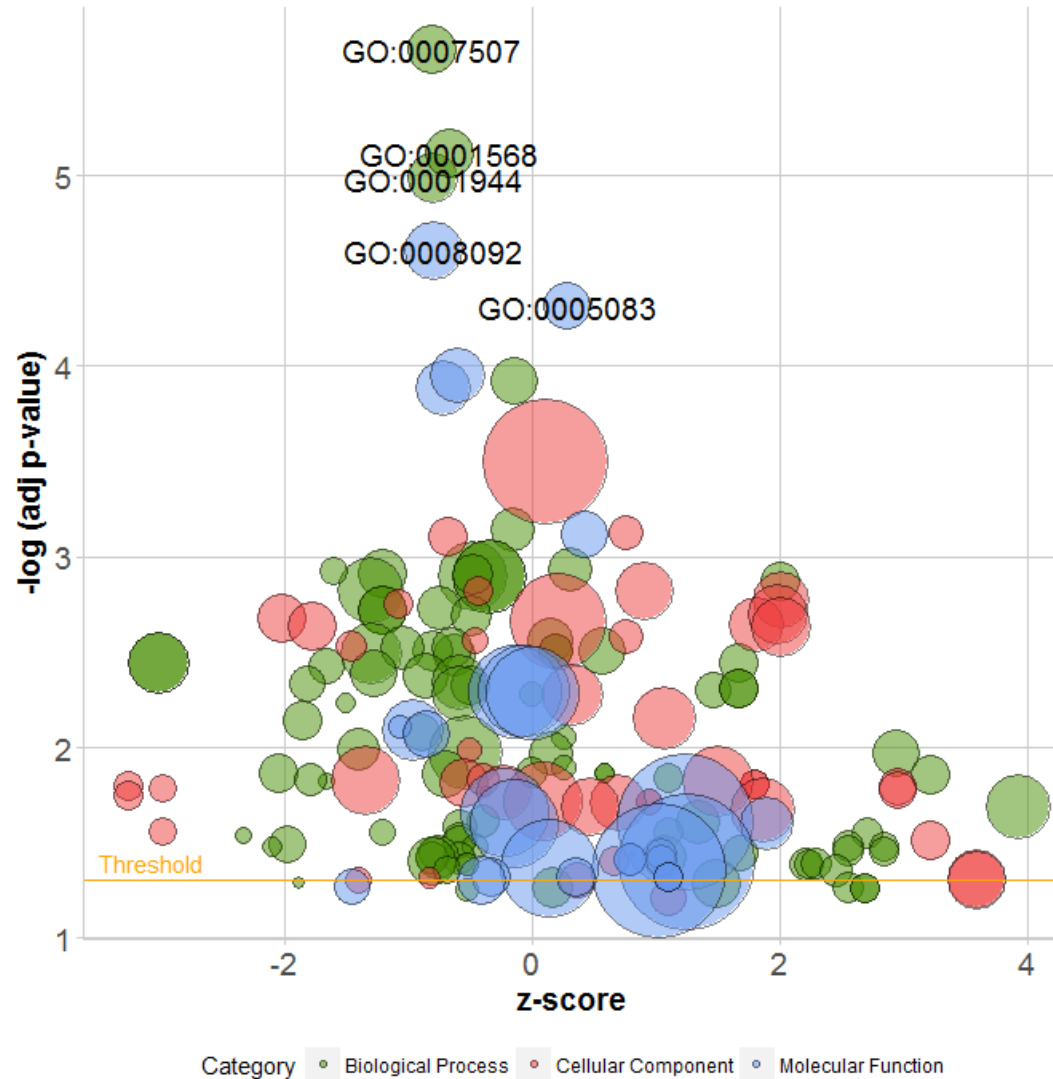
Tables are a natural way to present the key data

Name	P-value	Adjusted p-value	Z-score	Combined score
glycogen biosynthetic process (GO:0005978)	0.004329	0.5192	-3.53	2.31
maturation of 5.8S rRNA (GO:0000460)	0.02044	0.5192	-2.83	1.85
mitochondrion organization (GO:0007005)	0.007428	0.5192	-2.82	1.85
mitochondrial calcium ion transmembrane transport (GO:0006851)	0.05335	0.5192	-2.57	1.69
generation of precursor metabolites and energy (GO:0006091)	0.001993	0.5192	-2.56	1.68
retinal metabolic process (GO:0042574)	0.01428	0.5192	-2.44	1.60
protein polyubiquitination (GO:0000209)	0.1031	0.5192	-2.32	1.52
protein homotetramerization (GO:0051289)	0.01521	0.5192	-2.28	1.49

# Graphical Representations

- Need to add something over a table
  - Relationships between multiple result values
  - Representation of redundancy between categories
  - Relationship to original data
  - Context of surrounding pathway

# Plotting relationships between values



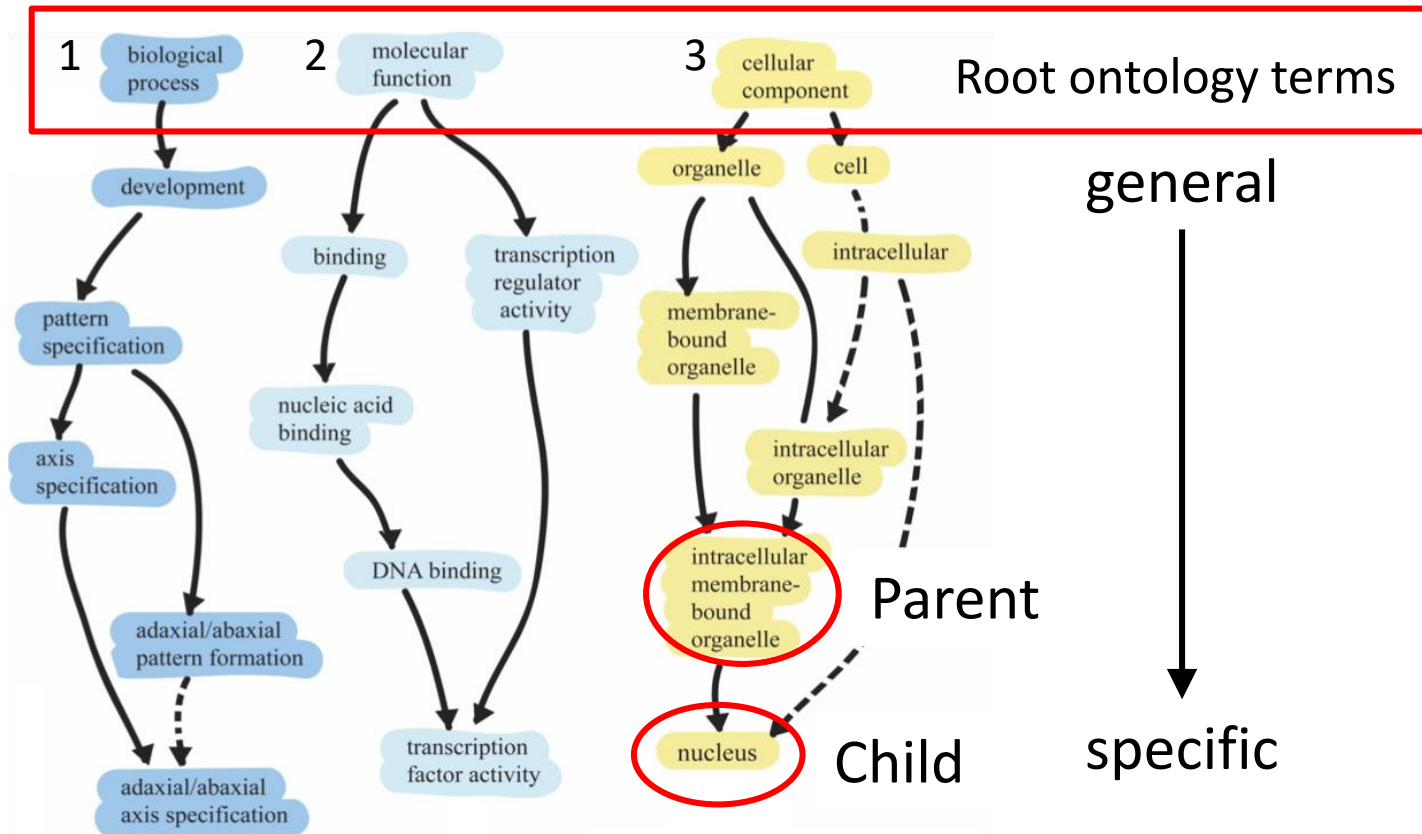
- P-value (y)
- Enrichment (x)
- Count (radius)
- Redundancy (colour)

Goplot

<http://wencke.github.io/>

# Redundancy in gene lists

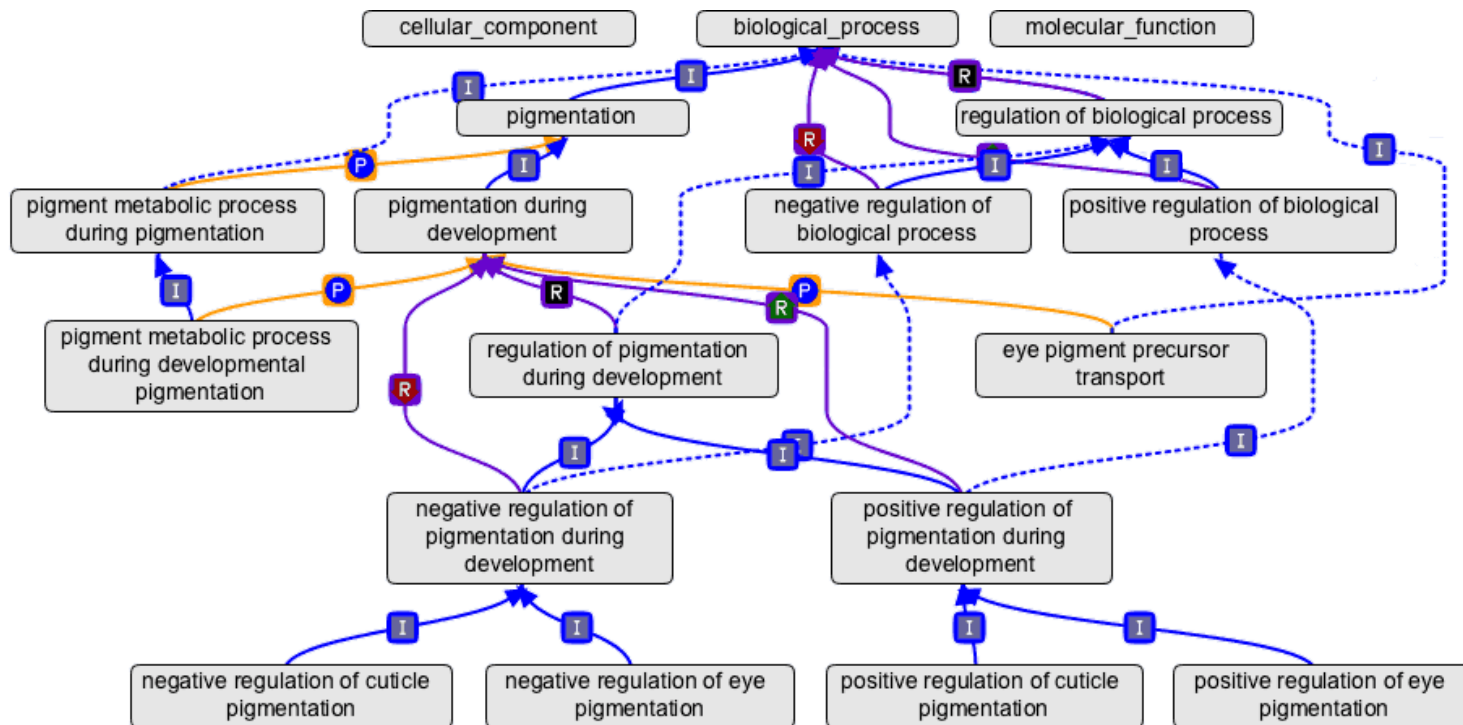
Gene ontology is hierarchical - a gene is placed in the most specific category and will also appear in all the parent categories



# Redundancy in gene lists

Most genes are associated with multiple GO terms

There are many annotation sources, not just gene ontologies



Ontology structure is more complex than just parent-child relationships

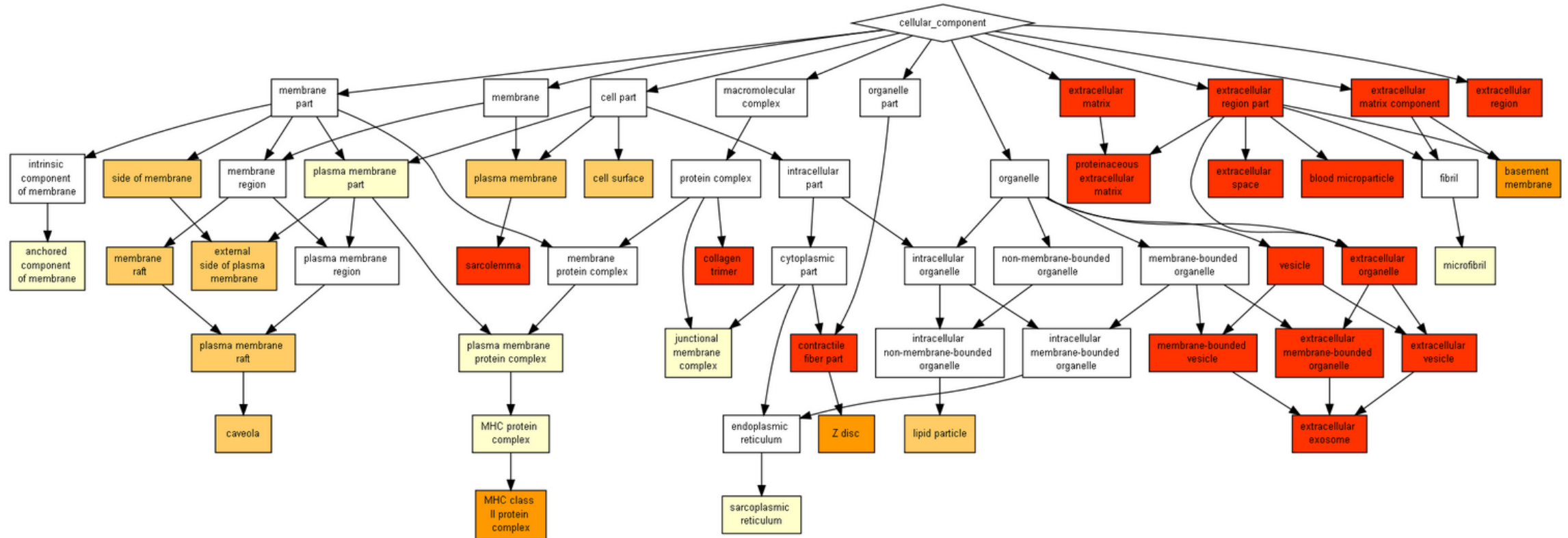


# Redundancy: DAVID clustering

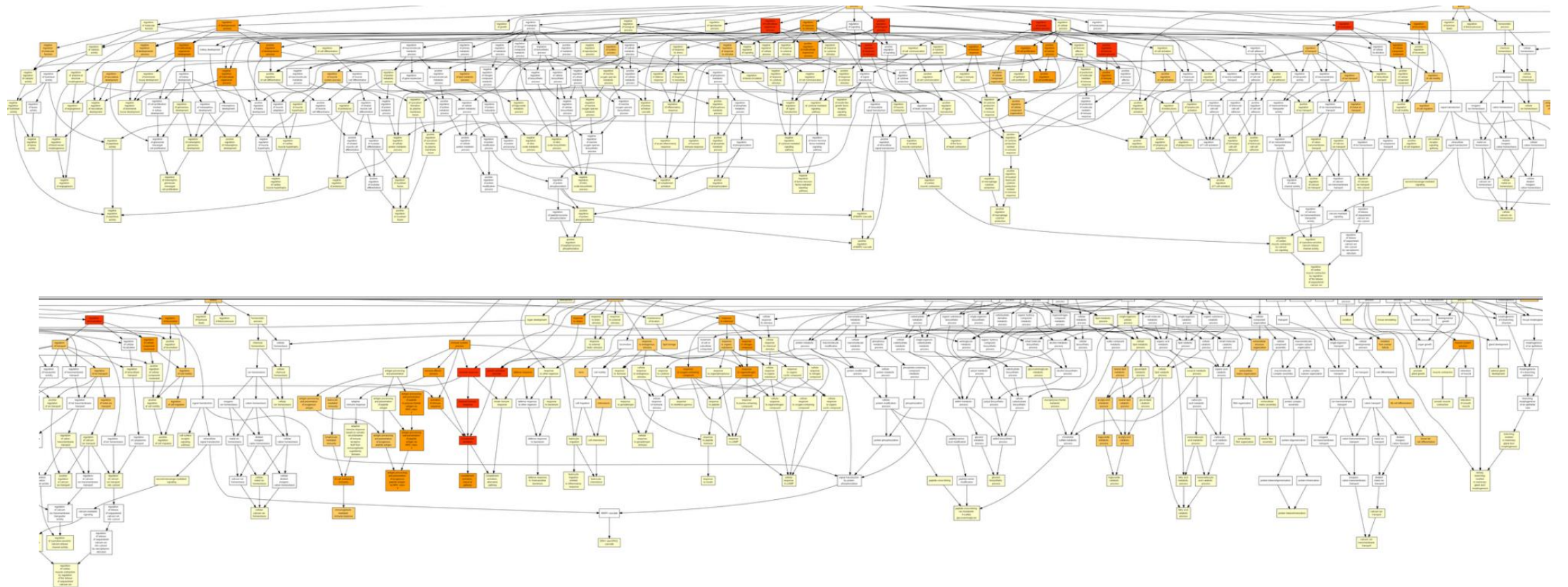
40 Cluster(s) [Download File](#)

Annotation Cluster 1		Enrichment Score: 16.36			Count	P_Value	Benjamini
<input type="checkbox"/>	SP_PIR_KEYWORDS	<a href="#">dna-binding</a>	RT		53	3.5E-24	4.5E-22
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">regulation of transcription</a>	RT		60	2.0E-20	1.8E-17
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">DNA binding</a>	RT		54	6.0E-20	7.9E-18
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">transcription regulator activity</a>	RT		45	2.0E-19	1.3E-17
<input type="checkbox"/>	SP_PIR_KEYWORDS	<a href="#">transcription regulation</a>	RT		49	5.9E-19	3.8E-17
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">sequence-specific DNA binding</a>	RT		30	1.5E-16	4.9E-15
<input type="checkbox"/>	SP_PIR_KEYWORDS	<a href="#">Transcription</a>	RT		48	8.1E-16	3.3E-14
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">transcription</a>	RT		48	1.9E-15	8.1E-13
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">transcription factor activity</a>	RT		33	2.8E-15	9.1E-14
<input type="checkbox"/>	SP_PIR_KEYWORDS	<a href="#">nucleus</a>	RT		69	1.1E-14	3.6E-13
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">regulation of RNA metabolic process</a>	RT		40	2.1E-12	6.1E-10
<input type="checkbox"/>	GOTERM_BP_FAT	<a href="#">regulation of transcription, DNA-dependent</a>	RT		39	6.4E-12	1.4E-9
Annotation Cluster 2		Enrichment Score: 10.03			Count	P_Value	Benjamini
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">sequence-specific DNA binding</a>	RT		30	1.5E-16	4.9E-15
<input type="checkbox"/>	GOTERM_MF_FAT	<a href="#">transcription factor activity</a>	RT		33	2.8E-15	9.1E-14
<input type="checkbox"/>	INTERPRO	<a href="#">Homeodomain-related</a>	RT		16	2.3E-10	4.3E-8
<input type="checkbox"/>	INTERPRO	<a href="#">Homeobox</a>	RT		15	1.8E-9	1.7E-7
<input type="checkbox"/>	INTERPRO	<a href="#">Homeobox, conserved site</a>	RT		14	3.4E-9	2.1E-7
<input type="checkbox"/>	SP_PIR_KEYWORDS	<a href="#">Homeobox</a>	RT		15	8.5E-9	1.8E-7
<input type="checkbox"/>	UP_SEQ_FEATURE	DNA-binding region:Homeobox	RT		13	2.6E-8	3.7E-6
<input type="checkbox"/>	SMART	<a href="#">HOX</a>	RT		15	4.7E-8	2.1E-6
Annotation Cluster 3		Enrichment Score: 5.86			Count	P_Value	Benjamini
<input type="checkbox"/>	INTERPRO	<a href="#">Transcription factor, fork head, conserved site</a>	RT		7	3.6E-7	1.7E-5
<input type="checkbox"/>	INTERPRO	<a href="#">Transcription factor, fork head</a>	RT		7	3.6E-7	1.7E-5
<input type="checkbox"/>	UP_SEQ_FEATURE	DNA-binding region:Fork-head	RT		7	9.1E-7	6.5E-5
<input type="checkbox"/>	SMART	<a href="#">FH</a>	RT		7	1.8E-6	4.0E-5
<input type="checkbox"/>	INTERPRO	<a href="#">Winged helix repressor DNA-binding</a>	RT		9	2.5E-5	6.6E-4

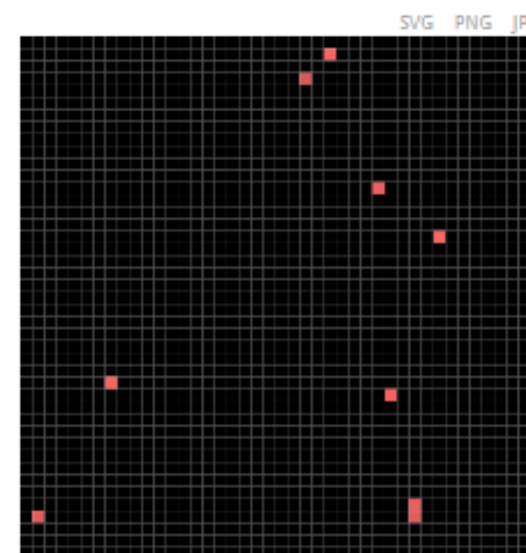
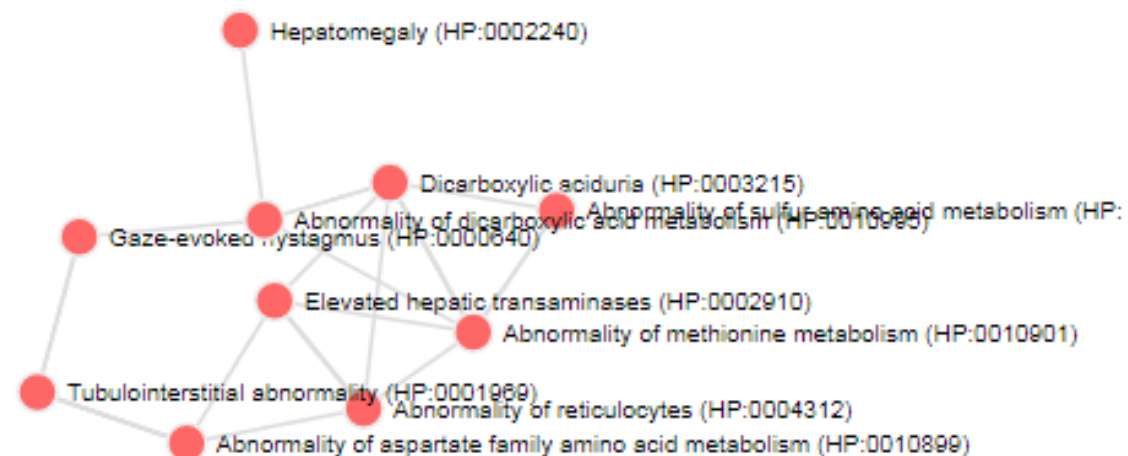
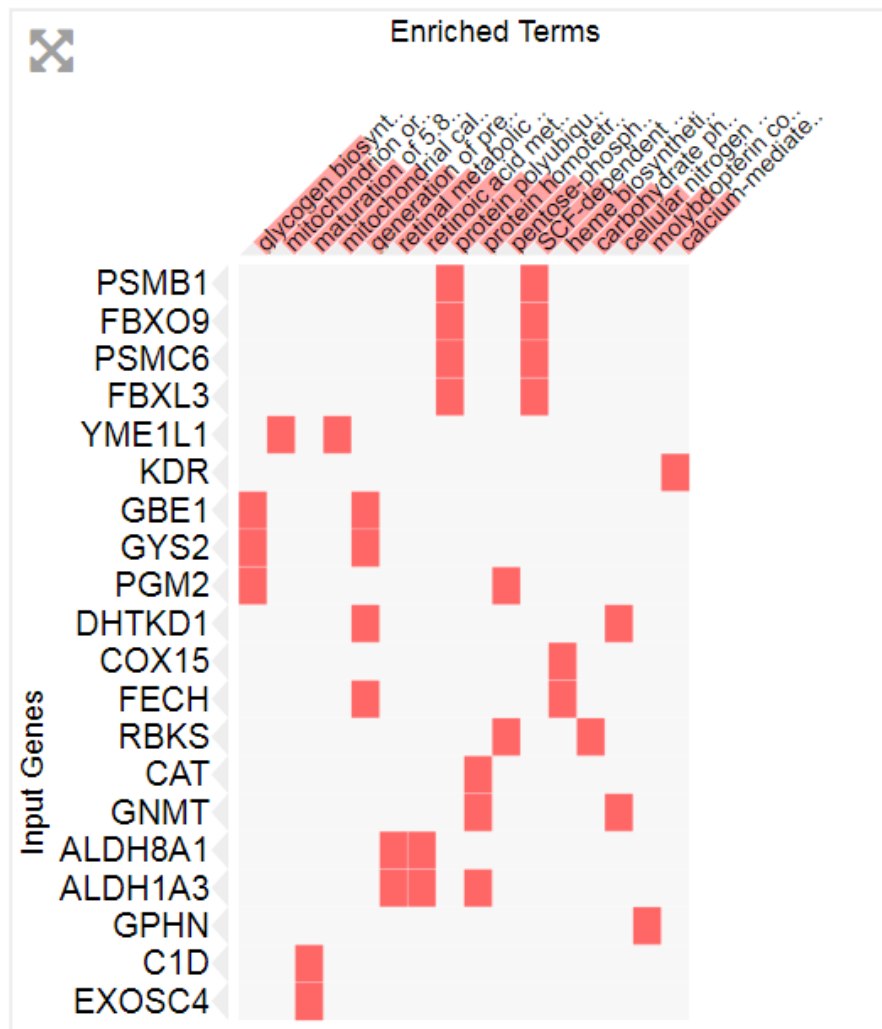
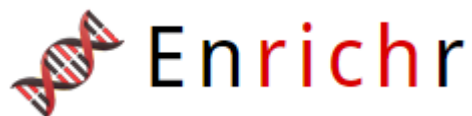
# Redundancy: Gorilla GO images



# Redundancy: Gorilla GO images



# Redundancy: Enrichr overlap plots



Z-score: -0.3496  
P-value: 0.3633

# Redundancy: REVIGO (from Gorilla)



<http://revigo.irb.hr/>

[Hide/show dispensable GO terms](#)

[Export results to text table \(CSV\)](#)

[Make R script for plotting](#)

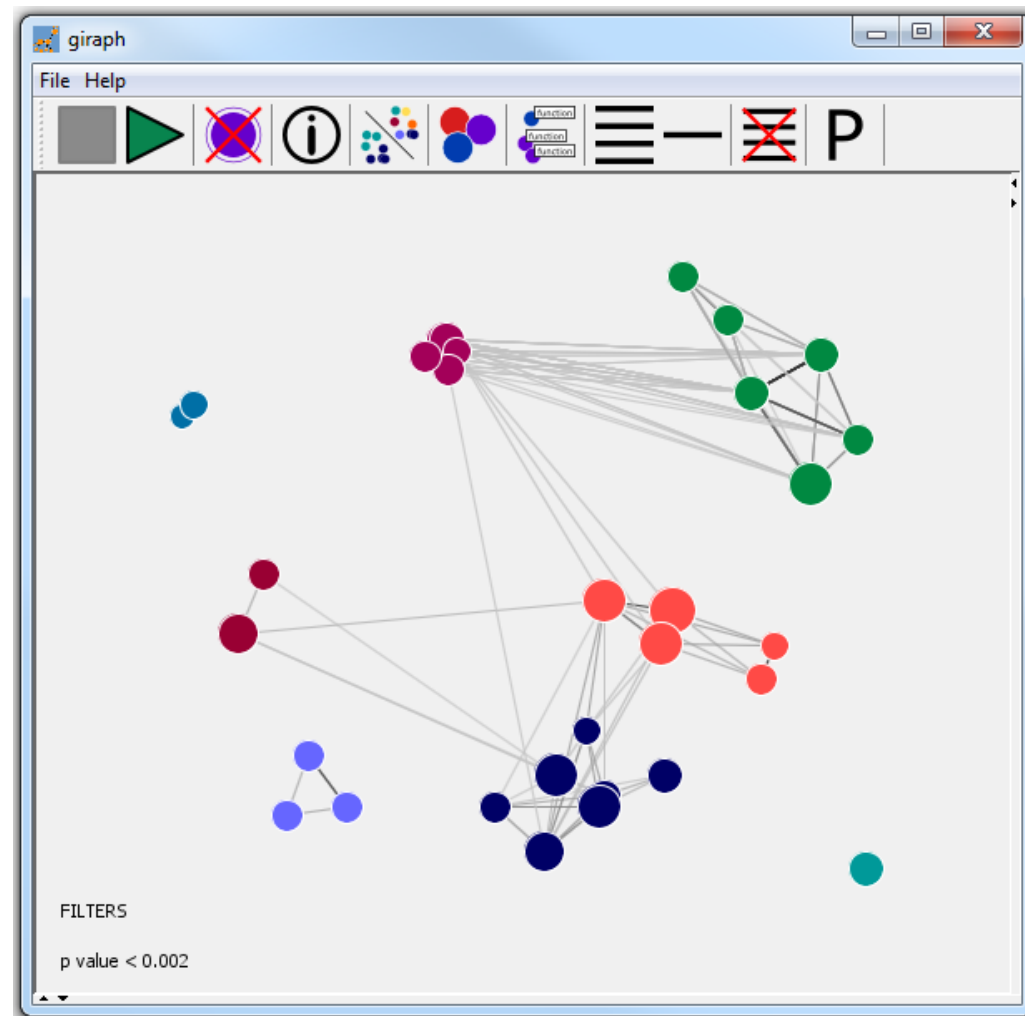
term ID	description	frequency	pin?	log <sub>10</sub> p-value	uniqueness	dispensability
<a href="#">GO:0007155</a>	cell adhesion	0.564 %		-7.1618	0.99	0.00
<a href="#">GO:0008150</a>	biological_process	100.000 %		-3.6162	1.00	0.00
<a href="#">GO:0022610</a>	biological adhesion	1.210 %		-7.5986	0.99	0.00
<a href="#">GO:0030198</a>	extracellular matrix organization	0.015 %		-18.6162	0.91	0.00
<a href="#">GO:0043206</a>	extracellular fibril organization	0.000 %	<input type="checkbox"/>	-3.1124	0.92	0.83
<a href="#">GO:0022617</a>	extracellular matrix disassembly	0.001 %	<input type="checkbox"/>	-8.4921	0.92	0.88
<a href="#">GO:0048251</a>	elastic fiber assembly	0.000 %	<input type="checkbox"/>	-3.8861	0.92	0.83
<a href="#">GO:0085029</a>	extracellular matrix assembly	0.001 %	<input type="checkbox"/>	-5.2993	0.92	0.87
<a href="#">GO:0032501</a>	multicellular organismal process	0.790 %		-9.7595	0.99	0.00
<a href="#">GO:0032502</a>	developmental process	1.387 %		-4.0458	0.99	0.00
<a href="#">GO:0044699</a>	single-organism process	53.975 %		-4.9066	1.00	0.00
<a href="#">GO:0048511</a>	rhythmic process	0.020 %		-3.6757	0.99	0.00

# Redundancy: Giraph (via David)

## Giraph plot

Proximity of circles is related to the proportion of overlapping genes between categories.

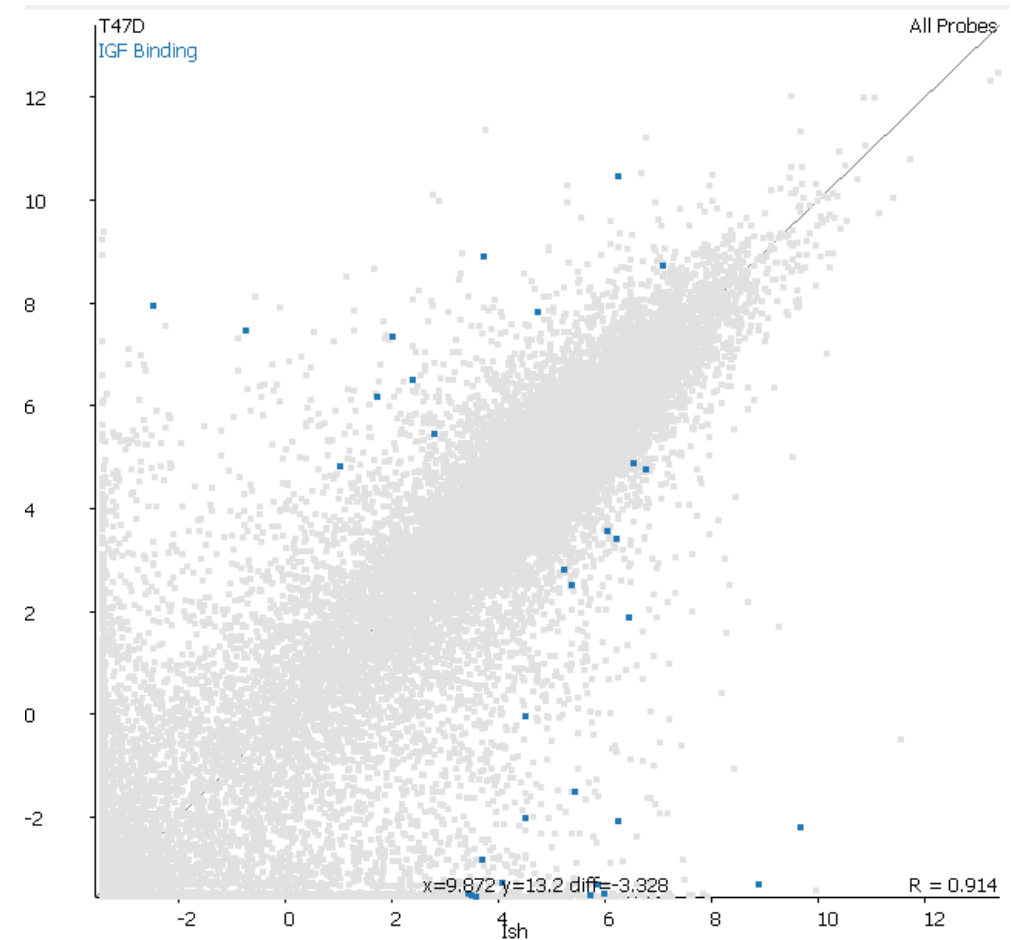
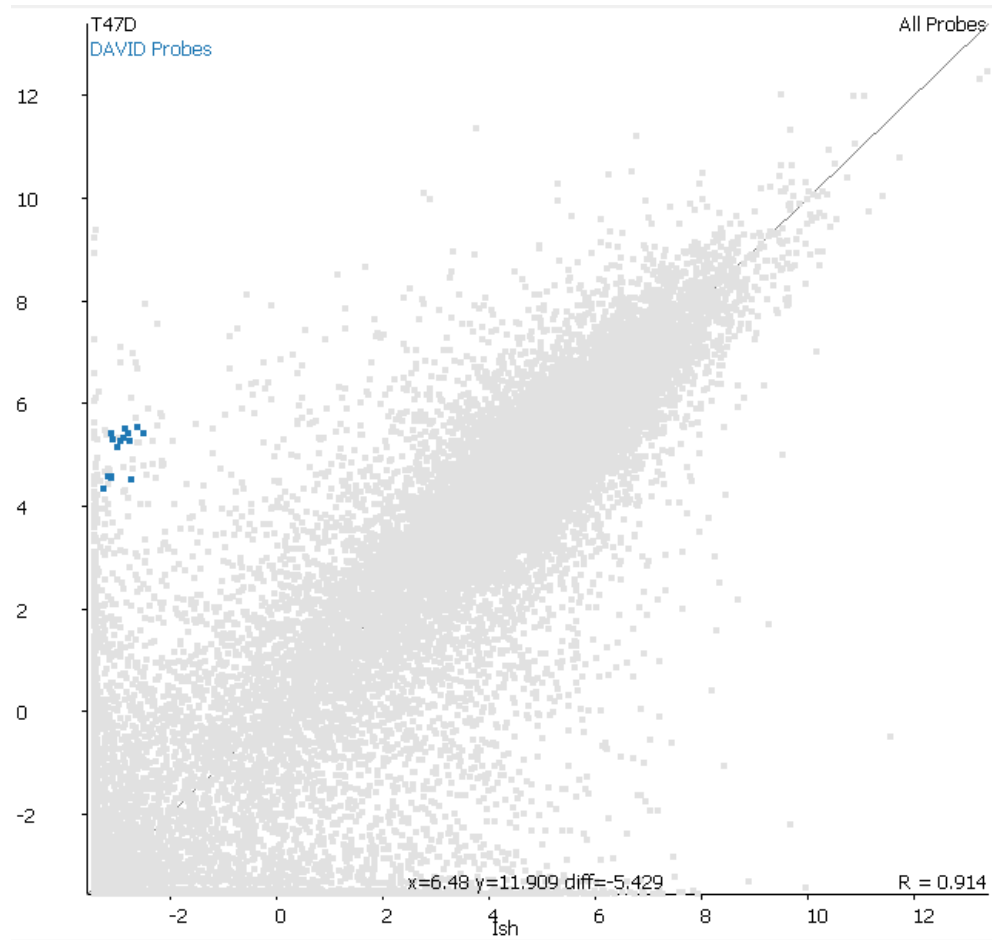
Size of circle represents the number of genes.



# Relationship to original data

- Quantitative values for genes in category
  - Direction and magnitude of change
- Look at genes in category which aren't hits
  - Relative numbers
  - Supportive changes?

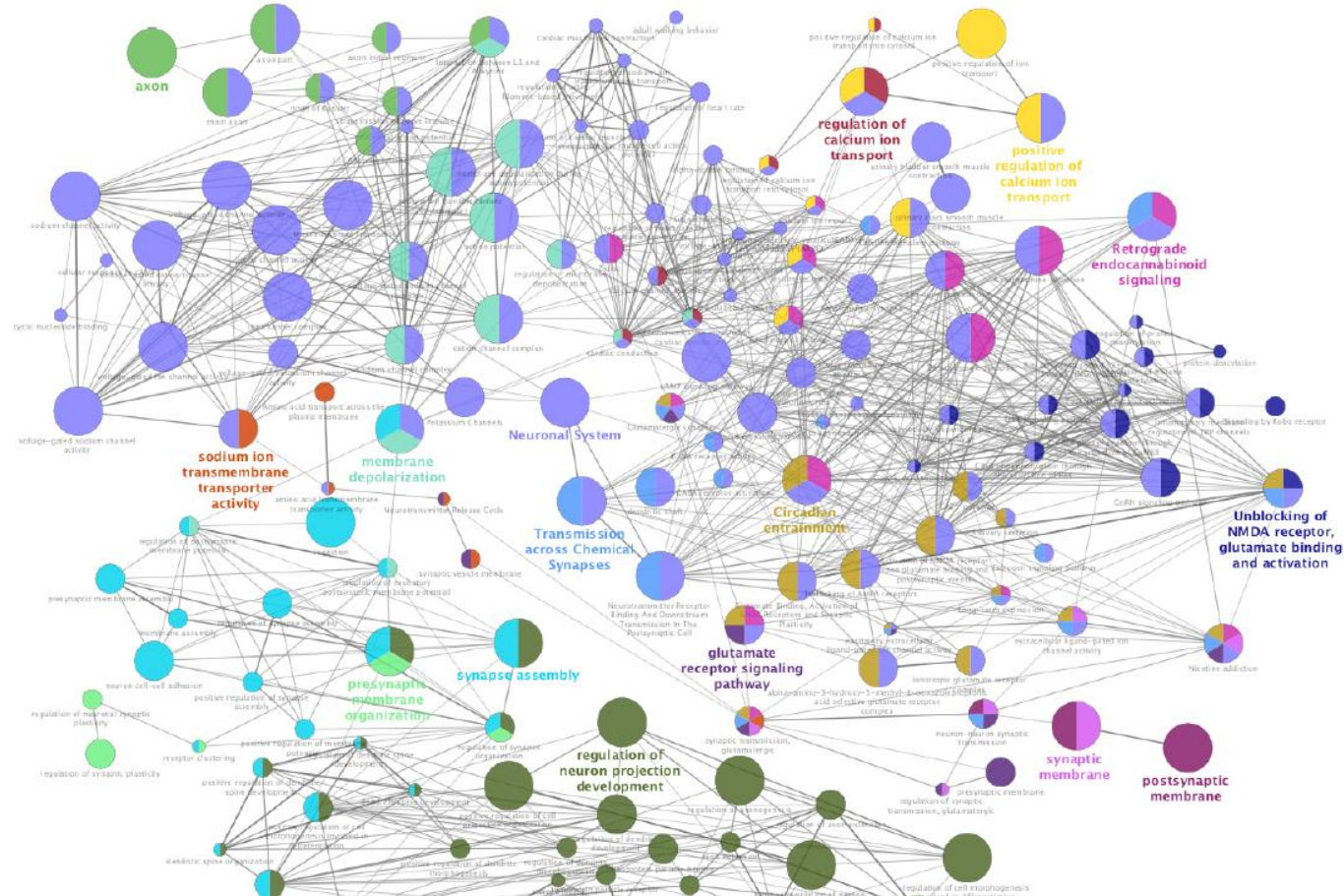
# Relationship to original data



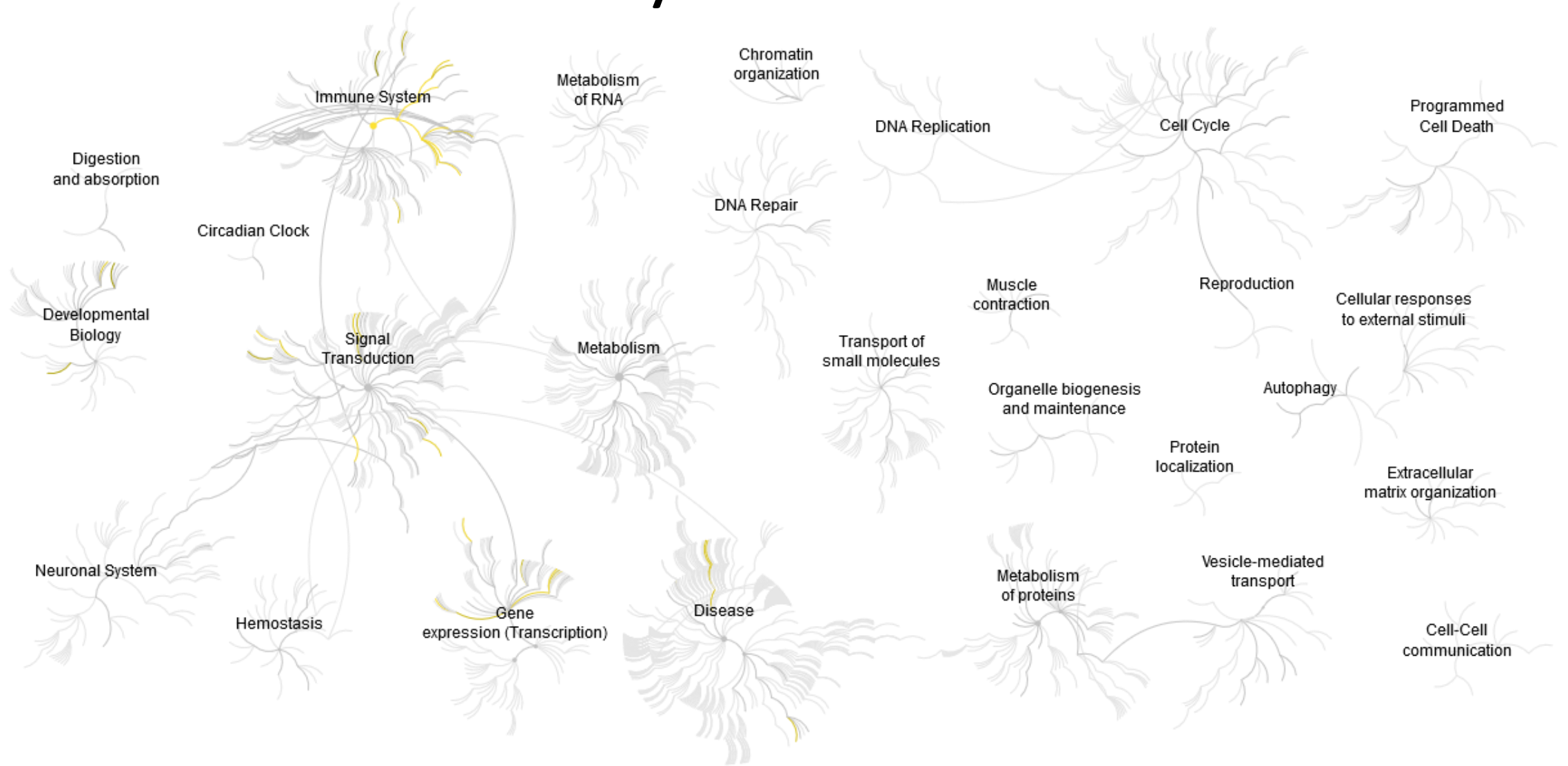


# Pathways: ClueGo

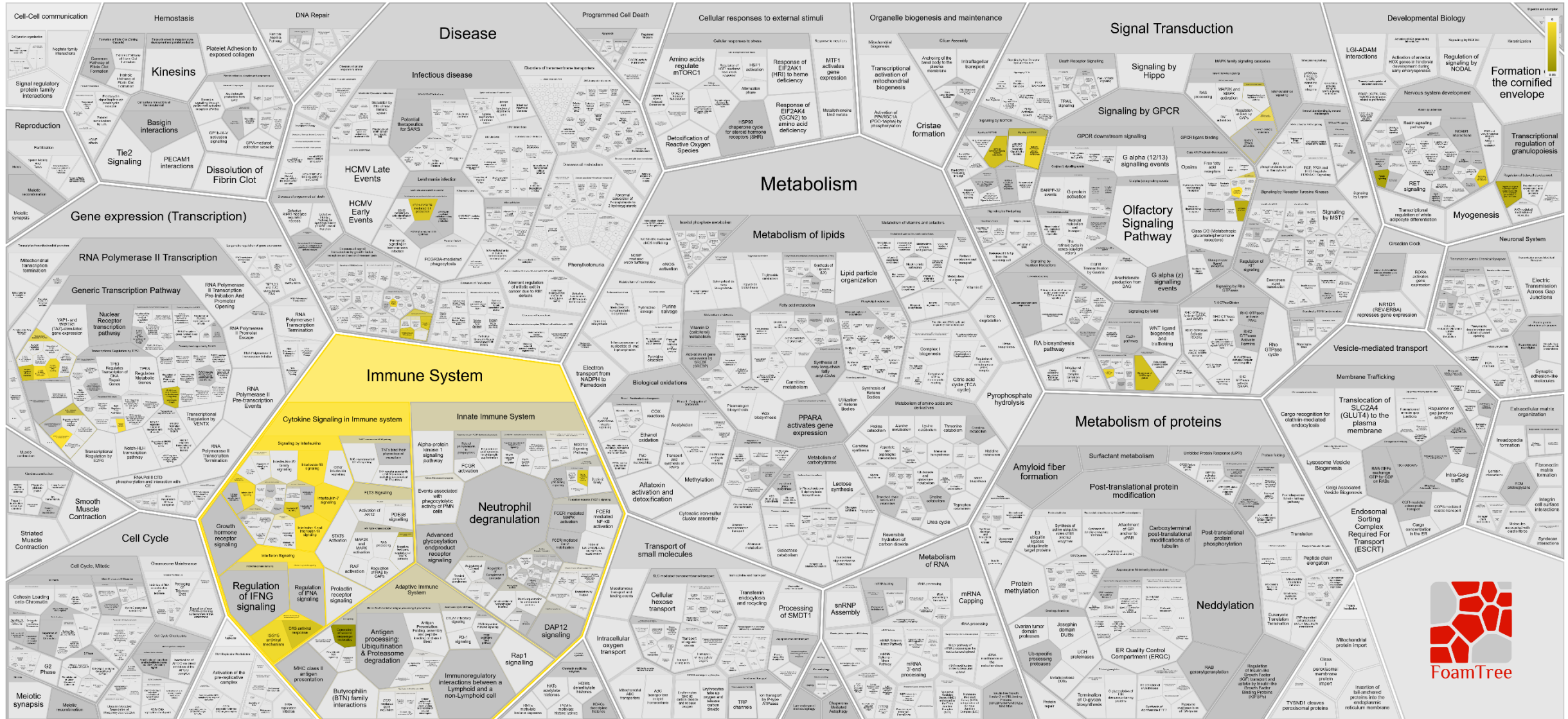
- App within Cytoscape
- ClueGO integrates GO terms as well as pathways
- Creates a functionally organized GO/pathway term network



# Pathways: Reactome



# Pathways: Reactome



FoamTree

# Summary

- Tables are often sufficient
  - Must include name, enrichment, corrected p-value
  - Other values are useful, but don't put in everything
- Figures can add extra information
  - Plotting multiple metrics
  - Illustrating redundancy
  - Relating to original data
  - Mapping to pathways