Functional Enrichment Analysis & Candidate Gene Ranking

Anil Jegga
Biomedical Informatics

Contact Information:
Anil Jegga
Biomedical Informatics
Room # 232, S Building 10th Floor
CCHMC
Homepage: http://anil.cchmc.org
Tel: 513-636-0261
E-mail: anil.jegga@cchmc.org

Slides and Example data sets available for download at:

http://anil.cchmc.org/dhc.html

Workshop Evaluation: Please provide your valuable feedback on the evaluation sheet provided along with the hand-outs

This workshop is about the analysis of transcriptome (identifying enriched biological processes, etc.) and ranking or prioritizing candidate genes. It **does not** cover microarray data analysis.

Contact Huan Xu (huan.xu@cchmc.org for GeneSpring related questions or microarray data analysis.

All the applications/servers/databases used in this workshop are **free** for academic-use. Applications that are not free for use (e.g., Ingenuity Pathway Analysis, etc.) are not covered here. However, we have licensed access to use some of these and please contact us if you are interested in using them.

What are we going to cover today?

1. Gene List Functional Enrichment Analysis
2. Multiple Gene Lists Functional Enrichment Analysis
3. Prioritizing or Ranking Candidate Genes
    - Based on functional annotations
    - Based on network connectivity

ToppGene Suite: http://toppgene.cchmc.org
ToppCluster: http://toppcluster.cchmc.org

Related Publications

(for methodology- and validation-related details)

ToppGene Suite

1. Chen J, Xu H, Aronow BJ, Jegga AG. 2007. Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics* 8:392.
2. Chen J, Aronow BJ, Jegga AG. 2009. Disease candidate gene identification and prioritization using protein interaction networks. *BMC Bioinformatics* 10:73.
3. Chen J, Bardes EE, Aronow BJ, Jegga AG 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research* doi: 10.1093/nar/gkp427.

ToppCluster

1. Kaimal V, Bardes E, Jegga AG, Aronow BJ. 2010. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Research* (*in press*).

I have a list of co-expressed mRNAs (Transcriptome)….

Now what?

1. **Identify putative shared regulatory elements**

- Known transcription factor binding sites (TFBS)
  - Conserved
  - Non-conserved
- Unknown TFBS or Novel motifs
  - Conserved
  - Non-conserved
- MicroRNAs

2. **Identify the underlying biological theme**

- Gene Ontology
- Pathways
- Phenotype/Disease Association
- Protein Domains
- Protein Interactions
- Expression in other tissues/experiments
- Drug targets
- Literature co-citation…

Annotation Databases
Gene Ontology, Pathways

Genome-wide Promoters
Putative Regulatory
Signatures

DNA Repair
XRCC1
OGG1
ERCC1
MPG….

Angiogenesis
HIF1A
ANGPT1
VEGF
KLF5….

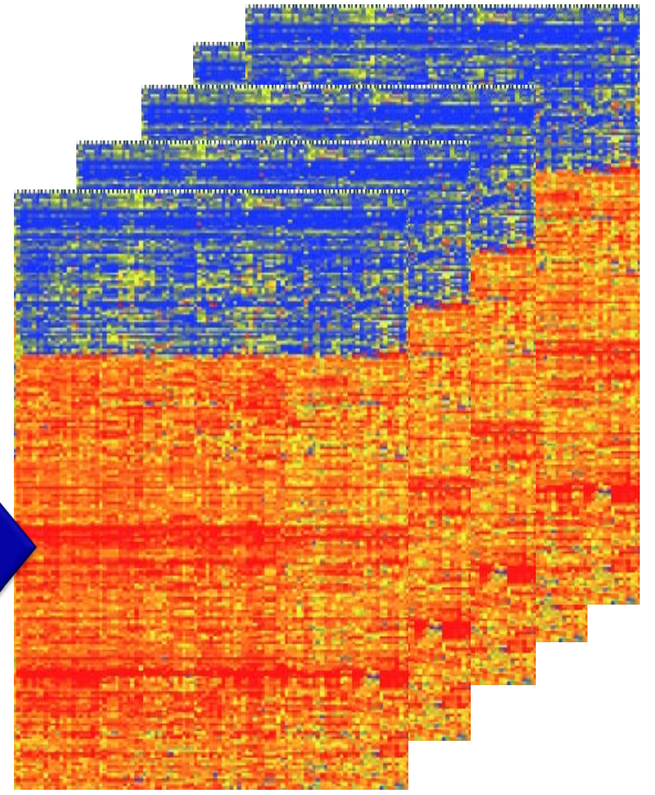
E2F
RB1
MCM4
FOS
SIVA…..

PDX1
GLUT2
PAX4
PDX1
IAPP….

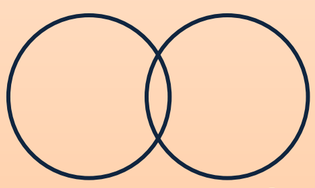
p53
CDKN1A
CTSD
CASP
DDB2….

Gene lists associated with similar
function/process/pathway

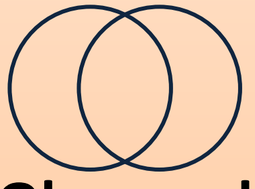
Expression Profile - Gene Lists



Enrichment Analysis



Expected

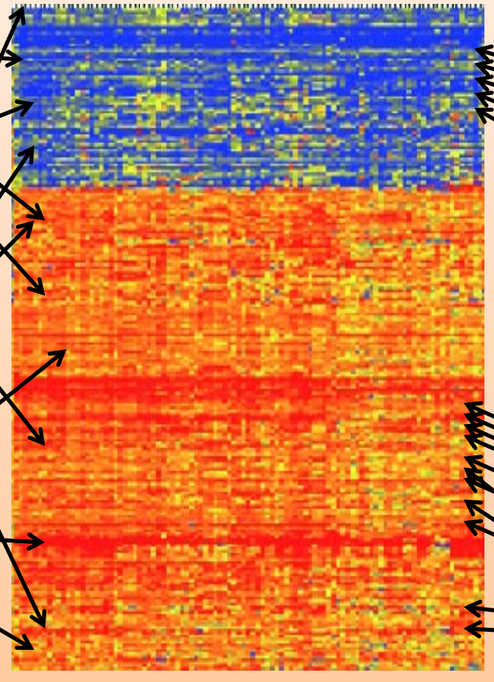


Observed

P53
CTSD
CASP
DDB2….

DNA Repair
XRCC1
OGG1
ERCC1
MPG….

Random Distribution



E2F
RB1
MCM4
FOS…

Angiogenesis
HIF1A
ANGPT1
VEGF…..

Significant Enrichment

I have a list of co-expressed mRNAs (Transcriptome)….

I want to find the shared cis-elements – Known and Novel

- Known transcription factor binding sites (TFBS)
  - Conserved
    - oPOSSUM
    - DiRE
  - Non-conserved
    - Pscan
    - MatInspector (*Licensed)
- Unknown TFBS or Novel motifs
  - Conserved
    - oPOSSUM
    - Weeder-H
  - Non-conserved
    - MEME
    - Weeder

1. Each of these applications support different forms of input. Very few support probeset IDs.
2. **Red Font**: Input sequence required; Do not support gene symbols, gene IDs, or accession numbers. The advantage is you can use them for scanning sequences from any species.
3. *Licensed software: We have access to the licensed version.

- **Covered in the last workshop (Sept. 2009).**
- **Will not be covered today.**
- **Training material is available on-line.**

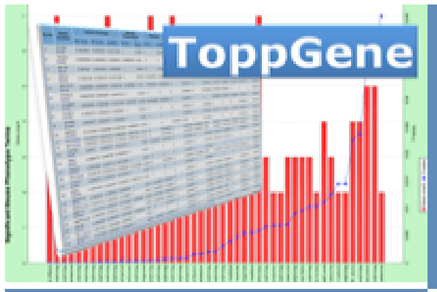
I have a list of co-expressed mRNAs (Transcriptome)….

Identify the underlying biological theme

What are my genes "enriched" for?

- Gene Ontology
- Pathways
- Phenotype/Disease Association
- Protein Domains
- TFBS and microRNA
- Protein Interactions
- Expression in other tissues/experiments
- Drug targets
- Literature co-citation…

ToppGene Suite (http://toppgene.cchmc.org)



- Home
- Links
- Database details
- Supplementary
- Help
- Publications
- Terms of Use
- Contacts

Supported by:

Computational Medicine Center

Cincinnati DHC Digestive Health Center

ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

- **ToppFun**: Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

  Detect functional enrichment of your gene list based on Transcriptome, Proteome, Regulome (TFBS and miRNA), Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), Pharmacome (Drug-Gene associations), literature co-citation, and other features.

- **ToppGene**: Candidate gene prioritization

  Prioritize or rank candidate genes based on functional similarity to training gene list.

- **ToppNet**: Relative importance of candidate genes in networks

  Prioritize or rank candidate genes based on topological features in protein-protein interaction network.

- **ToppGenet**: Prioritization of neighboring genes in protein-protein interaction network

  Identify and prioritize the neighboring genes of the seeds in protein-protein interaction network based on functional similarity to the "seed" list (ToppGene) or topological features in protein-protein interaction network (ToppNet).

1. Free for use, no log-in required.
2. Web-based, no need to install anything (except for applications to visualize or analyze networks)
3. Validated and published

ToppGene Suite (http://toppgene.cchmc.org) - ToppFun

ToppFun: Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

Select your gene identifier type, paste your sets below or select example set, then submit.

Entry Type: Entrez ID

Example gene sets: HGNC Symbol Entrez ID
(click on "HGNC Symbol" or "Entrez ID" to use the example training and test set of genes)

Training Gene Set:

```
259
5265
350
335
335
1558
1571
229
462
125
3240
5105
5265
3273
2244
2158
5053
125
1356
3827
383
```

Clear | Submit Query

1. Supports variety of inputs
2. Supports symbol correction
3. Eliminates any duplicates
4. Drawback: Supports human and mouse genes only

Input Gene List (81 / 97)

Entered	Human Symbol	Gene ID
259	AMBP	259
5265	SERPINA1	5265
350	APOH	350
335	APOA1	335
1558	CYP2C8	1558
1571	CYP2E1	1571
229	ALDOB	229
462	SERPINC1	462
125	ADH1B	125
3240	HP	3240
5105	PCK1	5105
3273	HRG	3273
2244	FGB	2244
2158	F9	2158
5053	PAH	5053
1356	CP	1356
3827	KNG1	3827
383	ARG1	383
5004	ORM1	5004
2168	FABP1	2168
325	APCS	325

Genes Not Found

Entered	Status
335	Duplicated
5265	Duplicated
125	Duplicated
1571	Duplicated
1571	Duplicated
1373	Duplicated
1356	Duplicated
462	Duplicated
2243	Duplicated
3827	Duplicated
125	Duplicated
229	Duplicated
6822	Duplicated
2328	Duplicated

ToppGene Suite (http://toppgene.cchmc.org) - ToppFun

Calculations

Feature	Correction	p-Value cutoff	Gene Limits		
All	Bonferroni	0.05	1	≤ n ≤	1500
GO: Molecular Function	Bonferroni	0.05	1	≤ n ≤	1500
GO: Biological Process	Bonferroni	0.05	1	≤ n ≤	1500
GO: Cellular Component	Bonferroni	0.05	1	≤ n ≤	1500
Human Phenotype	Bonferroni	0.05	1	≤ n ≤	1500
Mouse Phenotype	Bonferroni	0.05	1	≤ n ≤	1500
Domain	Bonferroni	0.05	1	≤ n ≤	1500
Pathway	Bonferroni	0.05	1	≤ n ≤	1500
Pubmed	Bonferroni	0.05	1	≤ n ≤	1500
Interaction	Bonferroni	0.05	1	≤ n ≤	1500
Cytoband	Bonferroni	0.05	1	≤ n ≤	1500
TFBS	Bonferroni	0.05	1	≤ n ≤	1500
Gene Family	Bonferroni	0.05	1	≤ n ≤	1500
Coexpression	Bonferroni	0.05	1	≤ n ≤	1500
Computational	Bonferroni	0.05	1	≤ n ≤	1500
MicroRNA	Bonferroni	0.05	1	≤ n ≤	1500
Drug	Bonferroni	0.05	1	≤ n ≤	1500
Disease	Bonferroni	0.05	1	≤ n ≤	1500

Home | Modify Query | Submit

1. Gene list analyzed for as many as 17 features!
2. Single-stop enrichment analysis server for both regulatory elements (TFBSs and miRNA) and biological themes
3. Back-end has an exhaustive, normalized data resources compiled and integrated
4. Bonferroni correction is “too stringent”; FDR with 0.05 is preferable.
5. TFBS are based on conserved cis-elements and motifs within ±2kb region of TSS in human, mouse, rat, and dog.
6. miRNA-targets are based on TargetScan, PicTar and miRrecords/Tarbase.

ToppGene Suite (http://toppgene.cchmc.org)

GO Biological Process

Annotations:	16,372
Genes:	15,079
	Updated Aug 26, 2009

GO Cellular Component

Annotations:	2,335
Genes:	16,728
	Updated Aug 26, 2009

GO Molecular Function

Annotations:	8,583
Genes:	15,948
	Updated Aug 26, 2009

Human Phenotype

Annotations:	9,551
Genes:	2,531
	Updated Sep 10, 2009

Mouse Phenotype

Annotations:	6,203
Genes:	5,590
	Updated Aug 25, 2009

1. Database updated regularly
2. Exhaustive collection of annotations

Pathways

Annotations:		1,672
	BioCyc	164
Aug 25, 2009	CGAP BioCarta	314
	GenMAPP	67
Jun 15, 2009	KEGG pathway	202
May 10, 2009	MSigDB	431
	PantherDB	150
Aug 25, 2009	Pathway Ontology	306
	Reactome	25
	SigmaAldrich	2
	Signalling Transduction KE	11
Genes:		6,697

Domains

Annotations:		10,223
	Gene3D	285
	InterPro	4,859
	PROSITE	1,351
	Pfam	2,774
	ProDom	385
	SMART	569
Genes:		12,430

Pubmed

Annotations:	221,282
Genes:	22,176
	Updated Aug 25, 2009

Interactions

Annotations:		18,047
	BIND	4,370
	BioGRID	7,602
	HPRD	6,075
Genes:		5,541

Cytoband

Annotations:	382
Genes:	29,821

TFBS

Annotations:	615
Genes:	9,770

miRNA

Annotations:		740
	MSigDB	313
	PicTar	178
	TargetScan	249
Genes:		11,618

Gene Families

Annotations:	151
Genes:	6,098

Coexpression

Annotations:		1,203
	Body Map	23
	mSigDB	1,180
Genes:		12,694

Computational Gene Set

Annotations:	427
Genes:	4,712

Drugs

Annotations:		13,141
Aug 28, 2009	CTD	4,977
Aug 25, 2009	Drug Bank	2,009
Aug 25, 2009	Stitch	6,155
Genes:		14,836

Disease

Annotations:		3,789
Aug 28, 2009	CTD	1,006
	GWAS	291
	OMIM	2,492
Genes:		4,385

Master Gene Info File

For All Annotations	35,449
	Updated Aug 28, 2009

ToppGene Suite (http://toppgene.cchmc.org) - ToppFun

Results

Go To Start Page

Input Parameters [Show Detail]

Training Results [Show All] [Download All] [Sparse Matrix]

1: GO: Molecular Function [Display Chart] [Show Detail]

2: GO: Biological Process [Display Chart] [Show Detail]

3: GO: Cellular Component [Display Chart] [Show Detail]

4: Human Phenotype [Display Chart] [Show Detail]

5: Mouse Phenotype [Display Chart] [Show Detail]

6: Domain [Display Chart] [Show Detail]

7: Pathway [Display Chart] [Show Detail]

8: Pubmed [Display Chart] [Show Detail]

9: Interaction [Display Chart] [Show Detail]

10: Cytoband [Display Chart] [Show Detail]

11: TFBS [Display Chart] [Show Detail]

12: Gene Family [Display Chart] [Show Detail]

13: Coexpression [Display Chart] [Show Detail]

14: Computational [Display Chart] [Show Detail]

15: MicroRNA [Display Chart] [Show Detail]

16: Drug [Display Chart] [Show Detail]

17: Disease [Display Chart] [Show Detail]

Input Parameters [Hide Detail]

Number of genes in training set:	81
Number of genes in test set:	0
Correction and Cutoff:	

category	Correction	Cutoff	Min	Max
GO: Molecular Function	Bonferroni	0.05	1	1500
GO: Biological Process	Bonferroni	0.05	1	1500
GO: Cellular Component	Bonferroni	0.05	1	1500
Human Phenotype	Bonferroni	0.05	1	1500
Mouse Phenotype	Bonferroni	0.05	1	1500
Domain	Bonferroni	0.05	1	1500
Pathway	Bonferroni	0.05	1	1500
Pubmed	Bonferroni	0.05	1	1500
Interaction	Bonferroni	0.05	1	1500
Cytoband	Bonferroni	0.05	1	1500
TFBS	Bonferroni	0.05	1	1500
Gene Family	Bonferroni	0.05	1	1500
Coexpression	Bonferroni	0.05	1	1500
Computational	Bonferroni	0.05	1	1500
MicroRNA	Bonferroni	0.05	1	1500
Drug	Bonferroni	0.05	1	1500
Disease	Bonferroni	0.05	1	1500

Random sampling size in analysis:	0
Minimun feature count in test set:	2
Analysis took:	2 seconds
Analysis finished at:	Sun Sep 27 16:45:06 EDT 2009

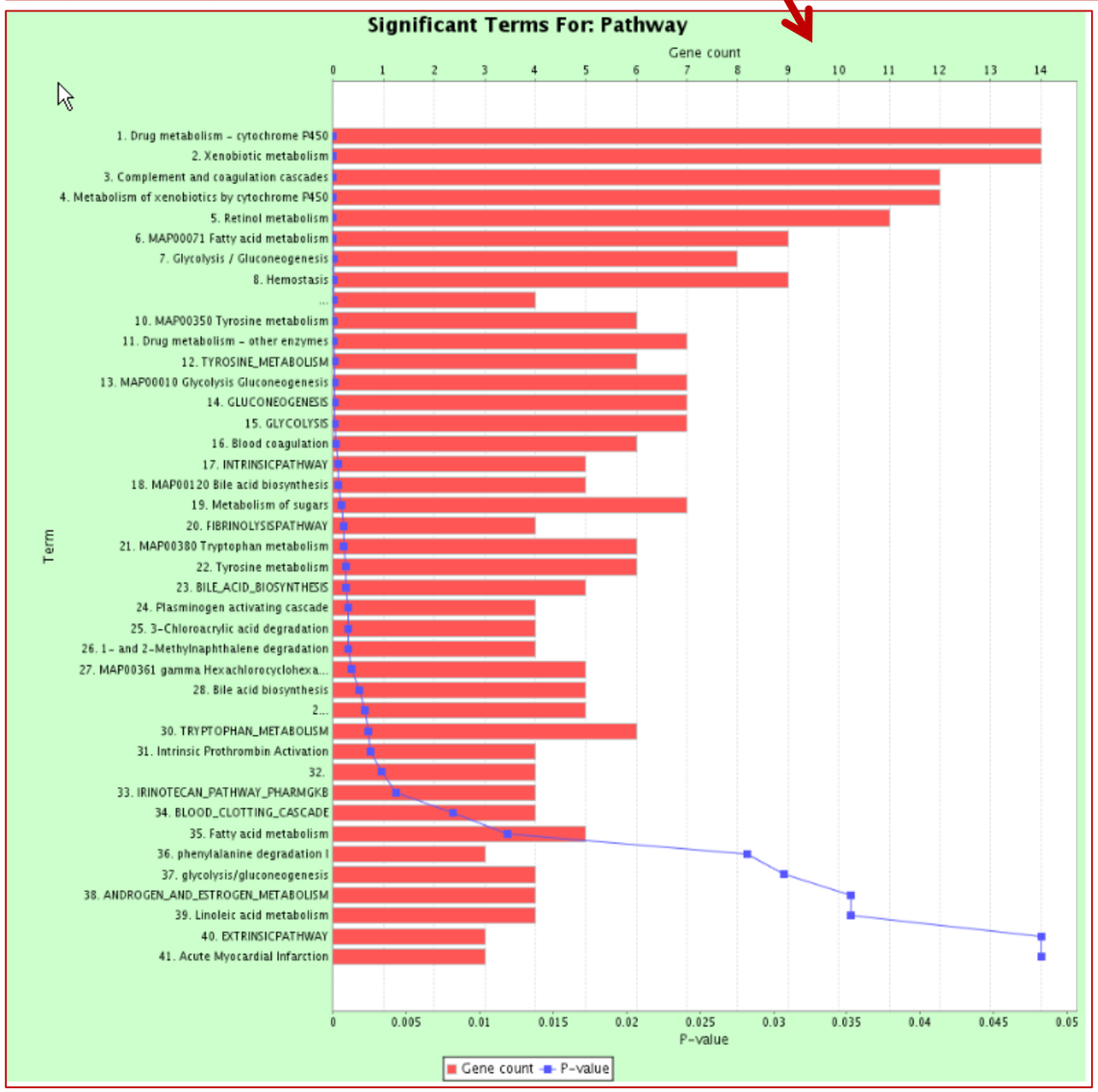
2: GO: Biological Process [Display Chart] [Hide Detail]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	GO:0009605	response to external stimulus		0	27	893
2	GO:0007596	blood coagulation		0	12	115
3	GO:0006629	lipid metabolic process		0	25	874
4	GO:0044255	cellular lipid metabolic process		0	23	720
5	GO:0050817	coagulation		0	12	119
6	GO:0007599	hemostasis		0	12	120
7	GO:0009611	response to wounding		0	20	542
8	GO:0042060	wound healing		0	13	185
9	GO:0050878	regulation of body fluid levels		0	12	151
10	GO:0055114	oxidation reduction		0	19	624
11	GO:0019752	carboxylic acid metabolic process		0	18	570

ToppGene Suite (http://toppgene.cchmc.org) - ToppFun

2: GO: Biological Process [Display Chart] [Hide Detail]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	GO:0009605	response to external stimulus		0	27	893
2	GO:0007596	blood coagulation		0	12	115
3	GO:0006629	lipid metabolic process		0	25	874
4	GO:0044255	cellular lipid metabolic process		0	23	720
5	GO:0050817	coagulation		0	12	119
6	GO:0007599	hemostasis		0	12	120
7	GO:0009611	response to wounding		0	20	542
8	GO:0042060	wound healing		0	13	185
9	GO:0050878	regulation of body fluid levels		0	12	151
10	GO:0055114	oxidation reduction		0	19	624
11	GO:0019752	carboxylic acid metabolic process		0	18	570



response to external stimulus; GO:0009605

	Entrez Gene ID	Gene Symbol	Gene Name	Original Symbol
1	126	ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide	126
2	335	APOA1	apolipoprotein A-I	335
3	350	APOH	apolipoprotein H (beta-2-glycoprotein I)	350
4	2158	F9	coagulation factor IX	2158
5	5950	RBP4	retinol binding protein 4, plasma	5950
6	197	AHSG	alpha-2-HS-glycoprotein	197
7	2243	FGA	fibrinogen alpha chain	2243
8	213	ALB	albumin	213
9	2244	FGB	fibrinogen beta chain	2244
10	629	CFB	complement factor B	629
11	3158	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	3158
12	5444	PON1	paraoxonase 1	5444
13	1361	CPB2	carboxypeptidase B2 (plasma)	1361
14	3078	CFHR1	complement factor H-related 1	3078
15	5265	SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265
16	3827	KNG1	kininogen 1	3827
17	325	APCS	amyloid P component, serum	325
18	2538	G6PC	glucose-6-phosphatase, catalytic subunit	2538
19	4153	MBL2	mannose-binding lectin (protein C) 2, soluble (opsonic defect)	4153
20	735	C9	complement component 9	735
21	462	SERPINC1	serpin peptidase inhibitor, clade C (antithrombin), member 1	462
22	3273	HRG	histidine-rich glycoprotein	3273
23	5340	PLG	plasminogen	5340
24	5004	ORM1	orosomucoid 1	5004
25	316	AOX1	aldehyde oxidase 1	316
26	3053	SERPIND1	serpin peptidase inhibitor, clade D (heparin cofactor), member 1	3053
27	1356	CP	ceruloplasmin (ferroxidase)	1356

ToppGene Suite (http://toppgene.cchmc.org) - ToppFun

ToppGene Result Page

Number of genes in training set:	81
Number of genes in test set:	0
Correction and Cutoff:	

category	Correction	Cutoff	Min	Max
GO: Molecular Function	Bonferroni	0.05	1	1500
GO: Biological Process	Bonferroni	0.05	1	1500
GO: Cellular Component	Bonferroni	0.05	1	1500
Human Phenotype	Bonferroni	0.05	1	1500
Mouse Phenotype	Bonferroni	0.05	1	1500
Domain	Bonferroni	0.05	1	1500
Pathway	Bonferroni	0.05	1	1500
Pubmed	Bonferroni	0.05	1	1500
Interaction	Bonferroni	0.05	1	1500
Cytoband	Bonferroni	0.05	1	1500
TFBS	Bonferroni	0.05	1	1500
Gene Family	Bonferroni	0.05	1	1500
Coexpression	Bonferroni	0.05	1	1500
Computational	Bonferroni	0.05	1	1500
MicroRNA	Bonferroni	0.05	1	1500
Drug	Bonferroni	0.05	1	1500
Disease	Bonferroni	0.05	1	1500

Random sampling size in analysis:	0
Minimun feature count in test set:	2
Analysis took:	2 seconds
Analysis finished at:	Sun Sep 27 16:45:06 EDT 2009

Training Results [Show All] [Download All] [Sparse Matrix]

1: GO: Molecular Function [Display Chart] [Show Detail]

2: GO: Biological Process [Display Chart] [Show Detail]

3: GO: Cellular Component [Display Chart] [Show Detail]

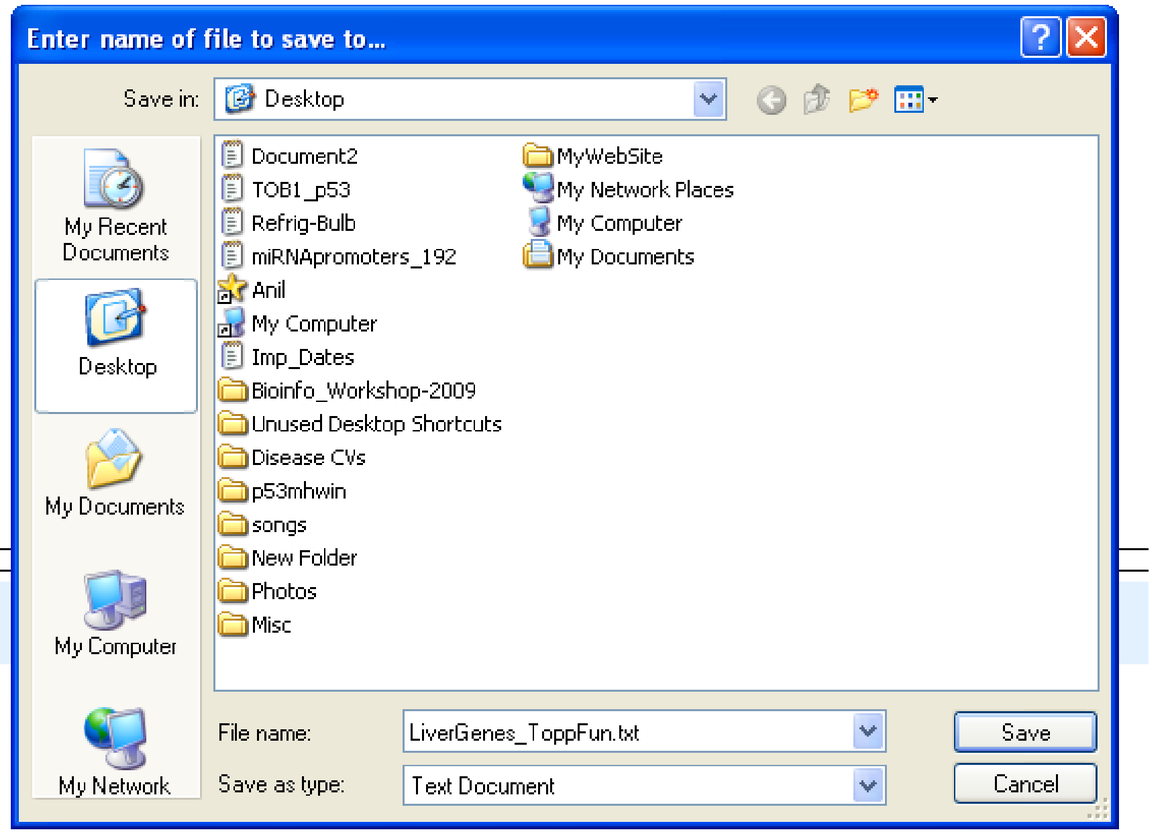
4: Human Phenotype [Display Chart] [Show Detail]

5: Mouse Phenotype [Display Chart] [Show Detail]

6: Domain [Display Chart] [Show Detail]

7: Pathway [Display Chart] [Show Detail]

8: Pubmed [Display Chart] [Show Detail]



Download Example Data Sets for Exercises From http://anil.cchmc.org/dhc.html

Two Excel Files:

1. **GeneLists.xls**: Has two worksheets
   a. Tissue_GeneLists: Has a list of overexpressed genes in some of the digestive system tissues
   b. miRNA-Targets_Validated: Has a list of validated target genes for some of the microRNAs
2. **CandidateGenes.xls**: Has two worksheets
   a. abnormal_dig_sys_morph_genes: Has a list of genes associated with the phenotype abnormal digestive system morphology in mouse
   b. miRNA_Putatitve_Targets: Has a list of predicted targets of some of the miRNAs from TargetScan (version 5.0)

Exercise 1: Use the different gene lists from the downloaded file ("GeneLists.xls") and find out:

Note: The "GeneLists.xls" file has two worksheets and within each worksheet there are several gene lists based on tissue-specificity or being microRNA targets (validated)

a. How many of the liver-overexpressed genes are associated with lipid metabolic process?
b. Are there any enriched TFBSs for liver overexpressed genes?
c. What are the enriched miRNAs in the colon-cecum overexpressed genes?
d. What gene families are enriched in esophagus overexpressed genes?
e. In which other regions are stomach (cardiac) genes overexpressed?
f. What biological process are miR-1 target genes enriched for?

What if I want to compare several gene lists at a time?

ToppCluster (http://toppcluster.cchmc.org)

Cincinnati Children's Biomedical Informatics

ToppCluster

A multiple gene list feature enrichment analyzer for the dissection of biological systems

Navigation
- Main
- Alternative Entry Methods

Information
- Disclaimer
- Documentation
- Supplementary

ToppGene
- Database Info
- Links

Paste input list Load Sample Data

Symbols are HGNC Symbol

Cluster Name Kidney

remove

Genes

ALDH8A1
LOC340094
LOC134147
SLC36A2
DZIP1
EHHADH
BTC
TMEM12
LOC388588
DNASE1L3
KLHDC7A
BBOX1
HOXD10
GLDC
AQP2
WDR72
KCNK5
VNN1
LOC441748
LOC348174
GATM
EMX1
DKFZP586D0919
EMX2
UNC5CL
BTBD5
HNF4G
LOC285733
PPARGC1A
DHDH

Symbols are HGNC Symbol

Cluster Name

remove

- HGNC Symbol
- HGNC Symbol and Synonyms
- Entrez ID
- Ensembl ID
- RefSeq
- Uniprot

CYP3A4
ALDH8A1
JUB
CYP3A5
A2M
CYP3A7
AQP9
MASP2
EHHADH
CYP2D6
TMEM12

ToppCluster (http://toppcluster.cchmc.org)

Options

Feature	Correction	p-Value cutoff	Gene Limits
All	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Molecular Function	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Biological Process	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Cellular Component	Bonferroni	0.05	1 ≤ n ≤ 1500
Human Phenotype	Bonferroni	0.05	1 ≤ n ≤ 1500
Mouse Phenotype	Bonferroni	0.05	1 ≤ n ≤ 1500
Domain	Bonferroni	0.05	1 ≤ n ≤ 1500
Pathway	Bonferroni	0.05	1 ≤ n ≤ 1500
Pubmed	Bonferroni	0.05	1 ≤ n ≤ 1500
Interaction	Bonferroni	0.05	1 ≤ n ≤ 1500
Cytoband	Bonferroni	0.05	1 ≤ n ≤ 1500
TFBS	Bonferroni	0.05	1 ≤ n ≤ 1500
Gene Family	Bonferroni	0.05	1 ≤ n ≤ 1500
Coexpression	Bonferroni	0.05	1 ≤ n ≤ 1500
Computational	Bonferroni	0.05	1 ≤ n ≤ 1500
MicroRNA	Bonferroni	0.05	1 ≤ n ≤ 1500
Drug	Bonferroni	0.05	1 ≤ n ≤ 1500
Disease	Bonferroni	0.05	1 ≤ n ≤ 1500

Chose Toppcluster output format: Network Generator

- Network Generator
- ***Interactive***
  - Comma Separated Values
  - GenePattern Format (GCT)
  - Tab Separated Values
  - HTML Table
  - Network Generator
- ***Batch***
  - Comma Separated Values
  - Tab Separated Values
  - Clustered Data (Zipped)
  - PDF Heatmap

Gene Sets

Kidney
303 known - 59 unknown

Original	Human Symbol	Entrez ID
ALDH8A1	ALDH8A1	64577
LOC340094	LOC340094	340094
SLC36A2	SLC36A2	153201
DZIP1	DZIP1	22873
EHHADH	EHHADH	1962
BTC	BTC	685
LOC388588	LOC388588	388588
DNASE1L3	DNASE1L3	1776
KLHDC7A	KLHDC7A	127707
BBOX1	BBOX1	8424
HOXD10	HOXD10	3236
GLDC	GLDC	2731
AQP2	AQP2	359

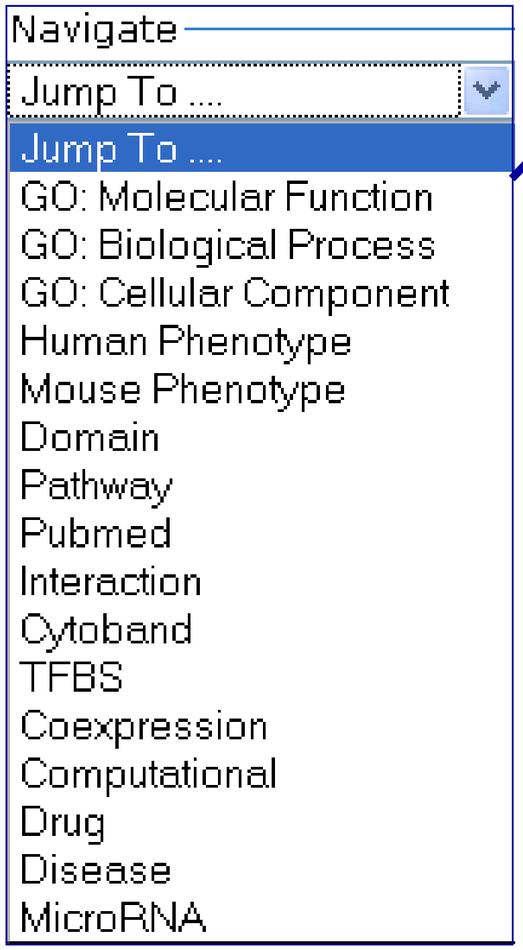
Liver
279 known - 30 unknown

Original	Human Symbol	Entrez ID
CYP3A4	CYP3A4	1576
ALDH8A1	ALDH8A1	64577
JUB	JUB	84962
CYP3A5	CYP3A5	1577
A2M	A2M	2
CYP3A7	CYP3A7	1551
AQP9	AQP9	366
MASP2	MASP2	10747
EHHADH	EHHADH	1962
CYP2D6	CYP2D6	1565
RFXAP	RFXAP	5994
DNASE1L3	DNASE1L3	1776
HEMGN	HEMGN	55363
AGTR1	AGTR1	185

Pancreas
134 known - 46 unknown

Original	Human Symbol	Entrez ID
APOBEC2	APOBEC2	10930
EDN3	EDN3	1908
SYT5	SYT5	6861
SLC44A4	SLC44A4	80736
C14orf50	C14orf50	145376
CDH22	CDH22	64405
TTR	TTR	7276
ERO1LB	ERO1LB	56605
FBXO27	FBXO27	126433
GPR44	GPR44	11251
PPP1R1A	PPP1R1A	5502
GP2	GP2	2813
CPA2	CPA2	1358
PLA2G1B	PLA2G1B	5319

ToppCluster (http://toppcluster.cchmc.org)



Navigate: Jump To .... | Links: Back to Start, Shareable Link | Highlighting: Highlight genes, Clear | Network Generator: Select nodes to be included in the network Then click next, Next »

Category	ID	Title (or Source)		Kidney_logP	Liver_logP	Pancreas_logP	Kidney_GeneSet
GO: Molecular Function				pValues			
GO: Molecular Function	GO:0048037	cofactor binding	☑	☑ 3.8132	☑ 10.0000		ACADSB AGXT2 ALDH6A1 CHDH DDC DMGDH EHHADH FMO1 GCSH GLDC GPT HAO2 HNF4A MIOX MOSC2 NOX4 OGDHL PAH
GO: Molecular Function	GO:0004252	serine-type endopeptidase activity	☑		☑ 10.0000	☑ 3.3520	
GO: Molecular Function	GO:0005215	transporter activity	☑	☑ 10.0000	☑ 3.0795		ABCC6 AQP2 ATP6V0A4 ATP6V0D2 ATP6V1G3 CLCNKB CLDN16 CUBN FXYD2 FXYD4 HCN3 KCNJ1 KCNJ15 KCNJ16 KCNK5 LOC153328 LRP2 OSTBETA PDZK1 RBP5 SCNN1G SLC10A2 SLC12A1 SLC12A3 SLC13A1 SLC13A2 SLC13A3 SLC16A12 SLC16A7 SLC16A9 SLC17A1 SLC17A3 SLC1A1 SLC22A2 SLC22A6 SLC22A7 SLC22A8 SLC23A1 SLC23A3 SLC26A7 SLC27A2 SLC28A1 SLC2A2 SLC2A9 SLC30A2 SLC34A1 SLC36A2 SLC3A1 SLC4A4 SLC4A9 SLC5A11 SLC5A12 SLC5A2 SLC5A8 SLC6A13 SLC6A19 SLC6A3 SLC7A13 SLC7A9 SLCO4C1 TRPM6
GO: Molecular Function	GO:0017171	serine hydrolase activity	☑		☑ 10.0000	☑ 2.7787	
GO: Molecular Function	GO:0031406	carboxylic acid binding	☑	☑ 2.1541	☑ 10.0000		DMGDH DPYS FOLR3 FTCD GCSH GLDC HNF4A NR1H4 PAH PCK1 SLC1A1
GO: Molecular Function	GO:0016491	oxidoreductase activity	☑	☑ 2.0183	☑ 10.0000		ACADSB ADH6 ALDH6A1 ALDH8A1 BBOX1 CHDH CYP17A1 CYP27B1 CYP4V2 DAO DHDH DIO1 DMGDH EHHADH FMO1 GLDC HAO2 HGD HSDL1 MIOX MOSC2 NDUFC1 NOX4 OGDHL PAH PIPOX PRODH2
GO: Molecular Function	GO:0005496	steroid binding	☑	☑ 1.5055	☑ 10.0000		AGXT2 CALB1 DDC GLDC GPT HNF4A HNF4G KL MOSC2 NR1H4
GO: Molecular Function	GO:0001871	pattern binding	☑		☑ 10.0000		
GO: Molecular Function	GO:0008324	cation transmembrane transporter activity	☑	☑ 10.0000			ATP6V0A4 ATP6V0D2 ATP6V1G3 CLDN16 FXYD4 HCN3 KCNJ1 KCNJ15 KCNJ16 KCNK5 SCNN1G SLC10A2 SLC12A1 SLC12A3 SLC13A1 SLC13A2 SLC13A3 SLC17A3 SLC1A1 SLC22A2 SLC22A8 SLC23A1 SLC28A1 SLC2A2 SLC2A9 SLC30A2 SLC36A2 SLC4A4 SLC5A11 SLC5A2 SLC6A13 SLC6A19 SLC6A3 TRPM6
GO: Molecular Function	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	☑		☑ 10.0000		

Network Generator

Warning: because of the network size, some options like layout, preview and PNG have been limited or hidden.

Summary

2126 checked boxes
2016 nodes in an ABSTRACTED network.
2600 nodes in a GENE LEVEL network.

Method

- (•) **ABSTRACTED** A Feature to Cluster network where the score is used as the edge weight..
- ( ) **GENE LEVEL** A gene-based network where each gene connects from a Cluster to a Feature.

Method Options

- [ ] Create a Category node linked to all features of that category

File Format

- (•) **XGMML** An XML based format compatible with Cytoscape. (more information)
- ( ) **TEXT** A Simple Text Format.
- ( ) **GEXF** An XML based format compatible with Gephi. (more information)

Begin

Network Generator

Summary

279 checked boxes
265 nodes in an ABSTRACTED network.
660 nodes in a GENE LEVEL network.

Method

- (•) **ABSTRACTED** A Feature to Cluster network where the score is used as the edge weight..
- ( ) **GENE LEVEL** A gene-based network where each gene connects from a Cluster to a Feature.

Method Options

- [ ] Create a Category node linked to all features of that category

Layout Algorithm

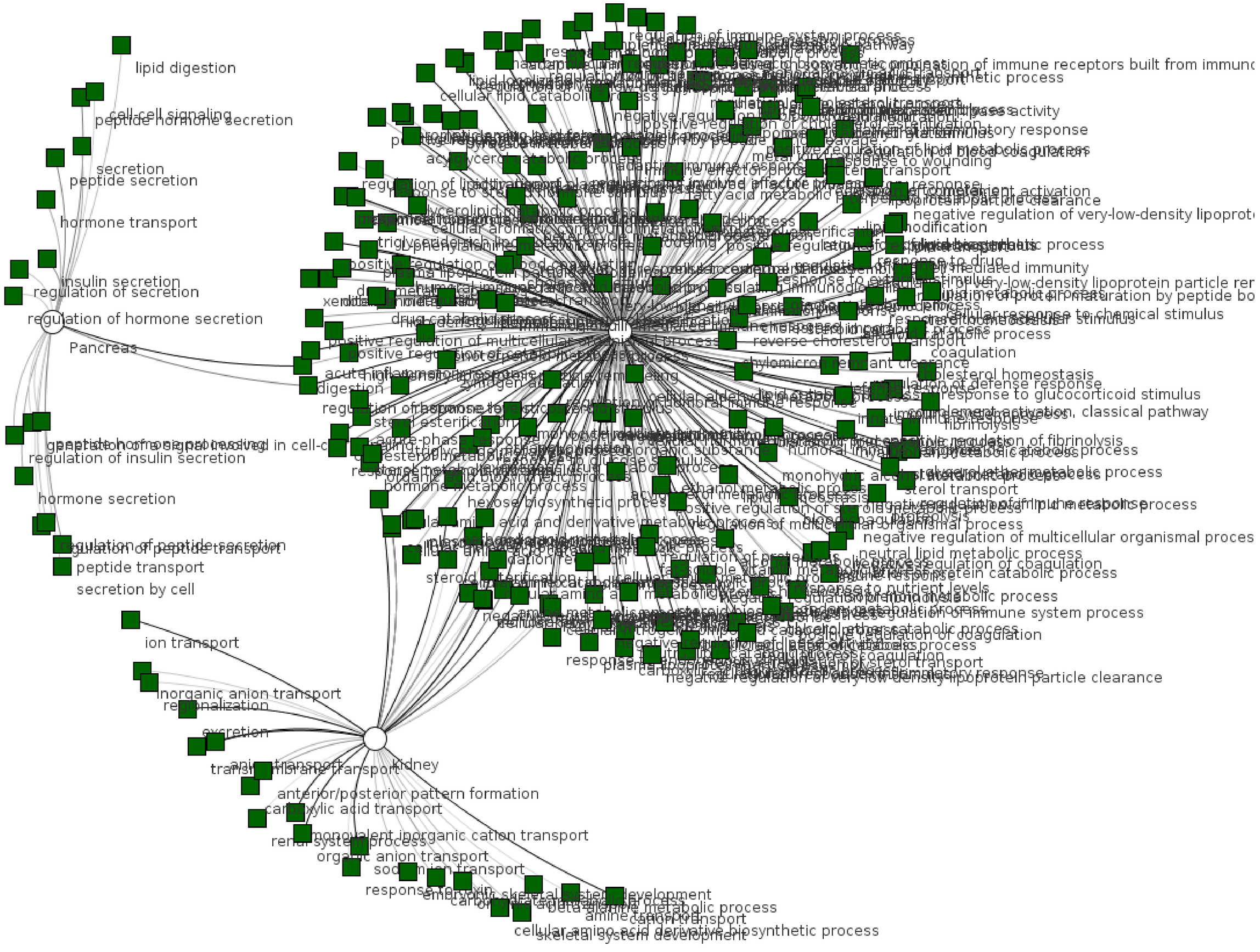
- (•) Kamada-Kawai.
- ( ) Fruchterman-Reingold.
- ( ) Spring.
- ( ) Circle.
- ( ) Meyer's Self-Organizing.

File Format

- (•) **XGMML** An XML based format compatible with Cytoscape. (more information)
- ( ) **PNG** A raster graphic format. (more information)
- ( ) **TEXT** A Simple Text Format.
- ( ) **GEXF** An XML based format compatible with Gephi. (more information)

Begin Preview

ToppCluster (http://toppcluster.cchmc.org)



ToppCluster (http://toppcluster.cchmc.org)

Network Generator

Summary

279 checked boxes
265 nodes in an ABSTRACTED network.
660 nodes in a GENE LEVEL network.

Method

- **ABSTRACTED** A Feature to Cluster network where the score is used as the edge weight..
- **GENE LEVEL** A gene-based network where each gene connects from a Cluster to a Feature.

Method Options

- Create a Category node linked to all features of that category

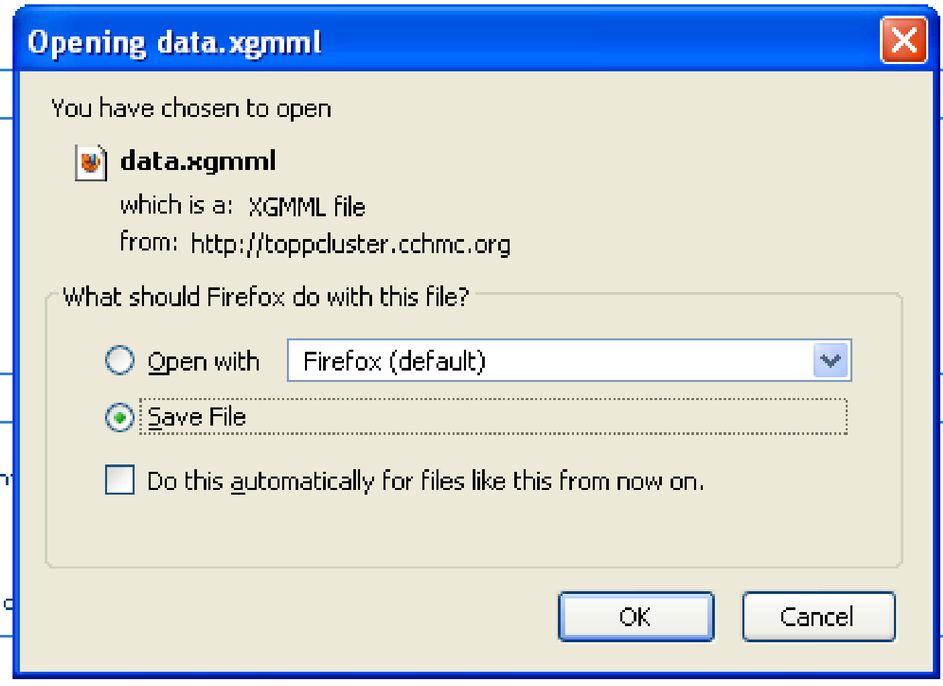
Layout Algorithm

- Kamada-Kawai.
- Fruchterman-Reingold.
- Spring.
- Circle.
- Meyer's Self-Organizing.

File Format

- **XGMML** An XML based format compatible with Cytoscape. (more in
- **PNG** A raster graphic format. (more information)
- **TEXT** A Simple Text Format.
- **GEXF** An XML based format compatible with Gephi. (more informatic

Begin **Preview**

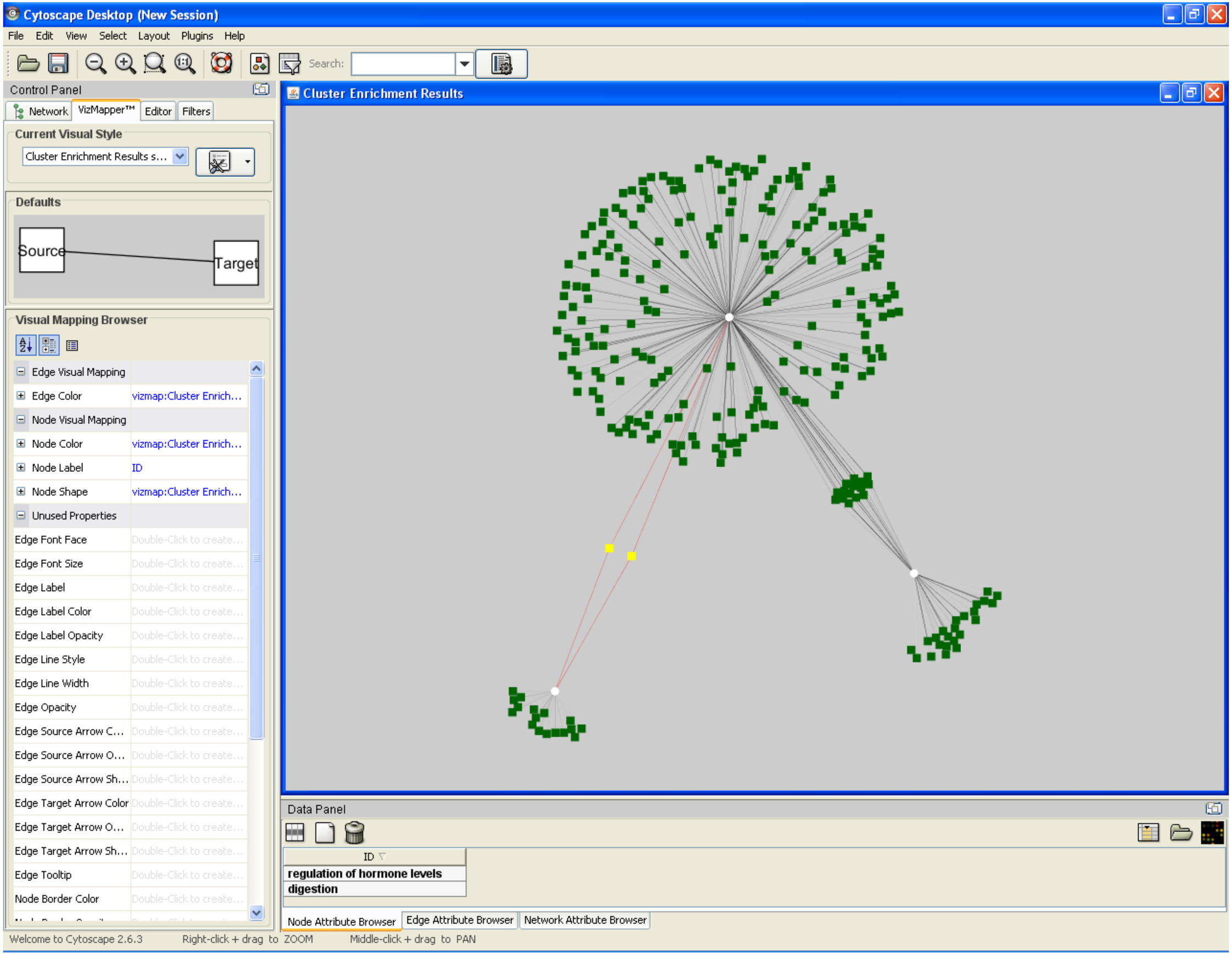


Cytoscape (http://cytoscape.org)

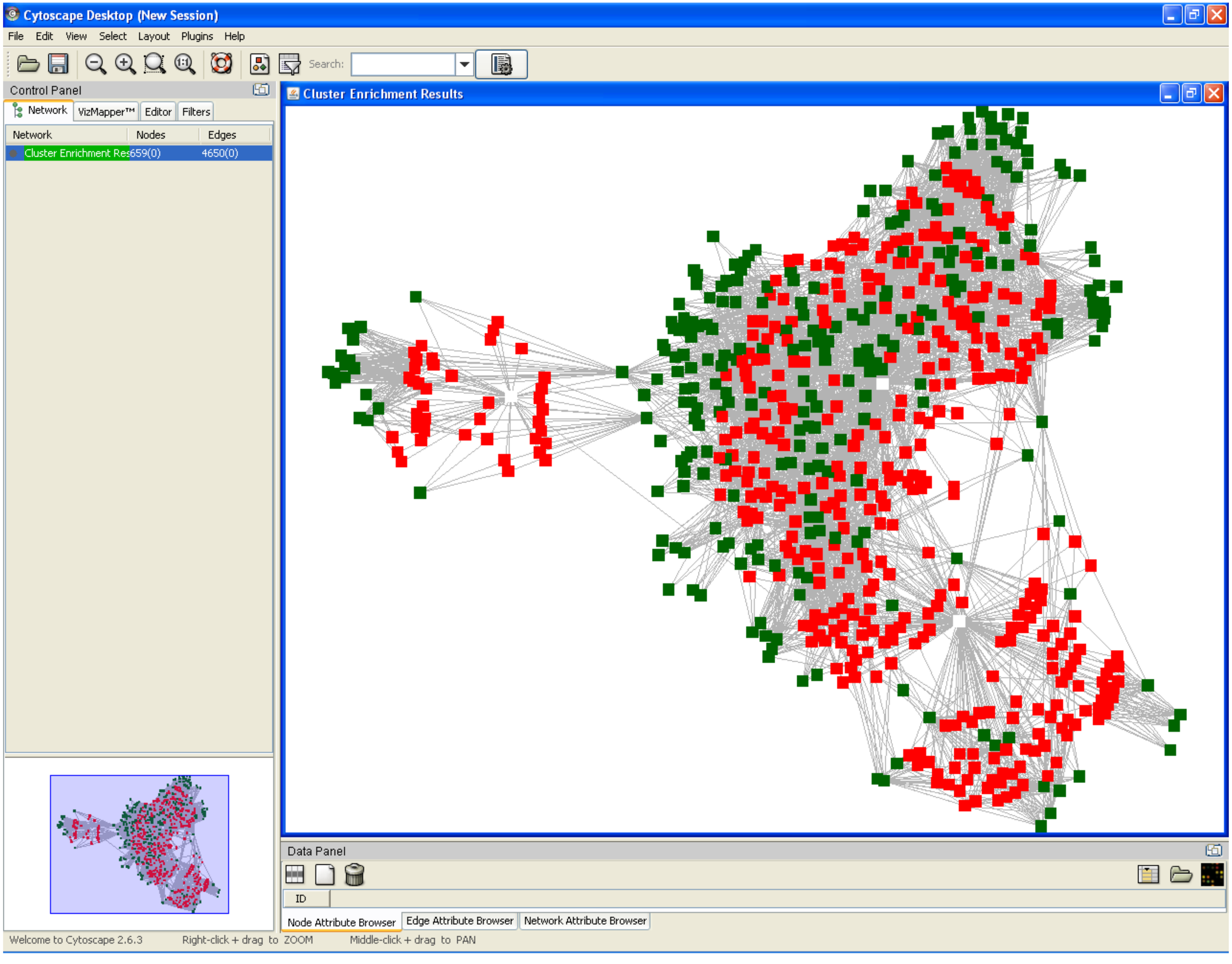
Gephi (http://gephi.org)

Should be installed on your computer and the downloaded files should be imported into these applications

Cytoscape Network (Abstract View)



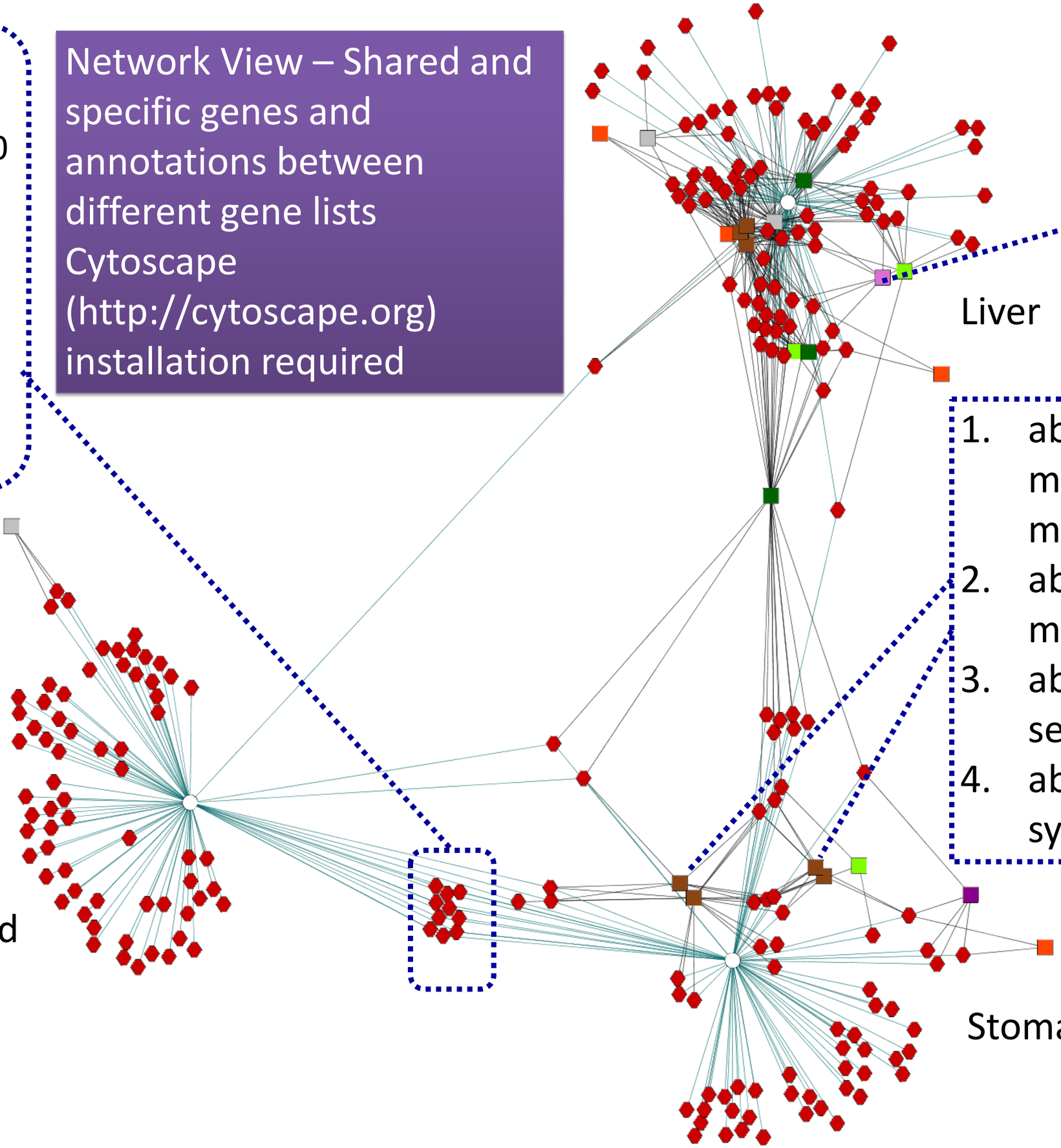
Cytoscape Network (GeneLevel View)



Cytoscape Network (GeneLevel View)

EHF
COL15A1
LOC100130100
IGHA1
LTF
IGKC
IGL@
FAM129A
ATP8B1
IGLC2

Network View – Shared and specific genes and annotations between different gene lists Cytoscape (http://cytoscape.org) installation required



V$HNF1

Liver

1. abnormal gastric mucosa morphology
2. abnormal stomach morphology
3. abnormal digestive secretion
4. abnormal digestive system physiology

Salivary Gland

Stomach

Exercise 2: Use the different gene lists from the downloaded file ("GeneLists.xls") and find out:

Note: The "GeneLists.xls" file has two worksheets and within each worksheet there are several gene lists based on tissue-specificity or being microRNA targets (validated)

a. What are the shared and specific biological processes between stomach and salivary glands?
b. Are there any enriched miRNAs for stomach? If so, which other tissues are enriched for this miRNA?
c. What are the functional similarities and differences between the 3 regions of the stomach (cardiac, fundus, and pylorus)?

ToppGene Suite (http://toppgene.cchmc.org)

I have a list of 200 over-expressed genes and I want to prioritize them for experimental validation (apart from using the fold change as a parameter)…..

ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

- **ToppFun**: Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

  Detect functional enrichment of your gene list based on Transcriptome, Proteome, Regulome (TFBS and miRNA), Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), Pharmacome (Drug-Gene associations), literature co-citation, and other features.

- **ToppGene**: Candidate gene prioritization

  Prioritize or rank candidate genes based on functional similarity to training gene list.

- **ToppNet**: Relative importance of candidate genes in networks

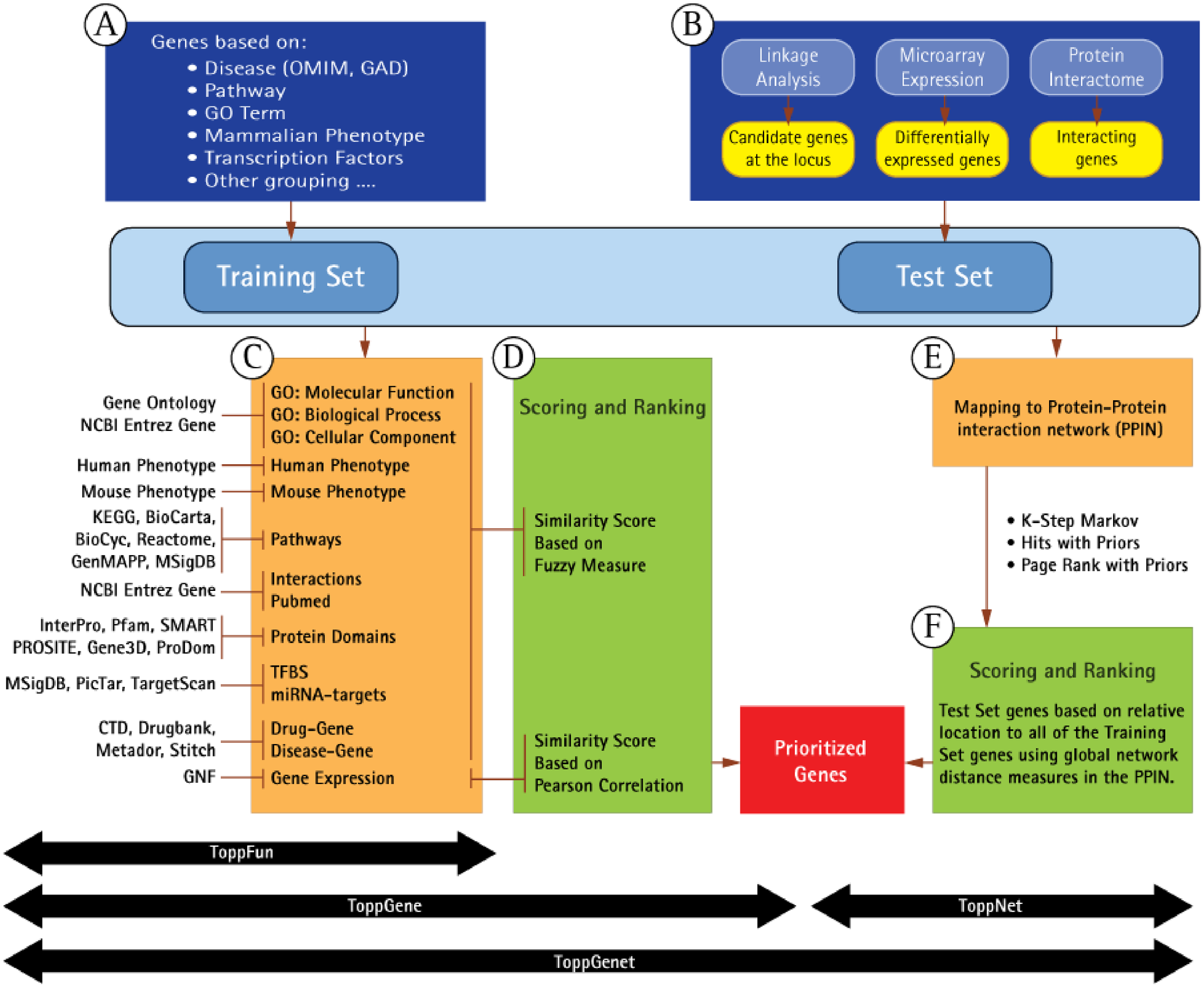
  Prioritize or rank candidate genes based on topological features in protein-protein interaction network.

- **ToppGenet**: Prioritization of neighboring genes in protein-protein interaction network

  Identify and prioritize the neighboring genes of the seeds in protein-protein interaction network based on functional similarity to the "seed" list (ToppGene) or topological features in protein-protein interaction network (ToppNet).

ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

I have a list of 200 over-expressed genes and I want to prioritize them for experimental validation (apart from using the fold change as a parameter).....



ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

ToppGene: Candidate gene prioritization

Select your gene identifier type, paste your training and test gene sets below or select example sets, then submit.

Example gene sets: HGNC Symbol Entrez ID
(click on "HGNC Symbol" or "Entrez ID" to use the example training and test set of genes)

Symbol Types	HGNC Symbol	HGNC Symbol
Training Gene Set:	NKX2-5 MEF2A GATA4 HAND1 HAND2 TBX5 SRF	
Test gene set:		ACVR1 ACVR2B ADAM19 ADM ADRA1A ADRA1B ADRBK1 ALDH1A2 ALPK3 ATP6VOA1 BMP10 BMP2 BMP4 BMPR1A CALCRL CASP3 CASP7 CASP8 CASQ2 CENTA2 CHD7

Clear Submit Query

ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

Training set (7 / 7)

Entered	Human Symbol	Gene ID
NKX2-5	NKX2-5	1482
MEF2A	MEF2A	4205
GATA4	GATA4	2626
HAND1	HAND1	9421
HAND2	HAND2	9464
TBX5	TBX5	6910
SRF	SRF	6722

Test set (146 / 158)

Entered	Human Symbol	Gene ID
ACVR1	ACVR1	90
ACVR2B	ACVR2B	93
ADAM19	ADAM19	8728
ADM	ADM	133
ADRA1A	ADRA1A	148
ADRA1B	ADRA1B	147
ADRBK1	ADRBK1	156
ALDH1A2	ALDH1A2	8854
ALPK3	ALPK3	57538
ATP6V0A1	ATP6V0A1	535
BMP10	BMP10	27302
BMP2	BMP2	650
BMP4	BMP4	652
BMPR1A	BMPR1A	657
CALCRL	CALCRL	10203
CASP3	CASP3	836
CASP7	CASP7	840
CASP8	CASP8	841
CASQ2	CASQ2	845
CHD7	CHD7	55636
CITED2	CITED2	10370

Entered	Suggestions
CENTA2	☑ ADAP2 - ArfGAP with dual PH domains 2 Human Synonym
CMYA1	☑ XIRP1 - xin actin-binding repeat ontaining 1 Human Synonym
GJA7	☑ GJC1 - gap junction protein, gamma 1, 45kDa Human Synonym
HOP	☑ HOPX - HOP homeobox Human Synonym
	☐ ST13 - suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein) Human Synonym
	☐ STIP1 - stress-induced-phosphoprotein 1 Human Synonym
PPARBP	☑ MED1 - mediator complex subunit 1 Human Synonym
RBPSUH	RBPJ Duplicated

Update ☐ Check All

Ignored

Entered	Status
CENTA2	Not Found
CMYA1	Not Found
GATA4	In Training Set
GJA7	Not Found
HAND1	In Training Set
HAND2	In Training Set
HOP	Not Found
NKX2-5	In Training Set
PPARBP	Not Found
RBPSUH	Not Found
SRF	In Training Set
TBX5	In Training Set

Find alternatives for missing symbols

ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

Training parameters

Feature	Correction	p-Value cutoff	Gene Limits
All	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Molecular Function	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Biological Process	Bonferroni	0.05	1 ≤ n ≤ 1500
GO: Cellular Component	Bonferroni	0.05	1 ≤ n ≤ 1500
Human Phenotype	Bonferroni	0.05	1 ≤ n ≤ 1500
Mouse Phenotype	Bonferroni	0.05	1 ≤ n ≤ 1500
Domain	Bonferroni	0.05	1 ≤ n ≤ 1500
Pathway	Bonferroni	0.05	1 ≤ n ≤ 1500
Pubmed	Bonferroni	0.05	1 ≤ n ≤ 1500
Interaction	Bonferroni	0.05	1 ≤ n ≤ 1500
Cytoband	Bonferroni	0.05	1 ≤ n ≤ 1500
TFBS	Bonferroni	0.05	1 ≤ n ≤ 1500
Gene Family	Bonferroni	0.05	1 ≤ n ≤ 1500
Coexpression	Bonferroni	0.05	1 ≤ n ≤ 1500
Computational	Bonferroni	0.05	1 ≤ n ≤ 1500
MicroRNA	Bonferroni	0.05	1 ≤ n ≤ 1500
Drug	Bonferroni	0.05	1 ≤ n ≤ 1500
Disease	Bonferroni	0.05	1 ≤ n ≤ 1500

Test parameter

Random sampling size: 1500 (6% of genome)

Min. feature count: 2

Home | Modify Query | Start prioritization

ToppGene is processing your query

Estimating p-Values

To see the training results before the test set is complete, click here.

ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

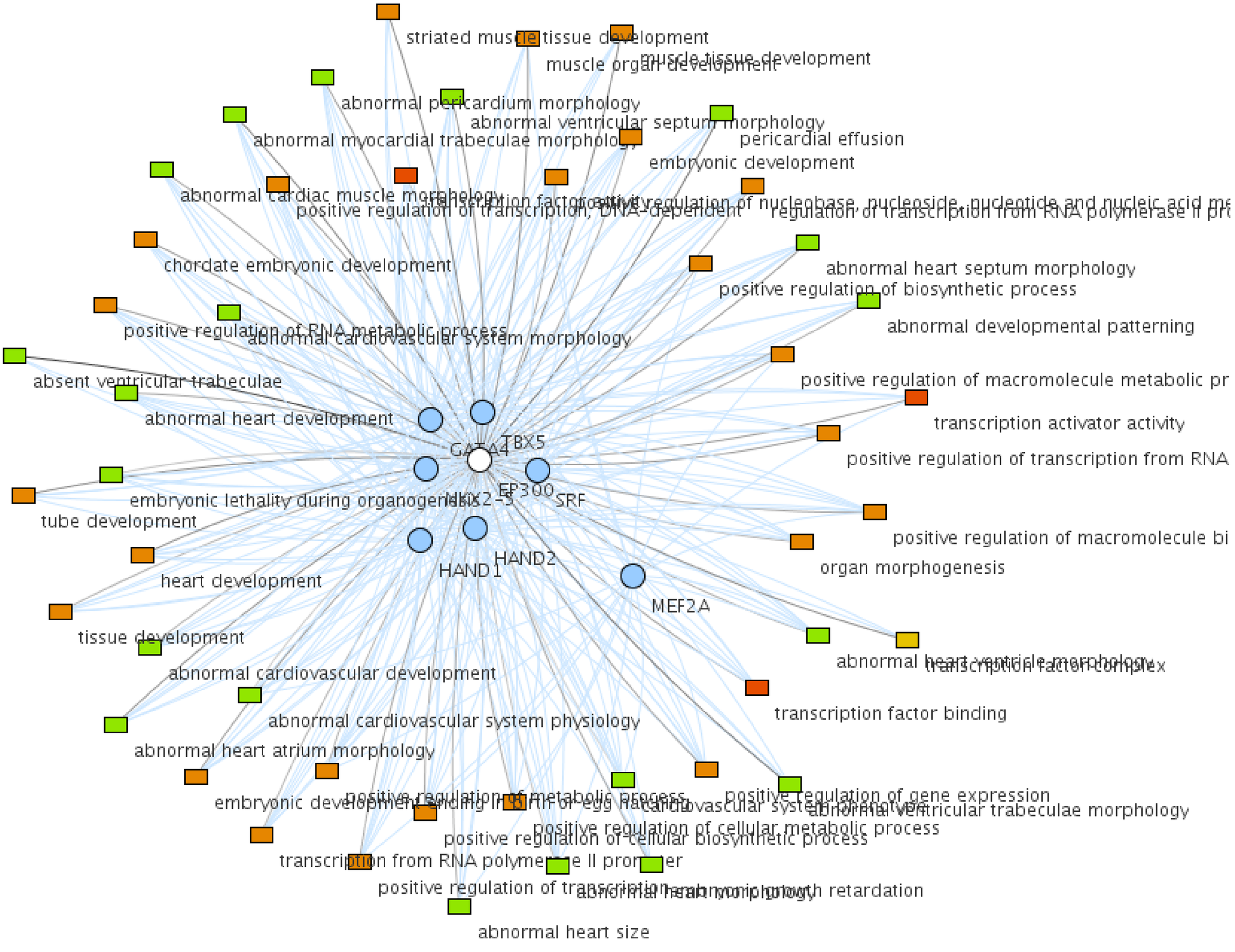
Rank	Gene Symbol	Gene ID	GO: Molecular Function		GO: Biological Process		GO: Cellular Component		Human Phenotype		Mouse Phenotype		Domain		Pathway		Pubmed		Interaction		Cytoband	
			Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue	Score	pValue
1	EP300	2033	0.7136999	0.0235641	0.9999726	0.0029455	0.4305814	0.001	0	0.5	0.9999881	0.005			0	0.5049834	0.8410056	0.001	0.7991753	0.001	0	0.5004907
2	TEAD1	7003	0.5804123	0.0309278	0.997771	0.0103093	0.4305814	0.001	0	0.5	0.9989337	0.035	0	0.5			0.7993714	0.001	0.7160863	0.001	0	0.5004907
3	HIF1A	3091	0.9391513	0.0029455	1	0.001	0.4305814	0.001			0.9997067	0.02	0.8885697	0.001	0	0.5049834	0	0.4995093	0	0.505	0	0.5004907
4	CTNNB1	1499	0.7136999	0.0235641	1	0.001	0.4305814	0.001	0	0.5	0.9529218	0.06			0.6371253	0.001	0	0.4995093	0	0.505	0	0.5004907
5	TBX20	57057	0.5804123	0.0309278	0.9999964	0.0014728	0	0.5022091			0.9999902	0.005	0	0.5			0	0.4995093			0	0.5004907
6	ZFPM2	23414	0.6308709	0.0250368	0.9999978	0.001	0	0.5022091	0	0.5	1	0.001	0	0.5			0	0.4995093			0	0.5004907
7	BMP4	652	0	0.5493373	1	0.001	0	0.5022091	0	0.5	0.9999435	0.01	0	0.5	0.6478057	0.001	0.7993714	0.001	0	0.505	0	0.5004907
8	TBX1	6899	0.9660807	0.001	0.999997	0.001	0	0.5022091	0	0.5	0.9996966	0.02	0	0.5			0	0.4995093			0	0.5004907
9	TBX2	6909	0.5418852	0.0397644	0.9943991	0.0162003	0.4305814	0.001			0.9993508	0.035	0	0.5	0	0.5049834	0	0.4995093			0	0.5004907
10	TGFB2	7042	0.8603852	0.005891	1	0.001	0	0.5022091			0.9999998	0.005	0	0.5	0.3937178	0.0033223	0	0.4995093	0	0.505	0	0.5004907

Rank	Gene Symbol
1	EP300
2	TEAD1
3	HIF1A
4	CTNNB1
5	TBX20
6	ZFPM2
7	BMP4
8	TBX1
9	TBX2
10	TGFB2

Average score	Overall P-value
0.3417445	0.0000003
0.3015437	0.0000058
0.3041435	0.0000062
0.2489552	0.0000788
0.3207447	0.0000893
0.2749466	0.000112
0.229808	0.0001787
0.2395215	0.0002528
0.2566618	0.0002615
0.2503156	0.0002619
0.3307561	0.0002975

ToppGene Suite (http://toppgene.cchmc.org) - ToppGene

Why is a test set gene ranked higher?



ToppGene Suite (http://toppgene.cchmc.org) - ToppNet

I have a list of 200 over-expressed genes and I want to prioritize them for experimental validation (apart from using the fold change as a parameter).....

ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

- **ToppFun**: Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis

  Detect functional enrichment of your gene list based on Transcriptome, Proteome, Regulome (TFBS and miRNA), Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), Pharmacome (Drug-Gene associations), literature co-citation, and other features.

- **ToppGene**: Candidate gene prioritization

  Prioritize or rank candidate genes based on functional similarity to training gene list.

- **ToppNet**: Relative importance of candidate genes in networks

  Prioritize or rank candidate genes based on topological features in protein-protein interaction network.

- **ToppGenet**: Prioritization of neighboring genes in protein-protein interaction network

  Identify and prioritize the neighboring genes of the seeds in protein-protein interaction network based on functional similarity to the "seed" list (ToppGene) or topological features in protein-protein interaction network (ToppNet).

ToppGene Suite (http://toppgene.cchmc.org) - ToppNet

Graph prioritization parameters

Prioritization method: k-Step Markov

Step Size(normally 4-8): 6

Training gene neighborhood subnetwork visualization parameters

Neighborhood distance: 1

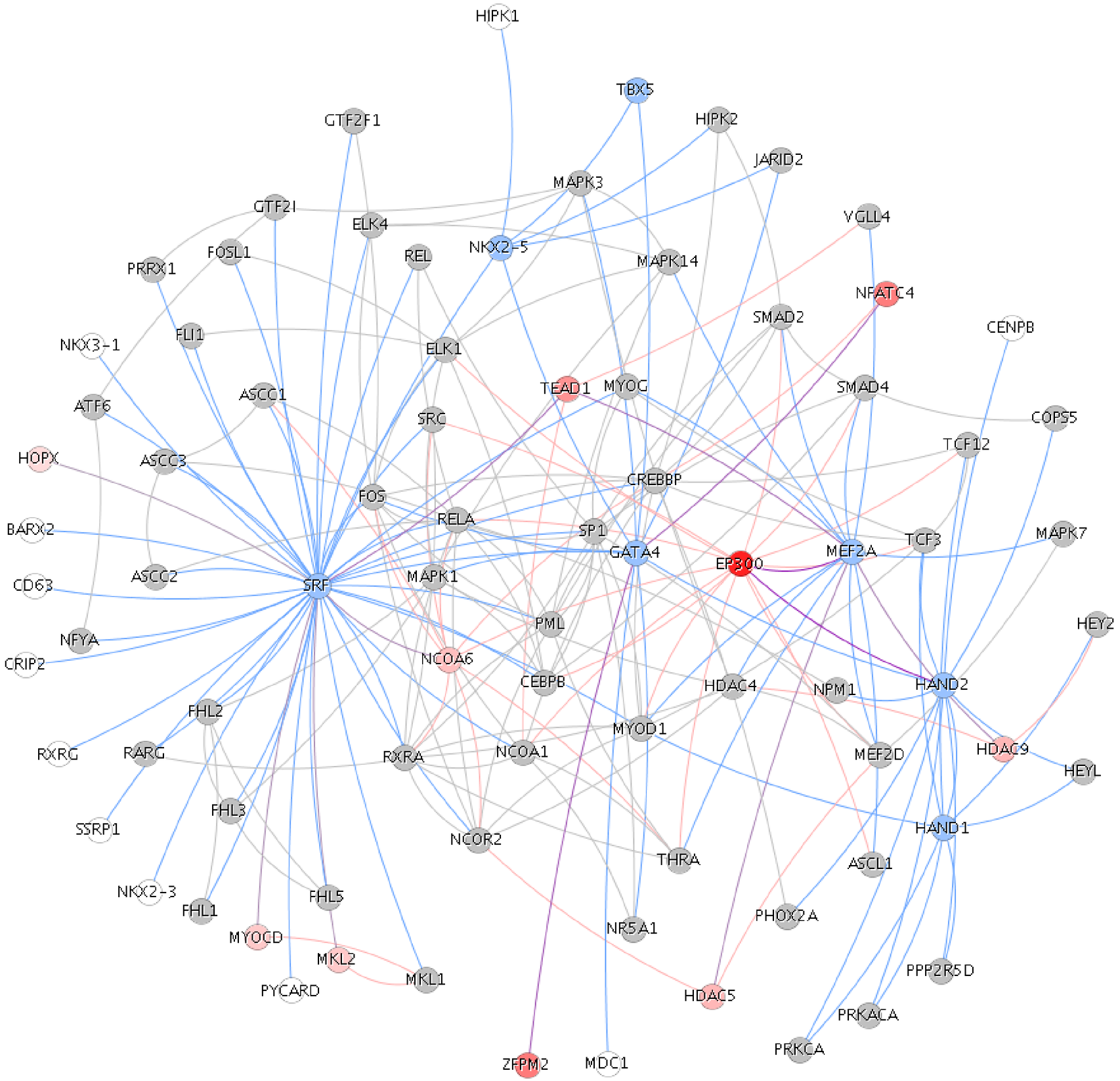
When training set is big, the training gene neighborhood subnetwork can be huge.

Home | Modify Query | Start prioritization

Test Genes [Hide All]

Rank	ID	Name	Interactant count	Score
1	2033	EP300	129	0.008192
2	23414	ZFPM2	4	0.004724
3	4776	NFATC4	4	0.004615
4	7003	TEAD1	9	0.003739
5	9734	HDAC9	20	0.002319
6	10014	HDAC5	33	0.002317
7	23054	NCOA6	49	0.001991
8	93649	MYOCD	2	0.0016
9	57496	MKL2	2	0.0016
10	1499	CTNNB1	138	0.001546

ToppGene Suite (http://toppgene.cchmc.org) - ToppNet



Exercise 3: Prioritize the 721 genes ("CandidateGenes.xls") using "stomach genes" from the "GeneLists.xls".

a. **What are the top 10 ranked genes using ToppGene and ToppNet?**
b. **What is the rank of TFF3 in ToppGene-based prioritization and why is it ranked among the top in ToppGene prioritization? What is its rank in ToppNet?**

Are there any other tools similar to these?

DAVID (http://david.abcc.ncifcrf.gov)

Database for Annotation, Visualization and Integrated Discovery

DAVID Bioinformatics Resources 2007

National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home | Start Analysis | Shortcut to DAVID Tools | Technical Center | Downloads & APIs | Term of Service | Why DAVID? | About Us

Shortcut to DAVID Tools

Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering new!, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. More

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More

Gene Name Batch Viewer new!

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. More

Welcome to DAVID Bioinformatics Resources 2003 - 2007

What's Special in DAVID 2007?

The Database for Annotation, Visualization and Integrated Discovery (**DAVID**) 2007 is the fifth ... programs of DAVID 2006 ... comprehensive set of functi... understand biological mean... gene list, DAVID tools are...

- Identify enriched biol...
- Discover enriched fu...
- Cluster redundant an...
- Visualize genes on Bi...
- Display related many...
- Search for other func...
- List interacting protei...
- Explore gene names ...
- Link gene-disease as...
- Highlight protein func...
- Redirect to related lit...
- Convert gene identife...
- And more

Screen Shot 1

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifers from one type to another.
- And more

DAVID (http://david.abcc.ncifcrf.gov)

Upload | List | Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

Clear

Or

B:Choose From a File

Browse...

Step 2: Select Identifier

AFFY_ID

Step 3: List Type

Gene List

Background

Step 4: Submit List

Submit List

Annotation Summary Results

Help and Tool Manual

Current Gene List: demolist1 — 171 DAVID IDs

Current Background: Homo sapiens — Check Defaults ☑ Clear All

⊞ Main Accessions (0 selected)

⊞ Other Accessions (0 selected)

⊟ Gene Ontology (3 selected)

☐ GOTERM_BP_1	79%	136	Chart
☐ GOTERM_BP_2	76%	131	Chart
☐ GOTERM_BP_3	74%	127	Chart
☐ GOTERM_BP_4	69%	119	Chart
☐ GOTERM_BP_5	60%	104	Chart
☑ GOTERM_BP_ALL	79%	136	Chart
☐ GOTERM_CC_1	70%	121	Chart
☐ GOTERM_CC_2	61%	106	Chart
☐ GOTERM_CC_3	55%	95	Chart
☐ GOTERM_CC_4	50%	86	Chart
☐ GOTERM_CC_5	38%	65	Chart
☑ GOTERM_CC_ALL	70%	121	Chart
☐ GOTERM_MF_1	75%	129	Chart
☐ GOTERM_MF_2	69%	119	Chart
☐ GOTERM_MF_3	60%	103	Chart
☐ GOTERM_MF_4	56%	97	Chart
☐ GOTERM_MF_5	45%	78	Chart
☑ GOTERM_MF_ALL	75%	129	Chart

⊞ Protein Domains (3 selected)

⊞ Pathways (3 selected)

⊞ General Annotations (0 selected)

⊞ Functional Categories (3 selected)

⊞ Protein Interactions (0 selected)

⊞ Literature (0 selected)

⊟ Disease (1 selected)

☑ GENETIC_ASSOCIATION_DB	13%	23
☐ OMIM_DISEASE	18%	32

Combined View for Selected Annotation

Functional Annotation Clustering new!

2D View

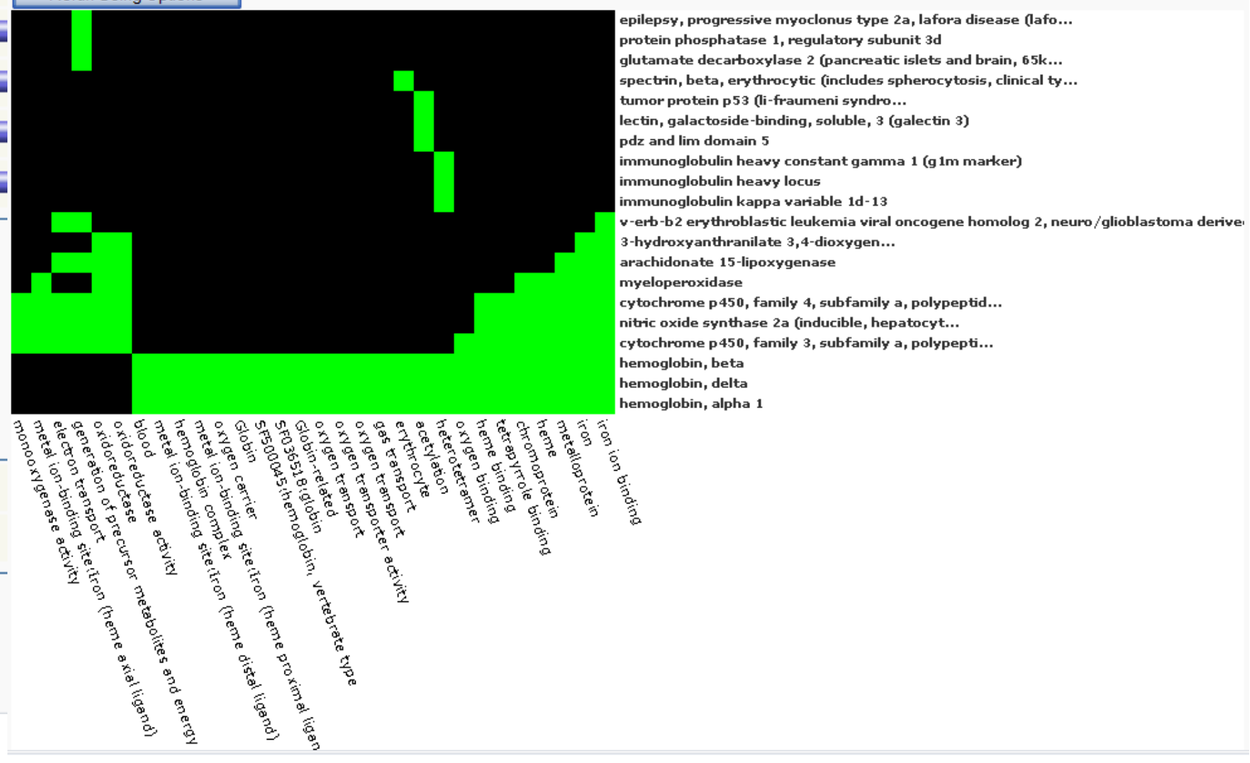
Help

corresponding gene-term association positively reported ■ corresponding gene-term association not reported yet

Download File ViewEditor

Options

Rerun Using Options



DAVID (http://david.abcc.ncifcrf.gov)

DAVID Bioinformatics Resources 2007
National Institute of Allergy and Infectious Diseases (NIAID), NIH

Home | Start Analysis | Shortcut to DAVID Tools | Technical Center | Downloads & APIs | Term of Service | Why DAVID? | About Us

Shortcut to DAVID Tools

Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering new!, BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and more

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. More

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. More

Gene Name Batch Viewer new!

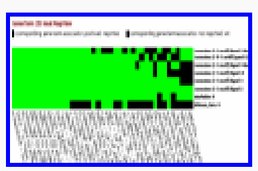
Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. More

Welcome to DAVID Bioinformatics Resources 2003 - 2007

What's Special in DAVID 2007?

The Database for Annotati… Visualization and Integrated Discovery
(**DAVID**) 2007 is the fifth …
programs of DAVID 2006 …
comprehensive set of functi…
understand biological mean…
gene list, DAVID tools are…

- Identify enriched biol…
- Discover enriched fu…
- Cluster redundant an…
- Visualize genes on Bi…
- Display related many…
- Search for other func…
- List interacting protei…
- Explore gene names …
- Link gene-disease as…
- Highlight protein func…
- Redirect to related lit…
- Convert gene identife…
- And more



Screen Shot 1

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifers from one type to another.
- And more

DAVID (http://david.abcc.ncifcrf.gov)

Convert NCBI Entrez Gene IDs to RefSeq Accession Numbers

Upload | List | Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

Clear

Or

B:Choose From a File

Browse...

Step 2: Select Identifier

AFFY_ID

Step 3: List Type

Gene List

Background

Step 4: Submit List

Submit List

Gene ID Conversion Tool

Submit your gene list to start conversion!

The Cross-Conversion of Gene ID Types:

- Entrez Gene ID
- Affy ID
- GenBank Accession
- Genpept Accession
- NCBI GI
- PIR Accession
- PIR ID
- PIR NREF ID
- RefSeq Genomic Accession
- RefSeq mRNA Accession
- RefSeq Protein Accession
- RefSeq RNA Accession
- Unigene
- UNIPROT Accession
- UNIPROT ID
- UNIREF100 ID
- Official Gene Symbol new!
- Not Sure new!

Upload | List | Background

Upload Gene List

Demolist 1 Demolist 2

Upload Help

Step 1: Enter Gene List

A: Paste a list

3493
3512
10562
3535
5284

Clear

Or

B:Choose From a File

Browse...

Step 2: Select Identifier

ENTREZ_GENE_ID

Step 3: List Type

Gene List

Background

Step 4: Submit List

Submit List

us how you like the tool

chnical notes of the tool

Contact us for questions

DAVID (http://david.abcc.ncifcrf.gov)

Upload | List | Background

Gene List Manager

Select to limit annotations by one or more species Help

- Use All Species -
HOMO SAPIENS(68)
UNKNOWN(4)
UNIDENTIFIED(1)

Select

List Manager Help

Uploaded List_1

Select List to:

Use | Rename | Remove | Combine | Show Gene List new!

View Unmapped Ids

Convert the gene list being selected in left panel to REFSEQ_MRNA

Submit

Gene ID Conversion Tool Result

Right-click to Download the result Help

Submit Converted List to DAVID as a Gene List | Submit Converted List to DAVID as a Background

Gene Accession Conversion Statistics Help

Conversion Summary		
ID Count	In DAVID DB	Conversion
60	Yes	Successful
8 IDs	Yes	None
0 IDs	No	None
0 IDs	Ambiguous	Pending
Total Unique User IDs: 68		

Summary of Ambiguous Gene IDs		
ID Count	Possible Source	Convert All

All Possible Sources For Ambiguous IDs		
Ambiguous ID	Possibility	Convert

From	To	Species	David Gene Name
202	NM_001624	HOMO SAPIENS	ABSENT IN MELANOMA 1
72	NM_001613	HOMO SAPIENS	ACTIN, ALPHA 2, SMOOTH MUSCLE, AORTA
72	NM_001615	HOMO SAPIENS	ACTIN, ALPHA 2, SMOOTH MUSCLE, AORTA
27299	NM_014479	HOMO SAPIENS	ADAM-LIKE, DECYSIN 1
125	NM_000667	HOMO SAPIENS	ALCOHOL DEHYDROGENASE 1A (CLASS I), ALPHA POLYPEPTIDE
125	NM_000668	HOMO SAPIENS	ALCOHOL DEHYDROGENASE 1A (CLASS I), ALPHA POLYPEPTIDE
126	NM_000669	HOMO SAPIENS	ALCOHOL DEHYDROGENASE 1A (CLASS I), ALPHA POLYPEPTIDE
126	NM_000668	HOMO SAPIENS	ALCOHOL DEHYDROGENASE 1A (CLASS I), ALPHA POLYPEPTIDE
125	NM_000669	HOMO SAPIENS	ALCOHOL DEHYDROGENASE 1A (CLASS I), ALPHA POLYPEPTIDE

Exercise 4: Convert affymetrix probeset IDs to gene symbols

Exercise 5: What are the enriched pathways and diseases for this gene set? Compare your results with ToppGene.

From the same example data set ("GeneLists.xls"), use the probe set IDs (1st column) and extract their RefSeq accession numbers

PANTHER (http://www.pantherdb.org/)

Protein ANalysis THrough Evolutionary Relationships

PANTHER Classification System

LOGIN REGISTER CONTACT US HELP

Home | Browse | Genes | Families and HMMs | Pathways | Ontologies | Tools | Workspace

Quick links

- Browse PANTHER
- Search PANTHER
- Batch search
- Browse pathways
- Community Curation
- My Workspace
- Gene expression tools
- HMM scoring
- cSNP analysis
- Downloads
- Site map

Find PANTHER-classified genes, transcripts, and proteins by uploading a list of IDs

Batch ID Search

Enter IDs: ACADSB, ACSL4, ADH6

separate IDs by a space or comma - supported IDs

Upload IDs: Browse...

- file format

Select upload ID type: Gene Symbol

Select File Type: ID List / Previously exported text search results

Result page: Genes / Transcripts/Proteins

Select datasets:

Celera:	H. sapiens	M. musculus	R. norvegicus
NCBI:	H. sapiens	M. musculus	R. norvegicus
FlyBase:	D. melanogaster		

Search

GENE EXPRESSION DATA ANALYSIS

Our expression analysis tools can be used for microarray data intrepretation. Multiple gene lists can be mapped to PANTHER molecular function and biological process categories, as well as to biological pathways. Our pathway visualization tool will display your experimental results on detailed diagrams of the relationships between genes/proteins in known pathways.

- **Compare gene lists**
  Upload lists of genes or gene products and statistically compare them to a reference list to look for under- and over-represented functional categories.
- **Analyze a list of genes with expression values**
  Upload a list of genes and their corresponding fold-change values from a differential expression experiment.

You can compare ***multiple*** lists!

Compare Classifications of Lists

Map lists of genes to a PANTHER ontology.For pathways, you can then view the gene expression values overlaid on top of a pathway diagram, where genes will be colored differently for different clusters of genes.

Use the binomial statistics tool to compare classifications of multiple clusters of lists to a reference list to statistically determine over- or under- representation of PANTHER classification categories. Each list is compared to the reference list using the binomial test (Cho & Campbell, TIGs 2000) for each molecular function, biological process, or pathway term in PANTHER.

Steps:

1. Select list(s) to analyze
2. Select reference list

1. Select Lists to Compare to a Reference List

For example, each selected list may be a cluster of co-expressed genes under a particular set of conditions.

Select list(s)

selected: FetalLiverSpecific.txt
FetalBrainSpecific.txt
AdultHeartSpecific.txt

2. Select Reference List

For example, the reference list may be the set of all genes in the experiment, or the set of all genes in the genome being analyzed.

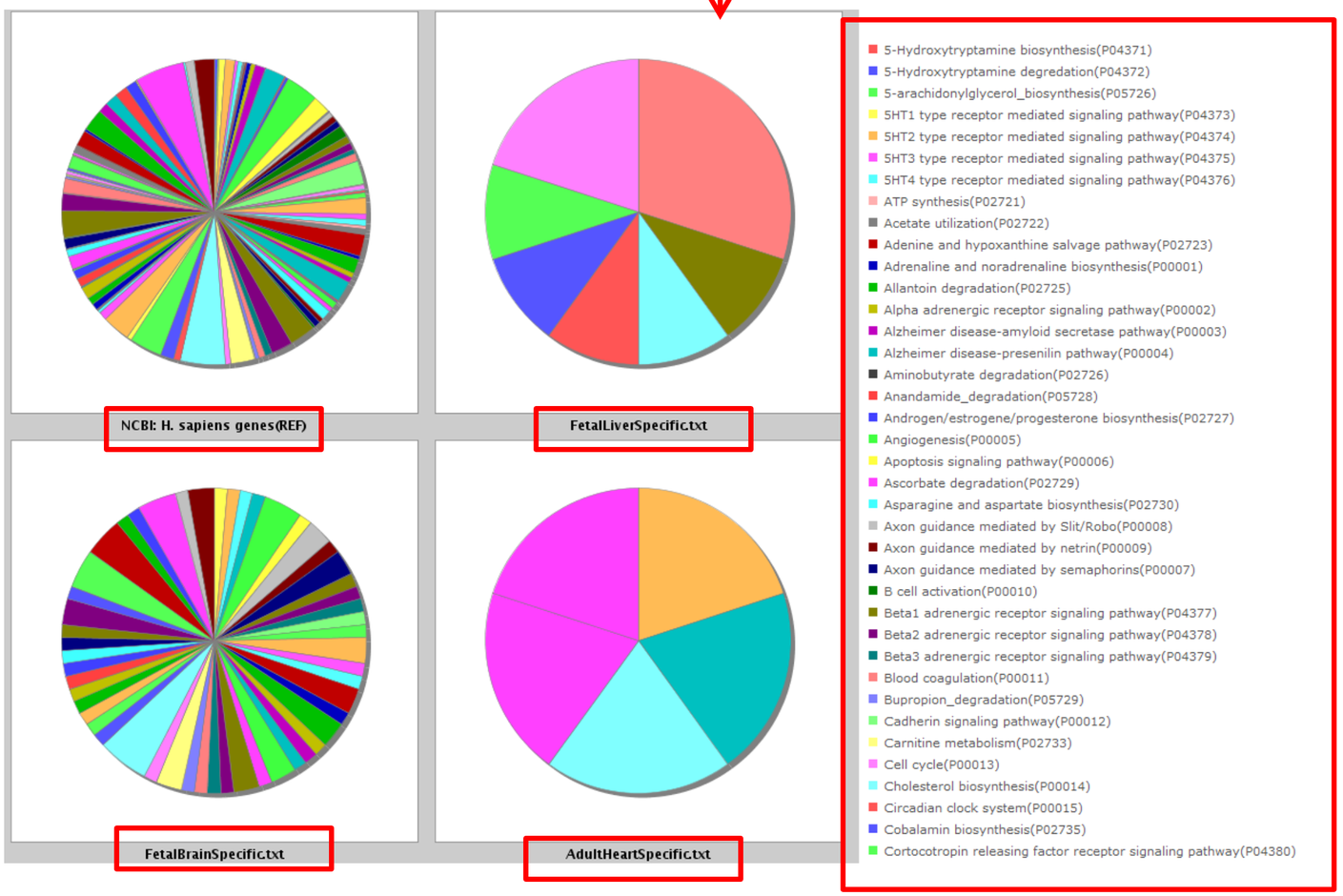
Select reference list default: NCBI: H. sapiens genes

Search options

PANTHER Ontology:

- (•) Pathways
- ( ) Biological Process
- ( ) Molecular Function

[✓] Use the Bonferroni correction for multiple testing



PANTHER (http://www.pantherdb.org/)

Protein ANalysis THrough Evolutionary Relationships

Results

Colors for viewing genes in pathway diagrams:

Example-Set-3.txt: red

gray: components only in the reference list

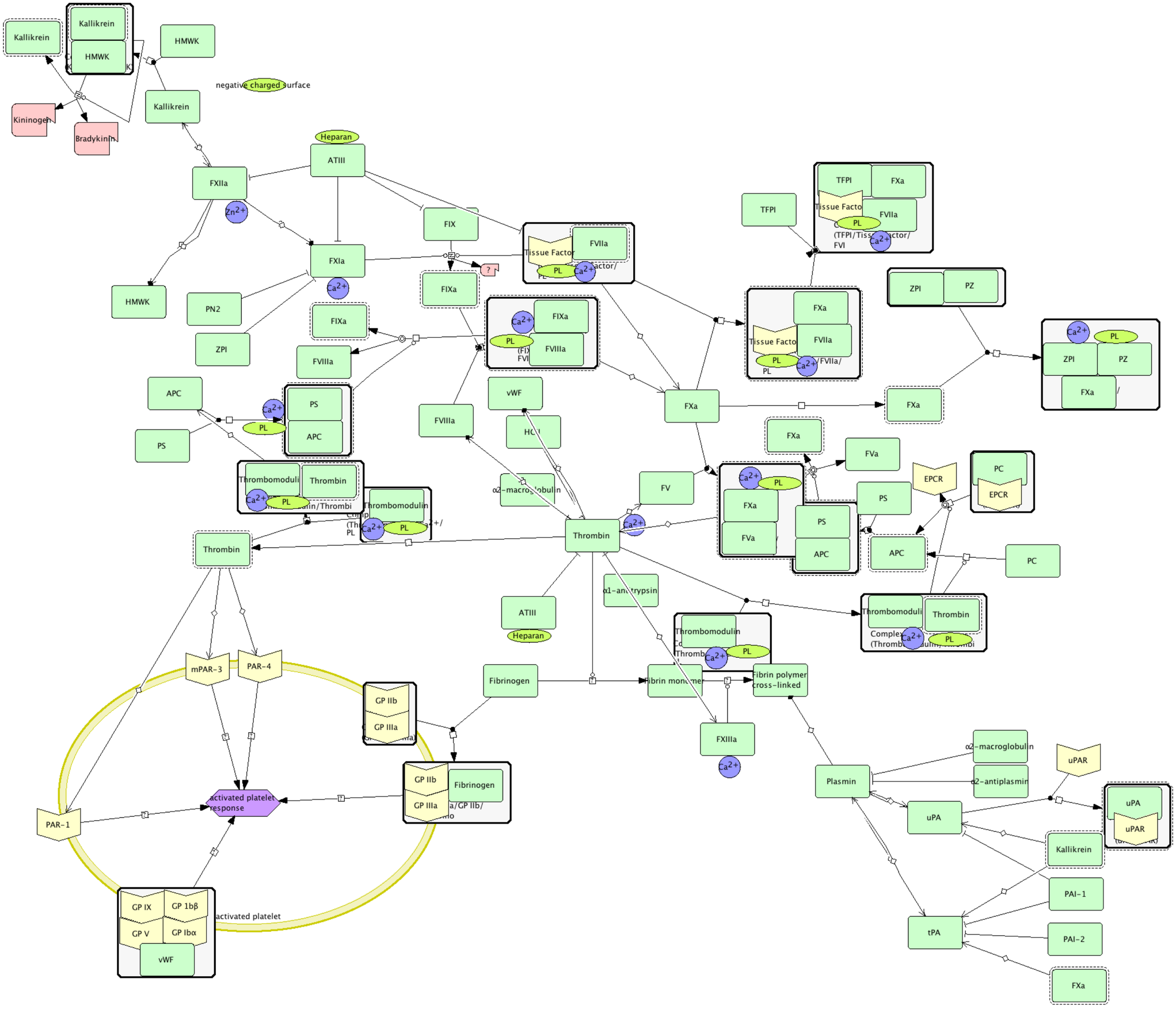
	Reference list	Example-Set-3.txt
Mapped IDs:	25431	135
Unmapped IDs:	0	15

Click on pathway name to see genes highlighted on pathway diagram

Export results View: Bar Chart of Gene Count

Pathways	NCBI: H. sapiens genes (REF) #	Example-Set-3.txt #	expected	+/-	P value
Blood coagulation	55	8	.29	+	1.38E-07
Plasminogen activating cascade	21	4	.11	+	9.37E-04

PANTHER (http://www.pantherdb.org/)



Gene Prioritization Tools

	Text (coocurrance)	Text (functional mining)	PPIs	Functional Annotations	Pathways	Expression	Sequence	Phenotype	Conservation/ Homology	Regulation	Disease	Drugs/Chemical
SUSPECT				x		x	x					
ToppGene	x		x	x	x	x	x	x		x	x	x
PolySearch		x	x	x	x							x
MimMiner		x	x	x			x	x				
PhenoPred			x	x			x	x				
PGMapper		x										
Endeavour		x	x	x	x	x	x			x	x	
G2D	x		x	x			x					
TOM				x		x						
SNPs3D		x	x	x	x		x	x				
GenTrepid		x	x		x		x					
GeneWanderer			x									
Bitola		x										
CANDID		x	x			x	x		x		x	
aGeneApart		x										
GeneProspector		x										
PosMed	x	x	x	x				x				x
GeneDistiller	x		x	x	x	x		x				

Adapted from Gene Prioritization Portal:
http://homes.esat.kuleuven.be/~bioiuser/gpp/index.php

RESOURCES - URLs: Summary

Application/Resource	URL
ToppGene	http://toppgene.cchmc.org
ToppCluster	http://toppcluster.cchmc.org
DAVID	http://david.abcc.ncifcrf.gov
PANTHER	http://www.pantherdb.org

Exercises - Summary

1. **Exercise 1**: Use the gene list from the downloaded file ("GeneLists.xls") and find out:
   - How many of the liver-overexpressed genes are associated with lipid metabolic process?
   - Are there any enriched TFBSs for liver overexpressed genes?
   - What are the enriched miRNAs in the colon-cecum overexpressed genes?
   - What gene families are enriched in esophagus overexpressed genes?
   - In which other regions are stomach (cardiac) genes overexpressed?
   - What biological process are miR-1 target genes enriched for?
2. **Exercise 2:** Use the different gene lists from the downloaded file ("GeneLists.xls") and find out:
   - What are the shared and specific biological processes between stomach and salivary glands?
   - Are there any enriched miRNAs for stomach? If so, which other tissues are enriched for this miRNA?
   - What are the functional similarities and differences between the 3 regions of the stomach (cardiac, fundus, and pylorus)?
3. **Exercise 3**: Prioritize the 721 genes ("CandidateGenes") using "stomach genes" from the "GeneLists.xls".
   - What are the top 10 ranked genes using ToppGene and ToppNet?
   - What is the rank of TFF3 and why is it ranked amongst the top? What is its rank in ToppNet?
4. **Exercise 4:** Convert affymetrix probeset IDs to gene symbols
5. **Exercise 5:** What are the enriched pathways and diseases for this gene set? Compare your results with ToppGene.

For additional exercises, see http://anil.cchmc.org/dhc.html