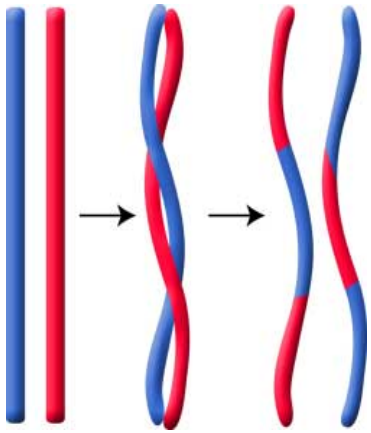
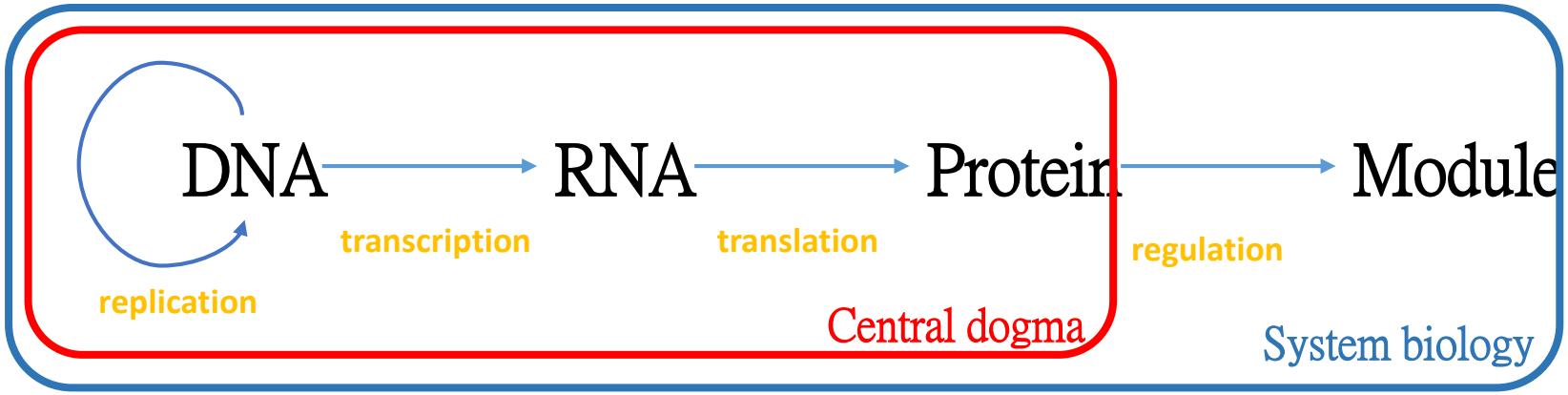


Transcriptome Functional Profiling

P. Tang (鄧致剛); YM Yeh (葉元鳴)
Genomic Medicine core Laboratory, CGMH
Bioinformatics Center, Chang Gung University.



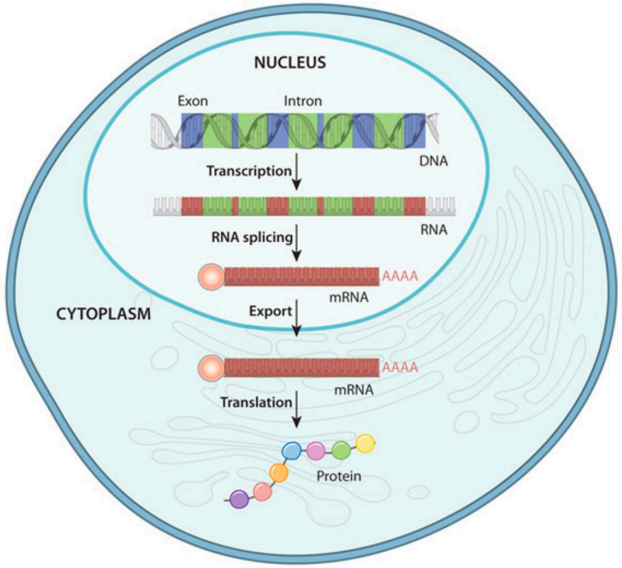


Figure 1: An overview of the flow of information from DNA to protein in a eukaryote
First, both coding and noncoding regions of DNA are transcribed into mRNA. Some regions are removed (introns) during initial mRNA processing. The remaining exons are then spliced together, and the spliced mRNA molecule (red) is prepared for export out of the nucleus through addition of an endcap (sphere) and a polyA tail. Once in the cytoplasm, the mRNA can be used to construct a protein.

© 2010 Nature Education All rights reserved.

Gene expression and regulation

- Differentially expressed
  - Different cell types (muscle cells, fibroblasts)
  - Environmental conditions (heat shock, nutrient deprivation)
  - Developmental phases (embryonic day 12)
  - Cell-cycle stages (S phase)
  - Disease states (tumor cells, virus-infected cells)
- Transcription level (primarily regulation
  - Determine the number of mRNA copies

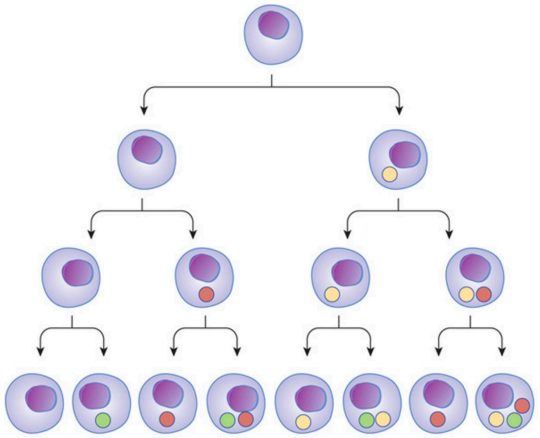
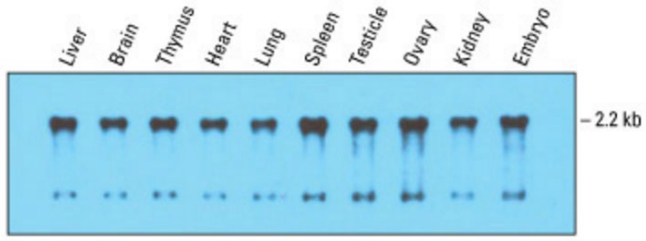
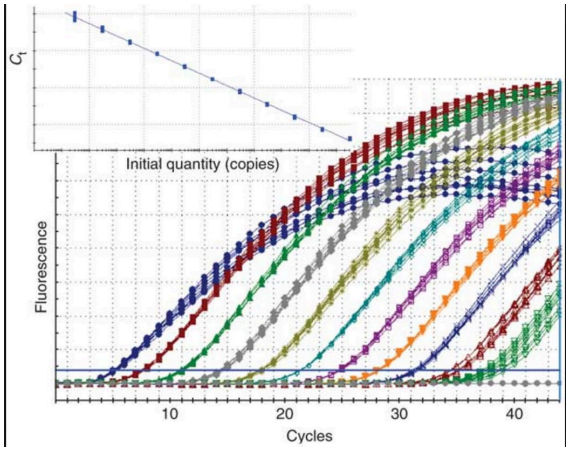


Figure 5: Transcriptional regulators can determine cell types
The wide variety of cell types in a single organism can depend on different transcription factor activity in each cell type. Different transcription factors can turn on at different times during successive generations of cells. As cells mature and go through different stages (arrows), transcription factors (colored balls) can act on gene expression and change the cell in different ways. This change affects the next generation of cells derived from that cell. In subsequent generations, it is the combination of different transcription factors that can ultimately determine cell type.
© 2010 Nature Education All rights reserved.

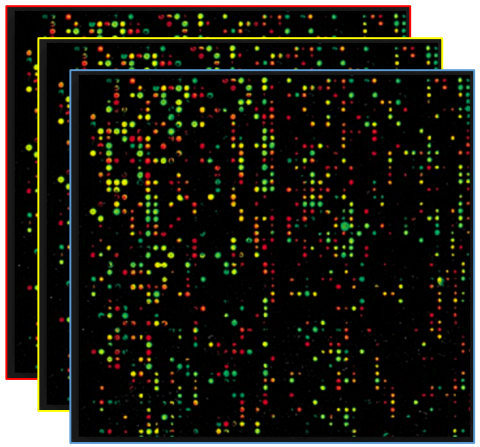
Evolution of transcriptomics technologies



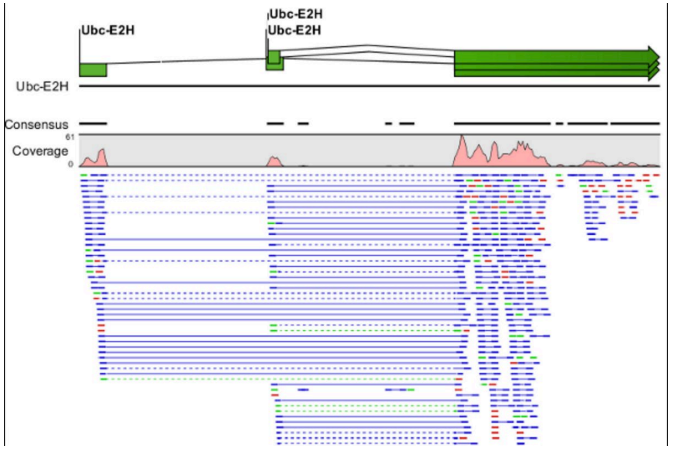
Northern Blot



Real time RT-PCR

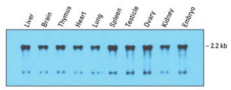


Microarray

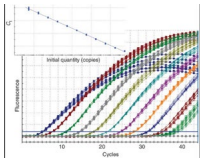


RNA seq

Traditional analysis-1



Northern Blot



Real time RT-PCR

GNAQ
GNAS
DGKZ
GUCY1A3
PDE4B
PDE4D
ATP2A2
ATP2A3
NOS1
CNN1
GSTO1
NOS3
CNN2
MYLK2
CALD1
ACTA1
MYL2

my favorite gene

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed GNAQ

RSS Save search Advanced

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Recently A

Article types
Review
More ...

Text availability
Abstract available
Free full text available
Full text available

Publication dates
5 years

See 225 articles about **GNAQ** gene function
See also: **GNAQ** guanine nucleotide binding protein (G protein), q
gnaq in Homo sapiens | Mus musculus | Rattus norvegicus | All

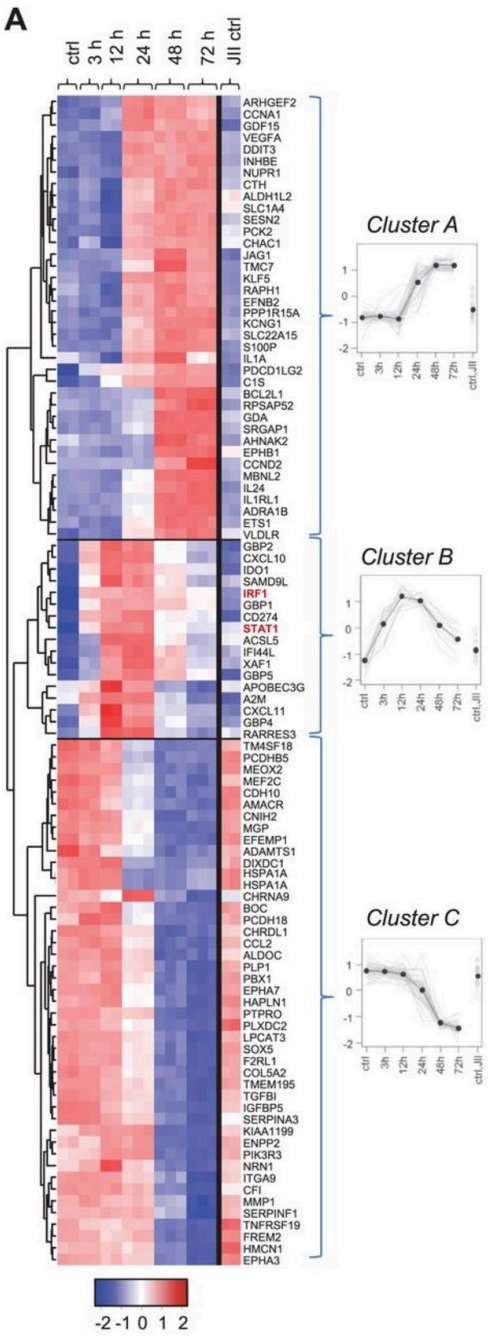
Results: 1 to 20 of 114

Sturge-Weber Syndrome and Port-Wine Stains Caused b
1. Shirley MD, Tang H, Gallione CJ, Baugher JD, Frelin LP, AM, Pevsner J.
...ay 8. [Epub ahead of print]
...Med - as supplied by publisher]



to Gary Bader

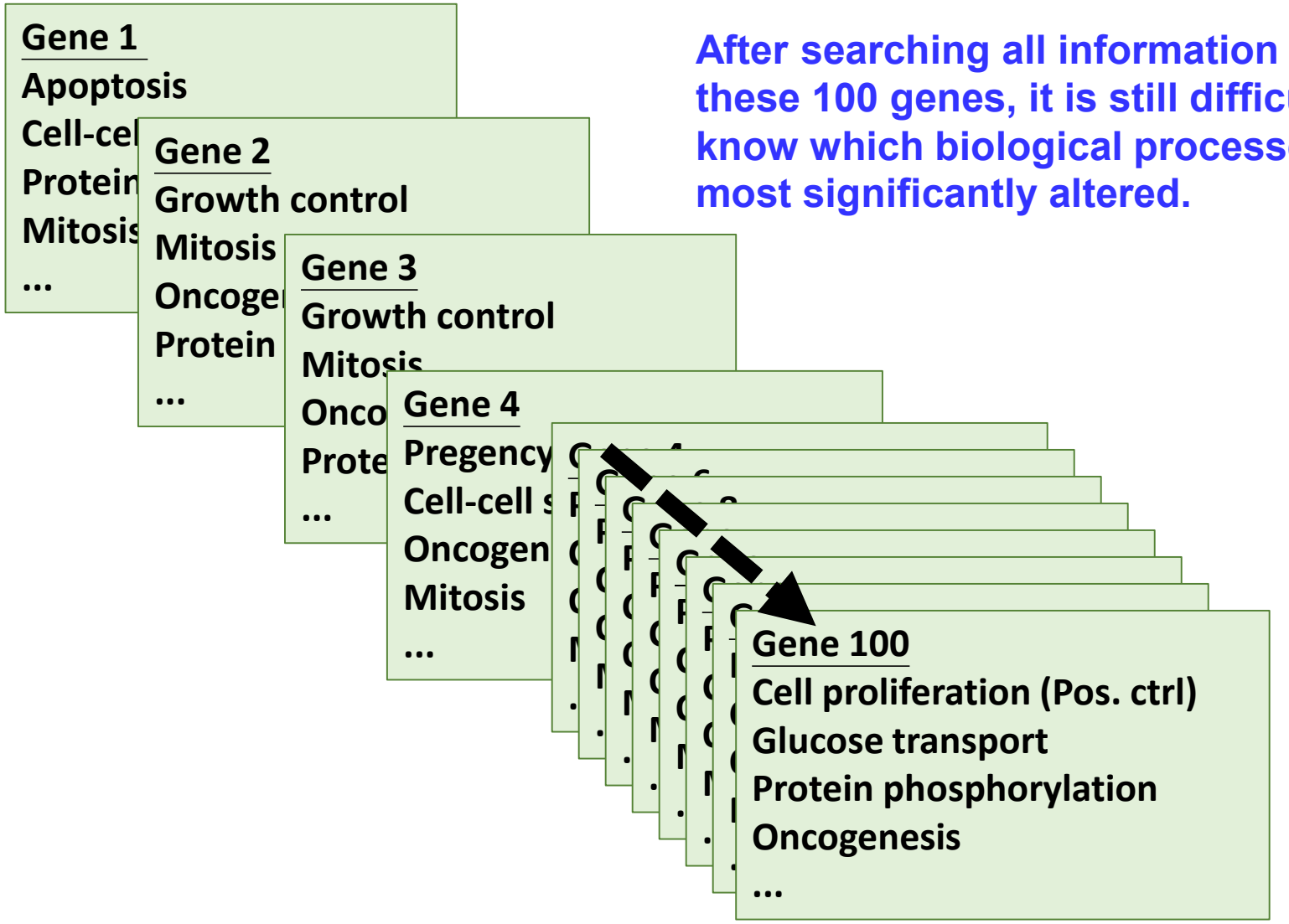
High throughput data



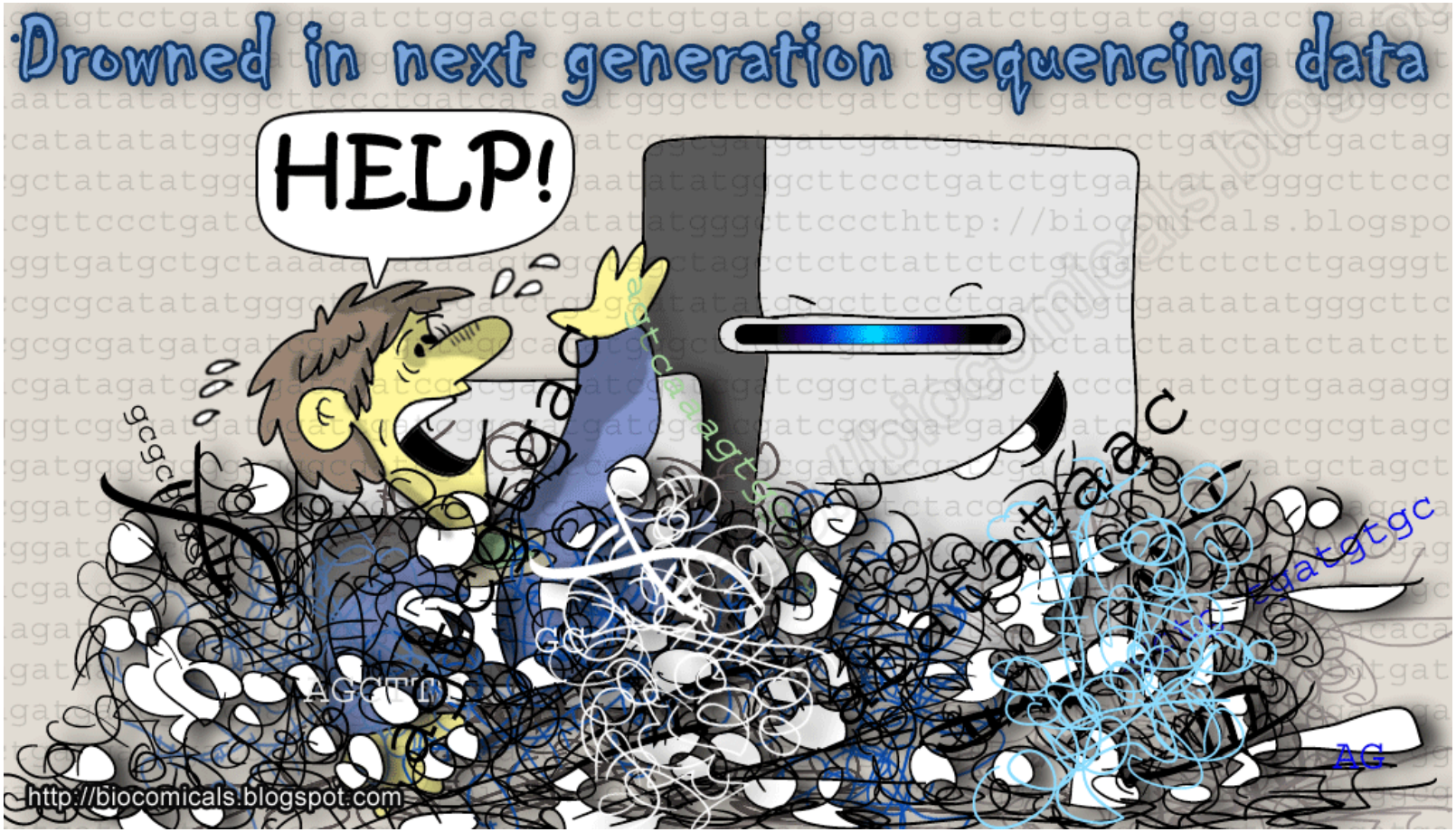
HT data shows changed expression of thousands of genes.

How will you spot the patterns?

Traditional analysis-2



After searching all information about these 100 genes, it is still difficult to know which biological processes are most significantly altered.


Drowned in next generation sequencing data
HELP!
http://biocomicals.blogspot.com

Gene Ontology (GO)

Gene Ontology Consortium

the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

- **molecular function(what)**

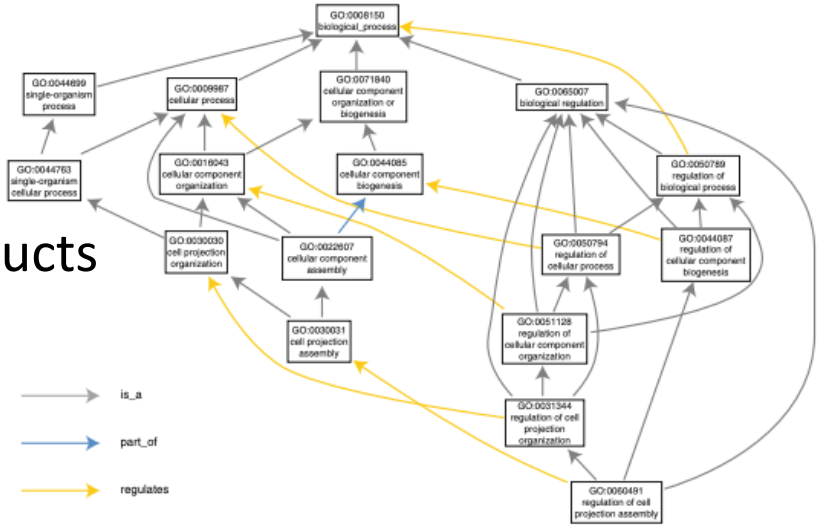
  molecular activities of gene products

- **cellular component(where)**

  where gene products are active

- **biological process(why)**

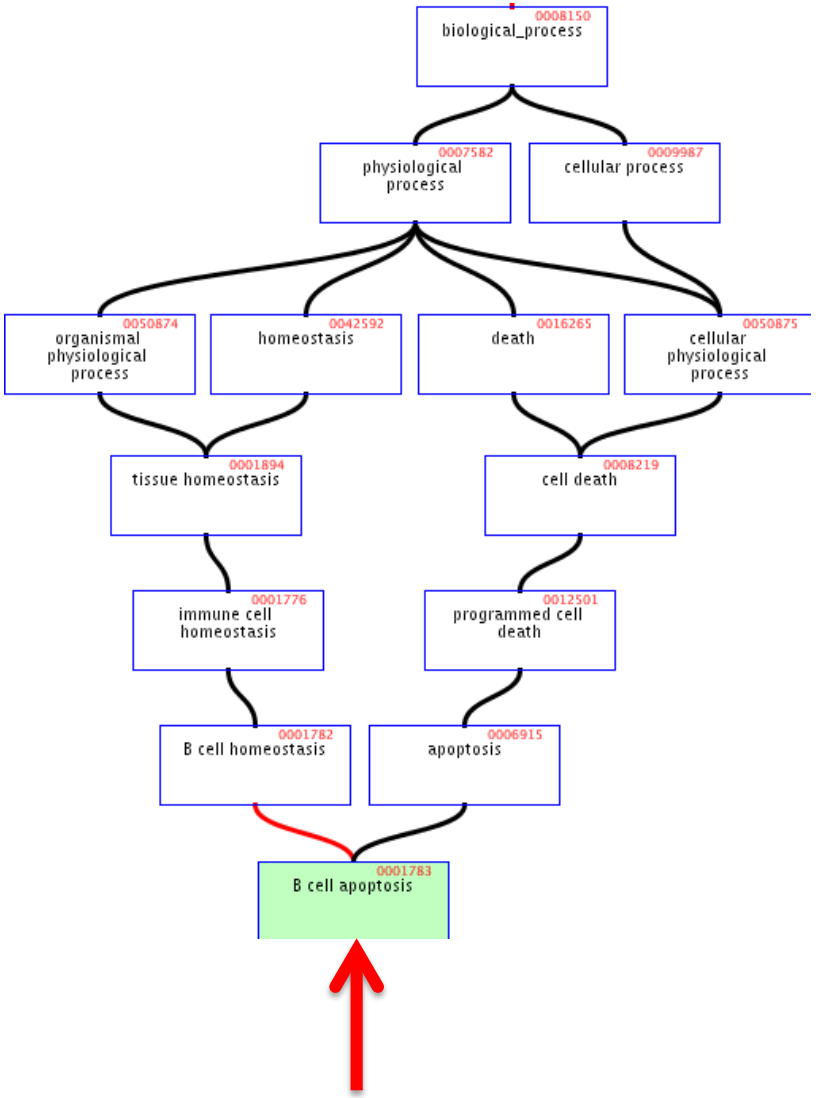
  pathways and larger processes made up of the activities of multiple gene products.



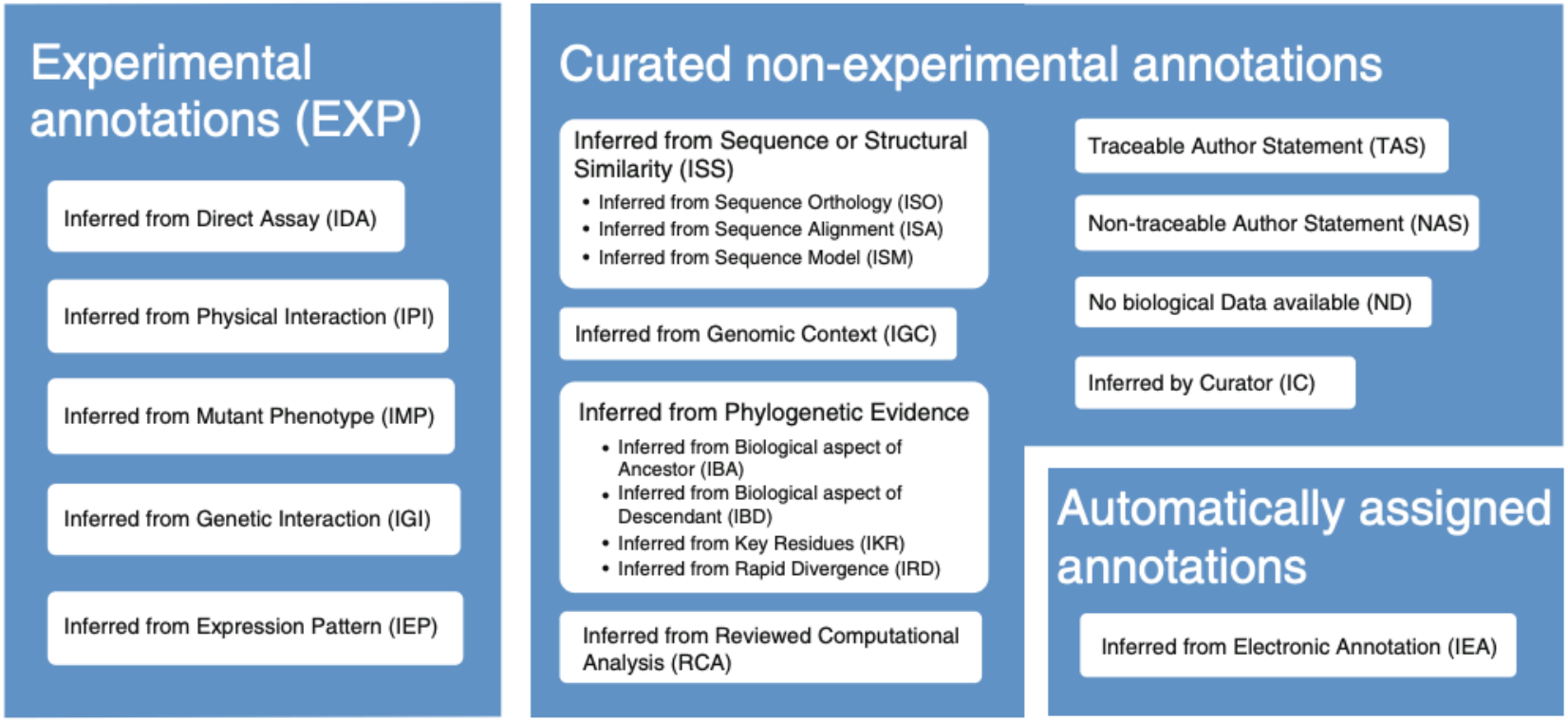
www.geneontology.org

GO annotations

- the model of biology.
- Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology.
- The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence.



Evidence Code of GO



http://geneontology.org/page/guide-go-evidence-codes

Gene Ontology Consortium | Home | Documentation | Downloads | Tools | About | Contact us

Enrichment analysis

Your gene IDs here...

biological process

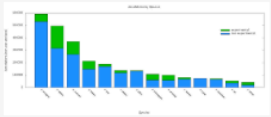
Homo sapiens

Submit

Help

Powered by PANTHER

Statistics



Other GOC tools

Explore other GOC tools in the AmiGO software suite.

Tweets about "#geneontology OR @news4go"

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products
cellular component
where gene products are active
biological process
pathways and larger processes made up of the activities of multiple gene products.
more

Annotations

Download annotations (standard files)

Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. more

Recent news

UCL Functional Gene Annotation Newsletter
Post date: 09/03/2018 - 06:52
Cardiovascular Gene Ontology Annotation Newsletter July 2018
Post date: 07/31/2018 - 06:58
Announcement: Change in GO xrefs to Reactome
Post date: 07/02/2018 - 14:49
Neurological Gene Ontology Annotation Newsletter July 2018
Post date: 06/29/2018 - 09:33
Now in pre-print: Article introducing the microRNA GO annotation resource from University College London
Post date: 06/14/2018 - 02:01
more

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. more

Search documentation

Search

What is the Gene Ontology?

- An introduction to the Gene Ontology
- What are annotations?
- Enrichment analysis
- Downloads

http://www.geneontology.org/

Grouping Genes by Biological Process

Apoptosis
Gene 1
Gene 23

Mitosis
Gene 3
Gene 26
Gene 45
Gene 68
...

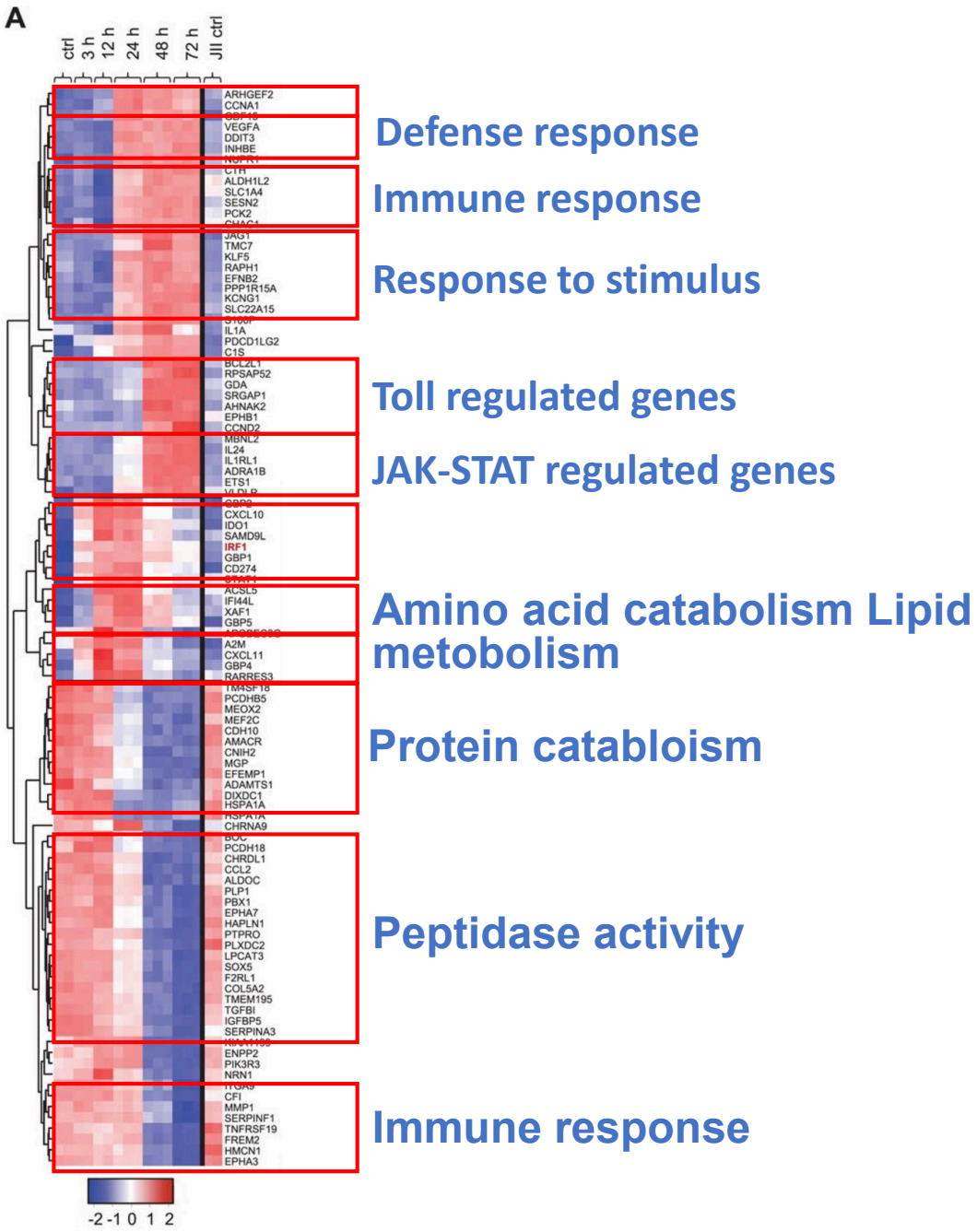
Glucose transport
Gene 5
Gene 9
Gene 11
Gene 35
Gene 66
Gene 79
...

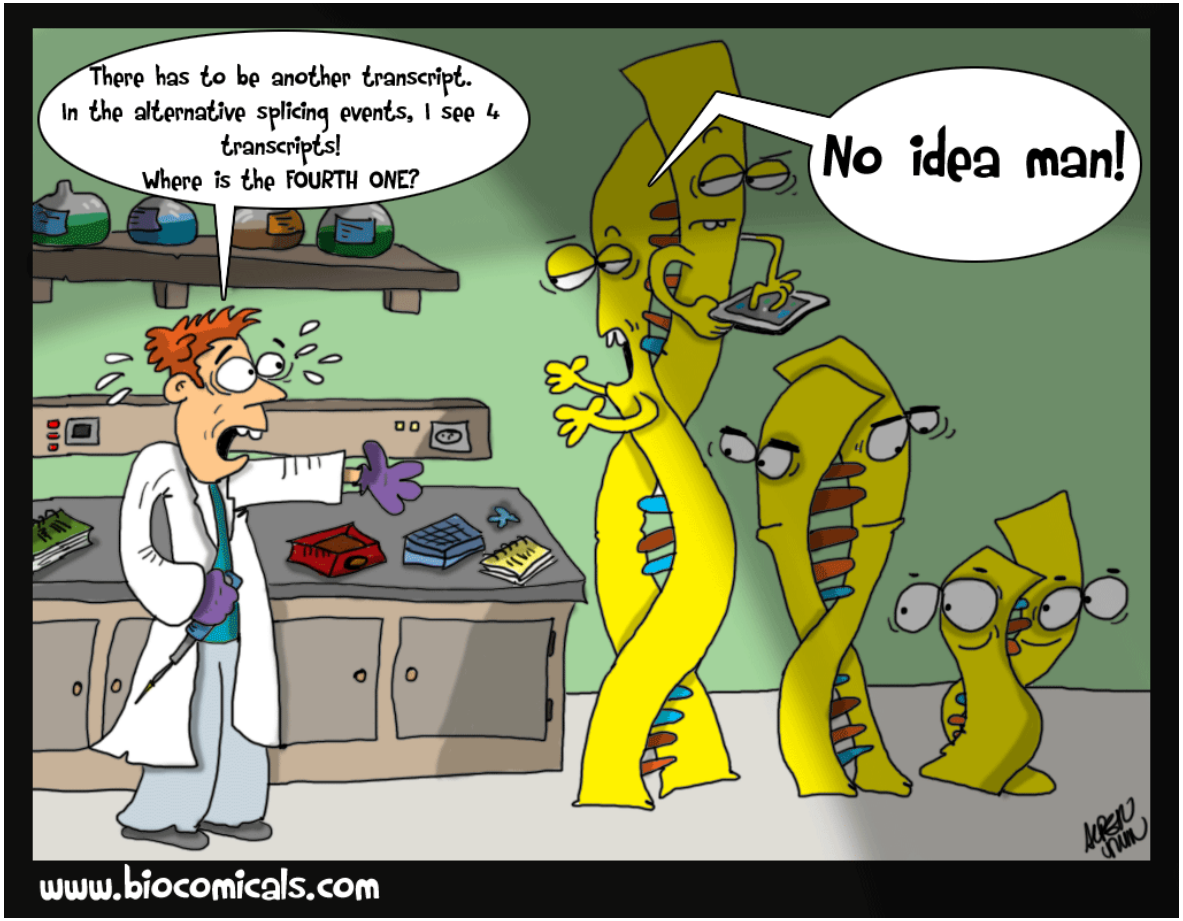
Protein phosphorylation
Gene 7
Gene 12
Gene 19
Gene 33
...

Cell-cell signaling
Gene 16
Gene 238
Gene 87
...

Cell proliferation
Gene 54
Gene 63
Gene 75
Gene 98
...

After GO annotation





How to spot biological functions embedded in a gene list?

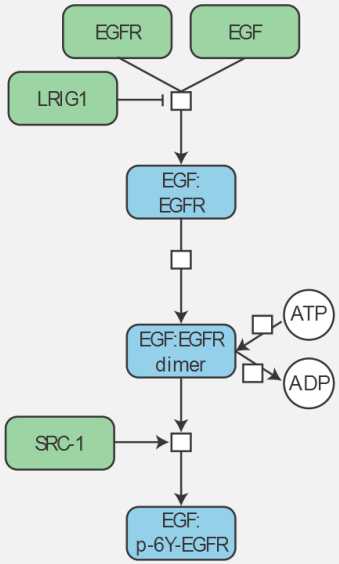
Pathway and Network Analysis

- Helps gain mechanistic insight into ‘omics data
  - Identifying a master regulator, drug targets, characterizing pathways active in a sample
- Any type of analysis that involves pathway or network information
- Most commonly applied to help interpret lists of genes
- Most popular type is pathway enrichment analysis, but many others are useful

to Gary Bader

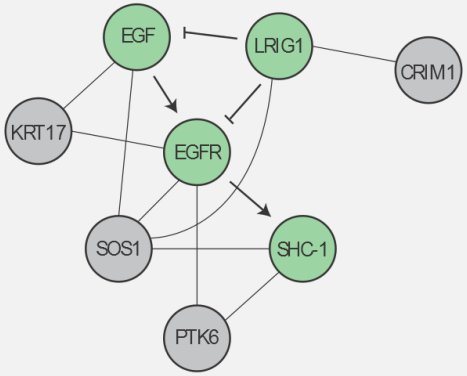
Pathways vs. Networks

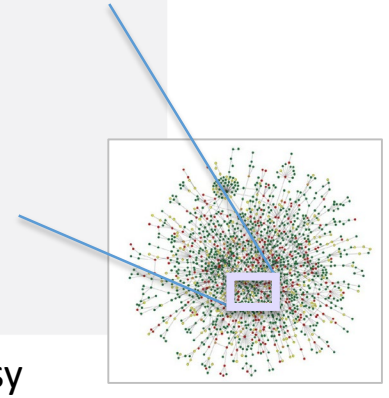
EGFR-centered Pathway



- Detailed, high-confidence consensus
- Biochemical reactions
- Small-scale, fewer genes
- Concentrated from decades of literature

EGFR-centered Network

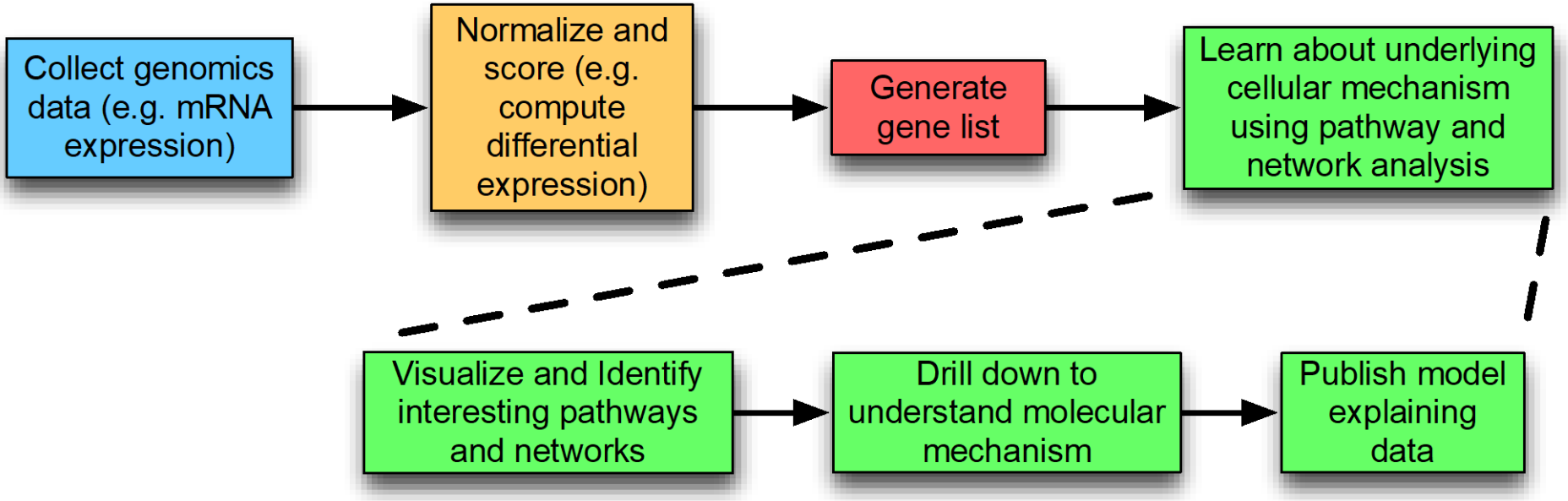




- Simplified cellular logic, noisy
- Abstractions: directed, undirected
- Large-scale, genome-wide
- Constructed from *omics* data integration

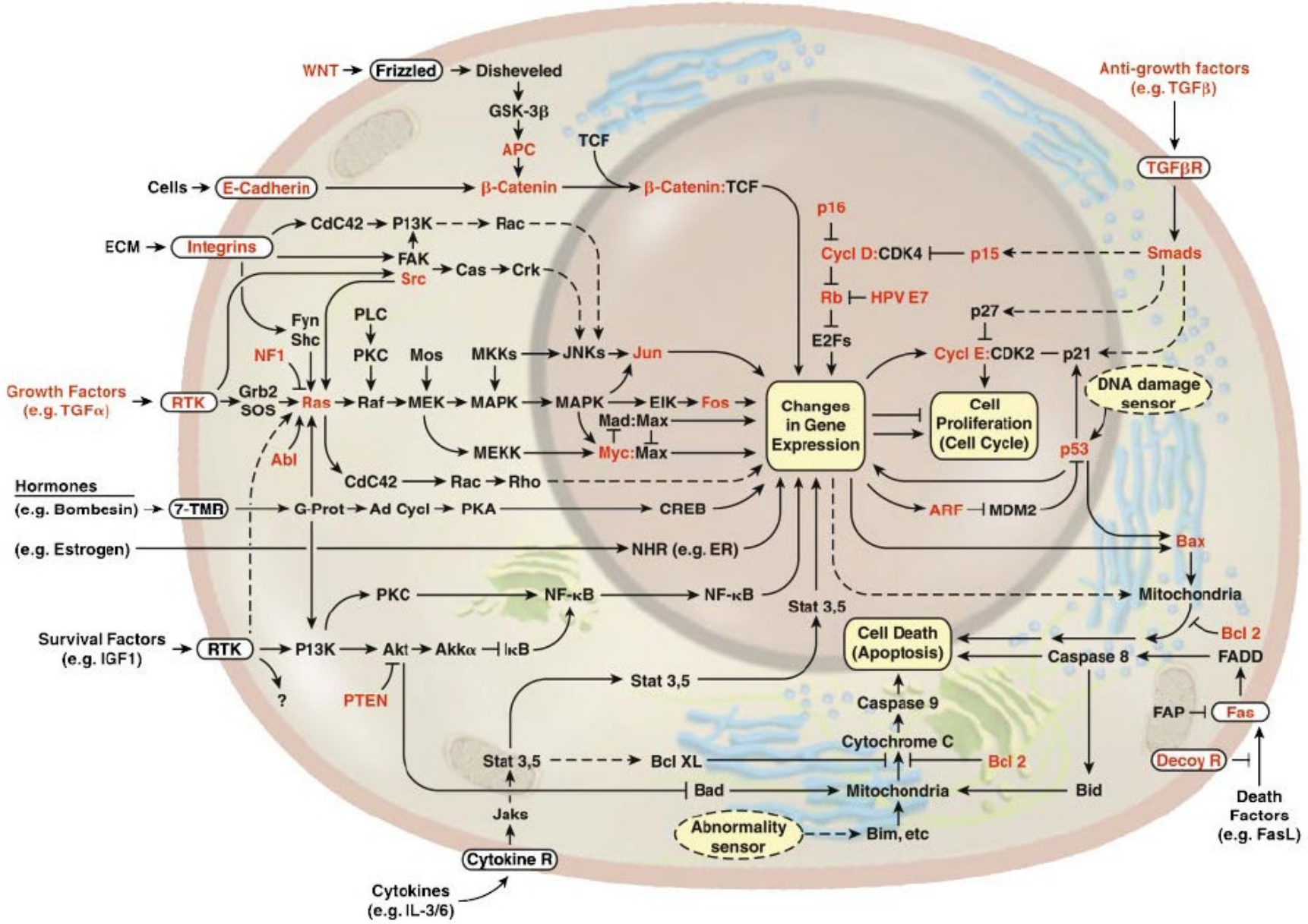
to Gary Bader

Pathway analysis workflow overview

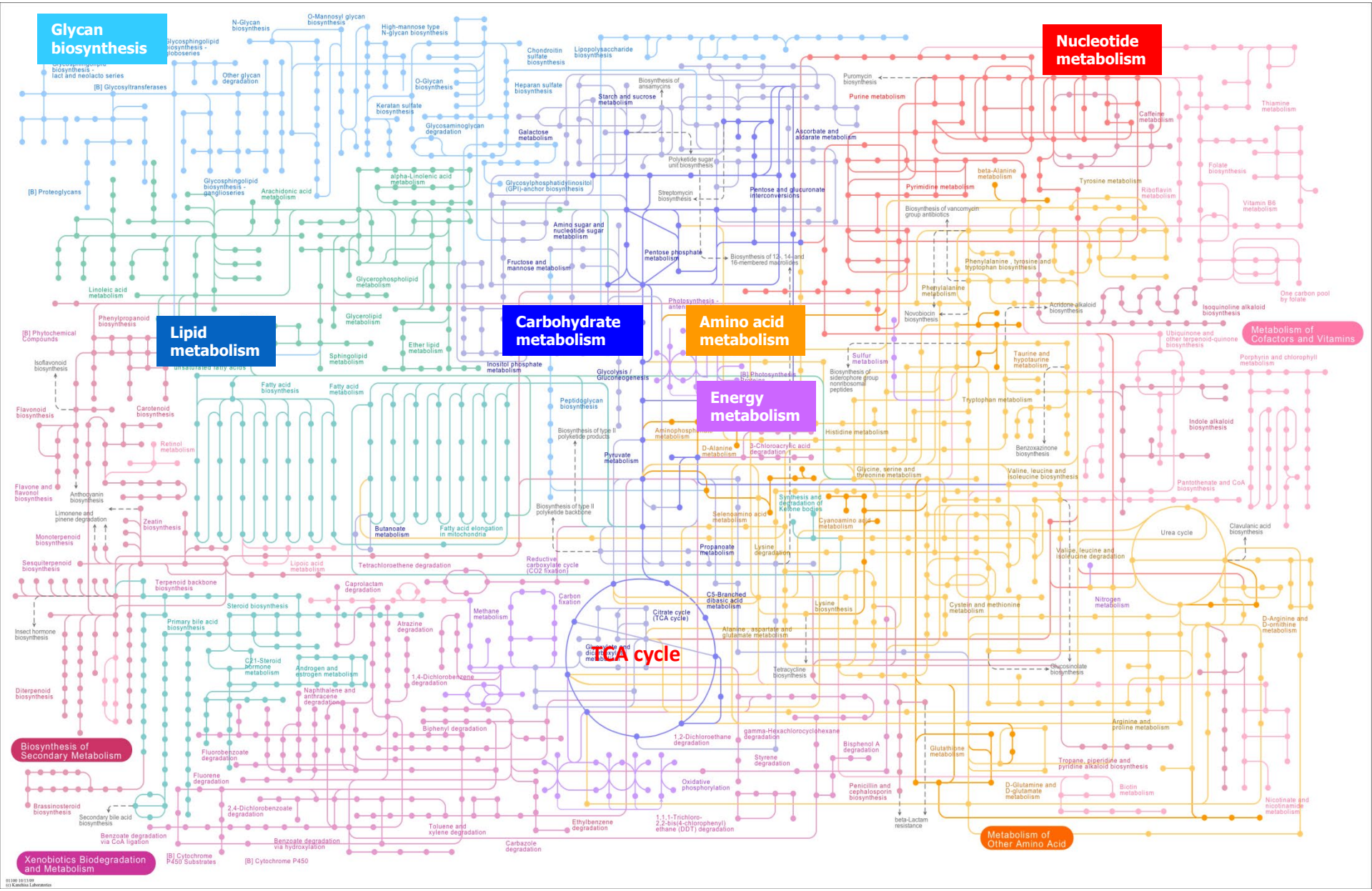


to Gary Bader

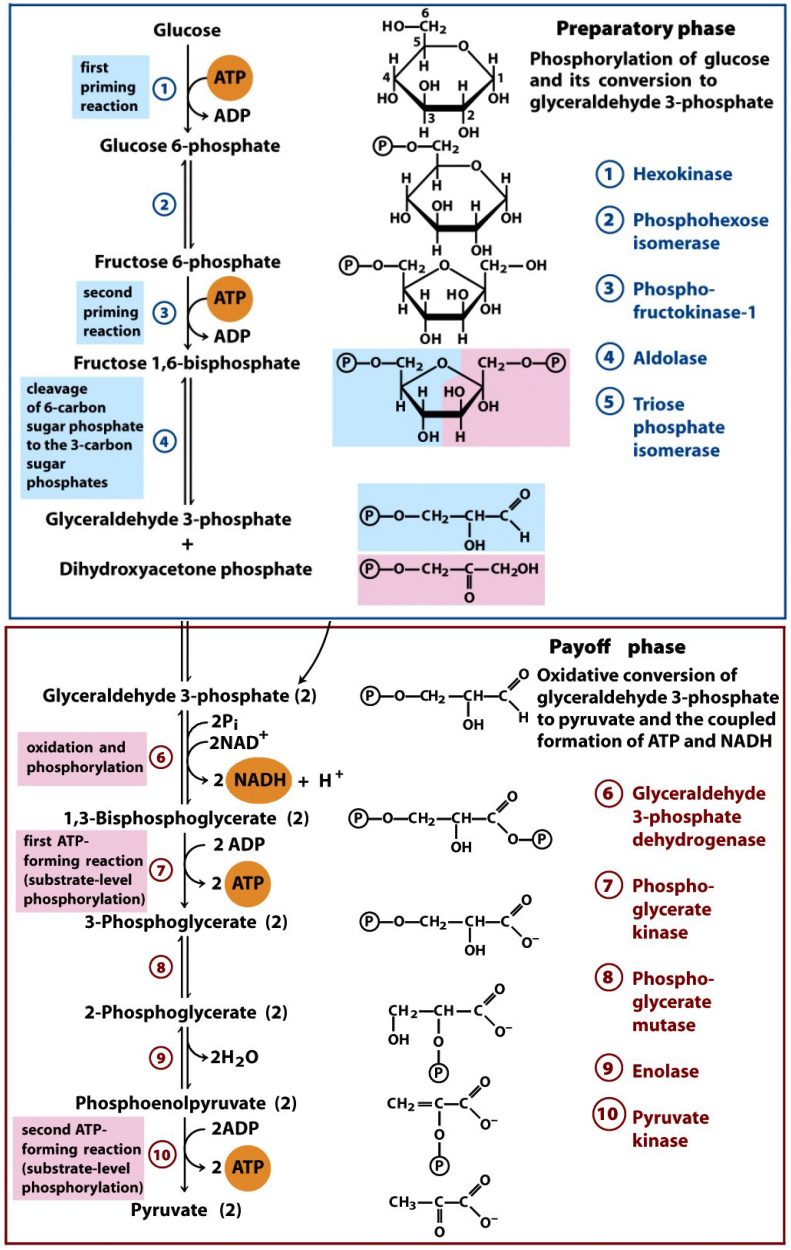
What's the reality in living cells: **Signaling Pathways**



What's the reality in living cells: **Cellular Metabolisms**



What's the reality in living cells: **Chemical Reactions**



Glycolysis pathways

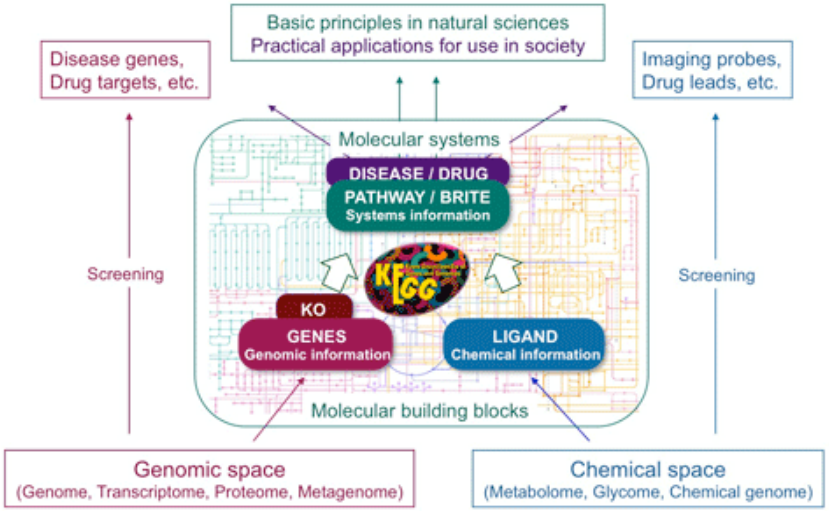
KEGG: Kyoto Encyclopedia of Genes and Genomes

Kanehisa Laboratories since 1995

- Manually create a knowledge base for such systemic functions by capturing and summarizing experimental knowledge in computable forms.
- To develop and improve the cross-species (ortholog-based) annotation procedure for linking genomes to the molecular networks.

- KEGG is a database of biological systems which contains:
  - Molecular building blocks of genes and proteins (*Genomic information*)
  - Chemical building blocks of both endogenous and exogenous substances (*Chemical information*)
  - Integrated with the knowledge on molecular wiring diagrams of interaction, reaction and relation networks (*Systems information*)

- KEGG provides a reference knowledge base for linking genomes to biological systems and also to environments by the processes of PATHWAY mapping and BRITE mapping.



KEGG is an integrated database resource consisting of the sixteen main databases subdivided into four categories

The KEGG Database

Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	
	KEGG GENOME	KEGG organisms (complete genomes)	
	KEGG GENES	Genes and proteins	
	KEGG SSDB	GENES sequence similarity	
Chemical information	KEGG COMPOUND	Small molecules	
	KEGG GLYCAN	Glycans	
	KEGG REACTION	Biochemical reactions	
	KEGG RCLASS	Reaction class	
	KEGG ENZYME	Enzyme nomenclature	
Health information	KEGG DISEASE	Human diseases	
	KEGG DRUG	Drugs	
	KEGG DGROUP	Drug groups	
	KEGG ENVIRON	Health-related substances	

Chemical information category is collectively called KEGG LIGAND
Health information category integrated with drug labels is called KEGG MEDICUS

Current statistics

KEGG Database as of 2016/10/19

Systems information		
KEGG PATHWAY	Pathway maps, reference (total)	498 (468,036)
KEGG BRITE	Functional hierarchies, reference (total)	231 (167,774)
KEGG MODULE	KEGG modules, reference (total)	764 (380,925)
Genomic information		
KEGG ORTHOLOGY	KEGG Orthology (KO) groups	20,206
KEGG GENOME	KEGG Organisms and selected viruses (346 eukaryotes, 3900 bacteria, 231 archaea, and 317 viruses)	4,794
KEGG GENES	Genes in KEGG Organisms and other categories (including 3,666 addendum, 273,878 viral)	20,355,507
KEGG SSDB	Best hit relations within GENES Bi-directional best hit relations within GENES	140,805,107,162 8,438,042,641
Chemical information		
KEGG COMPOUND	Metabolites and other small molecules	17,776
KEGG GLYCAN	Glycans	10,994
KEGG REACTION	Biochemical reactions	10,287
KEGG RCLASS	Reaction class	3,066
KEGG ENZYME	Enzyme nomenclature	6,791
Health information		
KEGG DISEASE	Human diseases	1,669
KEGG DRUG	Drugs	10,364
KEGG DGROUP	Drug groups	1,923
KEGG ENVIRON	Crude drugs and health-related substances	850

Pathway modules are subnetworks that are tighter functional units compared to pathway entries

KEGG Object Identifier

KEGG objects are biological entities from molecular to higher levels that are represented as database entries in KEGG, such as genes and protein, small molecules, reactions, pathways, diseases and drugs.

The KEGG object identifer or simply the **KEGG identifier** (kid) is a unique identifier for each KEGG object, which is also the database entry identifier. Generally it takes the form of a prefix followed a five-digit number as shown below:

Database	Object	Prefix	Example
KEGG PATHWAY	Pathway map	map, ko, ec, rn, (org)	hsa04930
KEGG BRITE	Functional hierarchy	br, jp, ko, (org)	ko01003
KEGG MODULE	KEGG module	M, (org)_M	M00010
KEGG ORTHOLOGY	KO	K	K04527
KEGG GENOME	KEGG organism	T	T01001 (hsa)
KEGG GENES	Gene / protein		hsa:3643
KEGG COMPOUND	Small molecule	C	C00031
KEGG GLYCAN	Glycan	G	G00109
KEGG REACTION	Reaction	R	R00259
KEGG RCLASS	Reaction class	RC	RC00046
KEGG ENZYME	Enzyme		ec:2.7.10.1
KEGG DISEASE	Human disease	H	H00004
KEGG DRUG	Drug	D	D01441
KEGG DGROUP	Drug group	DG	DG00710
KEGG ENVIRON	Crude drug, etc.	E	E00048

- **Each database entry is a KEGG object**
- **IDs for specific data types**
- **Use in KEGG or even in Google!**



http://www.kegg.jp

KEGG ▼ Search Help

» Japanese

KEGG Home
Release notes
Current statistics
Plea from KEGG

KEGG Database
KEGG overview
Searching KEGG
KEGG mapping
Color codes

KEGG Objects
Pathway maps
Brite hierarchies

KEGG Software
KegTools
KEGG API
KGML

KEGG FTP
Subscription

GenomeNet

DBGET/LinkDB

Feedback

Kanehisa Labs

KEGG: Kyoto Encyclopedia of Genes and Genomes

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.
See Release notes (October 1, 2016) for new and updated features.

Main entry point to the KEGG web service

KEGG2	KEGG Table of Contents [Update notes]

Data-oriented entry points

KEGG PATHWAY	KEGG pathway maps
KEGG BRITE	BRITE hierarchies and tables
KEGG MODULE	KEGG modules
KEGG ORTHOLOGY	KO functional orthologs
KEGG GENOME	Genomes [Release history]
KEGG GENES	Genes and proteins
KEGG COMPOUND	Small molecules
KEGG GLYCAN	Glycans
KEGG REACTION	Biochemical reactions
KEGG ENZYME	Enzyme nomenclature
KEGG DISEASE	Human diseases
KEGG DRUG	Drugs
KEGG MEDICUS	Health information resource [Drug labels search]

Subject-oriented entry points

KEGG Cancer
KEGG Pathogen
KEGG Virus
KEGG Plant
KEGG Annotation
KEGG RModule

Organism-specific entry points

KEGG Organisms Enter org code(s) Go hsa hsa eco

Analysis tools

KEGG Mapper	KEGG PATHWAY/BRITE/MODULE mapping tools
BlastKOALA	Genome annotation and KEGG mapping
GhostKOALA	Metagenome annotation and KEGG mapping
BLAST/FASTA	Sequence similarity search
SIMCOMP	Chemical structure similarity search

Copyright 1995-2016 Kanehisa Laboratories

Main database entry points

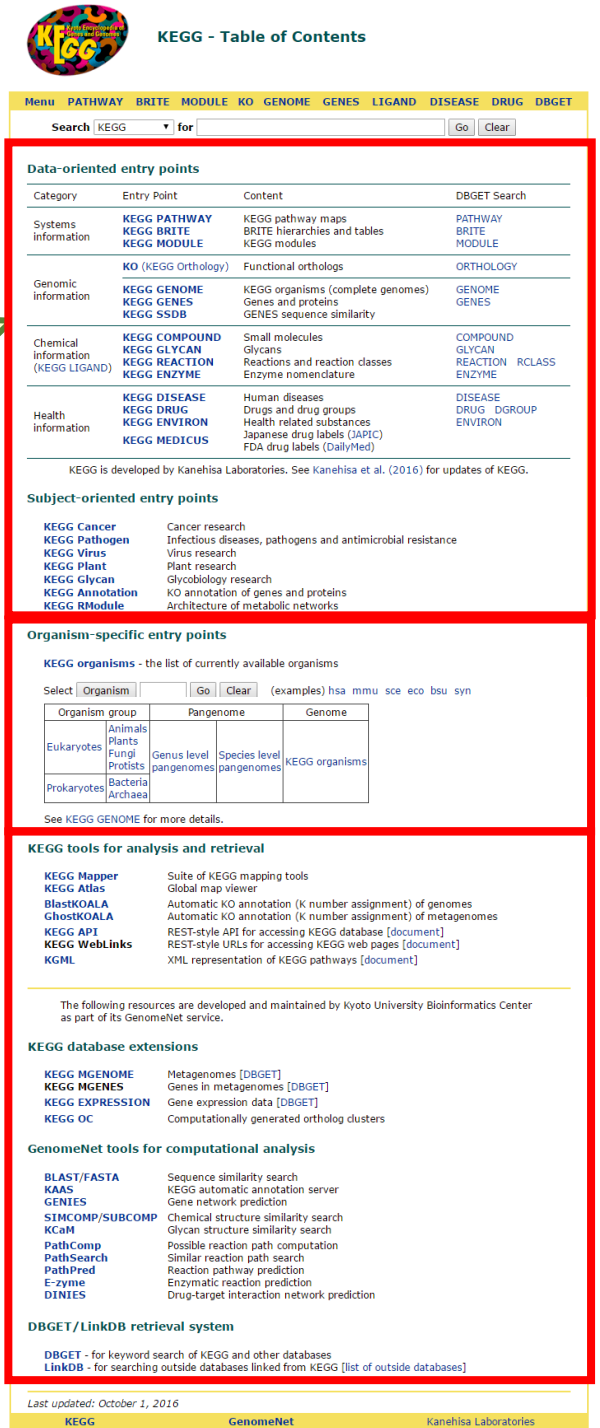
KEGG database: Table of contents

Systems information
Genomic information
Chemical information
Health information

Subject-oriented entry points

Organism-specific entry points

Web resources
Computational tools
Software development



- Access to KEGG data via “Entry Points”
- KEGG associated tools
  - Analysis tools
  - Software developers

http://www.kegg.jp/kegg/kegg2.html

KEGG database: Data-oriented entry points

Category	Entry Point	Content	DBGET Search
Systems information	**KEGG PATHWAY**	KEGG pathway maps	PATHWAY
	KEGG BRITE	BRITE hierarchies and tables	BRITE
	KEGG MODULE	KEGG modules	MODULE
Genomic information	**KO** (KEGG Orthology)	Functional orthologs	ORTHOLOGY
	KEGG GENOME	KEGG organisms (complete genomes)	GENOME
	KEGG GENES	Genes and proteins	GENES
	KEGG SSDB	GENES sequence similarity	
Chemical information (KEGG LIGAND)	**KEGG COMPOUND**	Small molecules	COMPOUND
	KEGG GLYCAN	Glycans	GLYCAN
	KEGG REACTION	Reactions and reaction classes	REACTION RCLASS
	KEGG ENZYME	Enzyme nomenclature	ENZYME
Health information	**KEGG DISEASE**	Human diseases	DISEASE
	KEGG DRUG	Drugs and drug groups	DRUG DGROUP
	KEGG ENVIRON	Health related substances	ENVIRON
	KEGG MEDICUS	Japanese drug labels (JAPIC) FDA drug labels (DailyMed)	

KEGG is developed by Kanehisa Laboratories. See Kanehisa et al. (2016) for updates of KEGG.

Total 15 main databases subdivided into four categories



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Menu PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix: map | Organism

Enter keywords: Search by key words | Go | Help

[New pathway maps | Update history]

Pathway Maps

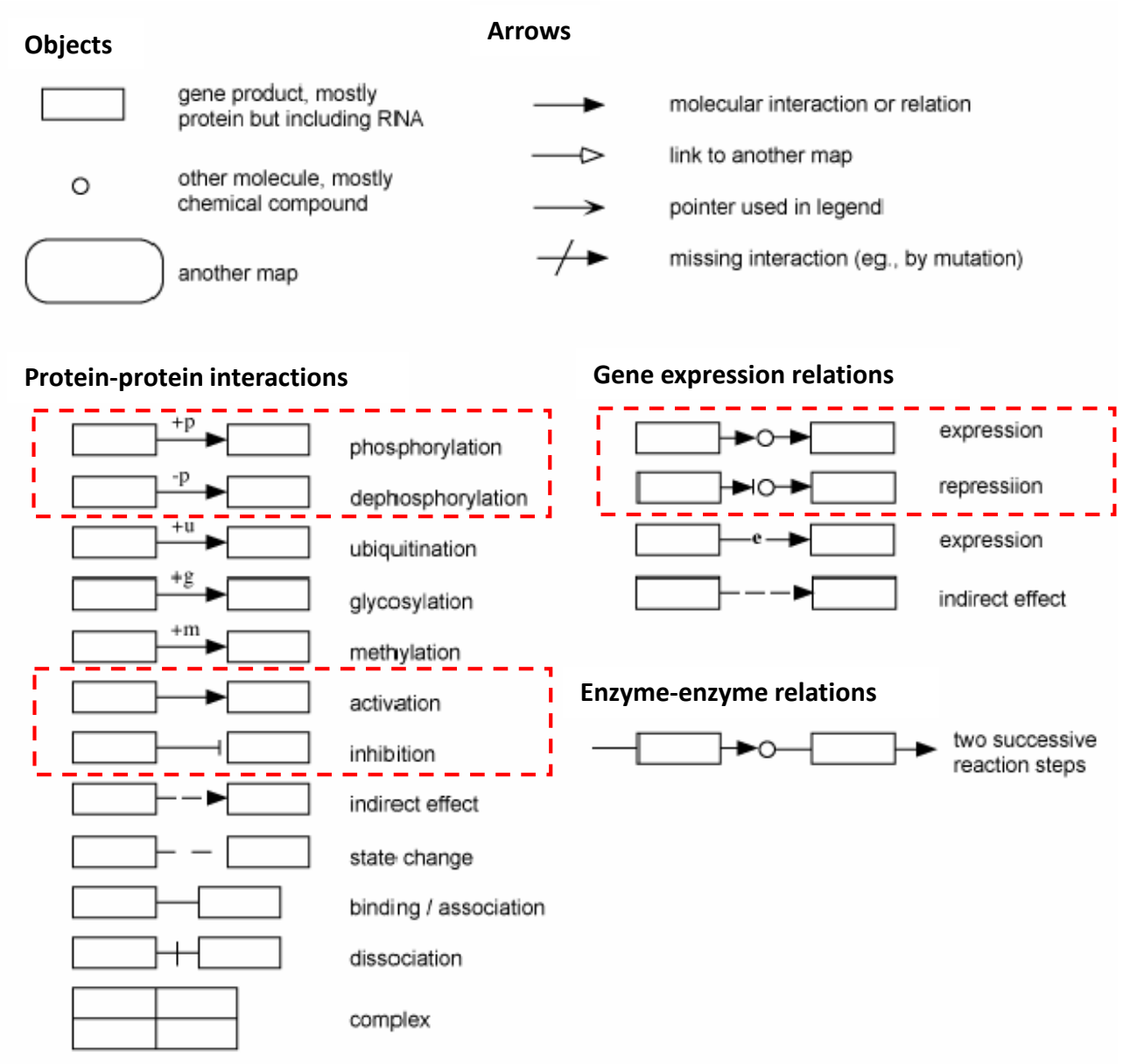
KEGG PATHWAY is a collection of manually drawn pathway maps representing our knowledge on the molecular interaction and reaction networks for:

1. **Metabolism**
   Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

7. **Drug Development**

Definition of KEGG symbols



Partial list of metabolism pathways

1. Metabolism

1.1 Carbohydrate Metabolism
- Glycolysis / Gluconeogenesis
- Citrate cycle (TCA cycle)
- Pentose phosphate pathway
- Pentose and glucuronate interconversions
- Fructose and mannose metabolism
- Galactose metabolism
- Ascorbate and aldarate metabolism
- Starch and sucrose metabolism
- Amino sugar and nucleotide sugar metabolism
- Pyruvate metabolism
- Glyoxylate and dicarboxylate metabolism
- Propanoate metabolism
- Butanoate metabolism
- C5-Branched dibasic acid metabolism
- Inositol phosphate metabolism

1.2 Energy Metabolism
- Oxidative phosphorylation
- Photosynthesis
- Photosynthesis - antenna proteins
- Carbon fixation in photosynthetic organisms
- Reductive carboxylate cycle (CO2 fixation)
- Methane metabolism
- Nitrogen metabolism
- Sulfur metabolism

1.3 Lipid Metabolism
- Fatty acid biosynthesis
- Fatty acid elongation in mitochondria
- Fatty acid metabolism
- Synthesis and degradation of ketone bodies
- Steroid biosynthesis
- Primary bile acid biosynthesis
- Secondary bile acid biosynthesis
- C21-Steroid hormone metabolism
- Androgen and estrogen metabolism
- Glycerolipid metabolism

1.4 Nucleotide Metabolism
- Purine metabolism
- Pyrimidine metabolism

1.5 Amino Acid Metabolism
- Alanine, aspartate and glutamate metabolism
- Glycine, serine and threonine metabolism
- Cysteine and methionine metabolism
- Valine, leucine and isoleucine degradation
- Valine, leucine and isoleucine biosynthesis
- Lysine biosynthesis
- Lysine degradation
- Arginine and proline metabolism
- Histidine metabolism
- Tyrosine metabolism
- Phenylalanine metabolism
- Tryptophan metabolism
- Phenylalanine, tyrosine and tryptophan biosynthesis

1.6 Metabolism of Other Amino Acids
- beta-Alanine metabolism
- Taurine and hypotaurine metabolism
- Aminophosphonate metabolism
- Selenoamino acid metabolism
- Cyanoamino acid metabolism
- D-Glutamine and D-glutamate metabolism
- D-Arginine and D-ornithine metabolism
- D-Alanine metabolism
- Glutathione metabolism

1.7 Glycan Biosynthesis and Metabolism
- N-Glycan biosynthesis
- High-mannose type N-glycan biosynthesis
- O-Glycan biosynthesis
- O-Mannosyl glycan biosynthesis
- Chondroitin sulfate biosynthesis
- Heparan sulfate biosynthesis
- Keratan sulfate biosynthesis
- Lipopolysaccharide biosynthesis
- Peptidoglycan biosynthesis

5. Organismal Systems

5.1 Immune system
- Hematopoietic cell lineage
- Complement and coagulation cascades *Major update!*
- Platelet activation
- Toll-like receptor signaling pathway
- NOD-like receptor signaling pathway
- RIG-I-like receptor signaling pathway
- Cytosolic DNA-sensing pathway
- Natural killer cell mediated cytotoxicity
- Antigen processing and presentation
- T cell receptor signaling pathway
- B cell receptor signaling pathway
- Fc epsilon RI signaling pathway
- Fc gamma R-mediated phagocytosis
- Leukocyte transendothelial migration
- Intestinal immune network for IgA production
- Chemokine signaling pathway

5.2 Endocrine system
- Insulin secretion
- Insulin signaling pathway
- Glucagon signaling pathway
- Regulation of lipolysis in adipocytes
- Adipocytokine signaling pathway
- PPAR signaling pathway
- GnRH signaling pathway
- Ovarian steroidogenesis
- Estrogen signaling pathway
- Progesterone-mediated oocyte maturation
- Prolactin signaling pathway
- Oxytocin signaling pathway
- Thyroid hormone synthesis
- Thyroid hormone signaling pathway
- Melanogenesis
- Renin secretion
- Renin-angiotensin system
- Aldosterone synthesis and secretion *New!*

5.3 Circulatory system
- Cardiac muscle contraction
- Adrenergic signaling in cardiomyocytes
- Vascular smooth muscle contraction

5.4 Digestive system
- Salivary secretion
- Gastric acid secretion
- Pancreatic secretion
- Bile secretion
- Carbohydrate digestion and absorption
- Protein digestion and absorption
- Fat digestion and absorption
- Vitamin digestion and absorption
- Mineral absorption

5.5 Excretory system
- Vasopressin-regulated water reabsorption
- Aldosterone-regulated sodium reabsorption
- Endocrine and other factor-regulated calcium reabsorption
- Proximal tubule bicarbonate reclamation
- Collecting duct acid secretion

5.6 Nervous system
- Glutamatergic synapse
- GABAergic synapse
- Cholinergic synapse
- Dopaminergic synapse
- Serotonergic synapse
- Long-term potentiation
- Long-term depression
- Retrograde endocannabinoid signaling
- Synaptic vesicle cycle
- Neurotrophin signaling pathway

5.7 Sensory system
- Phototransduction
- Phototransduction - fly
- Olfactory transduction
- Taste transduction *Major update!*
- Inflammatory mediator regulation of TRP channels

5.8 Development
- Dorso-ventral axis formation
- Axon guidance
- Osteoclast differentiation

5.9 Aging
- Longevity regulating pathway - mammal *New!*
- Longevity regulating pathway - worm *New!*
- Longevity regulating pathway - multiple species *New!*

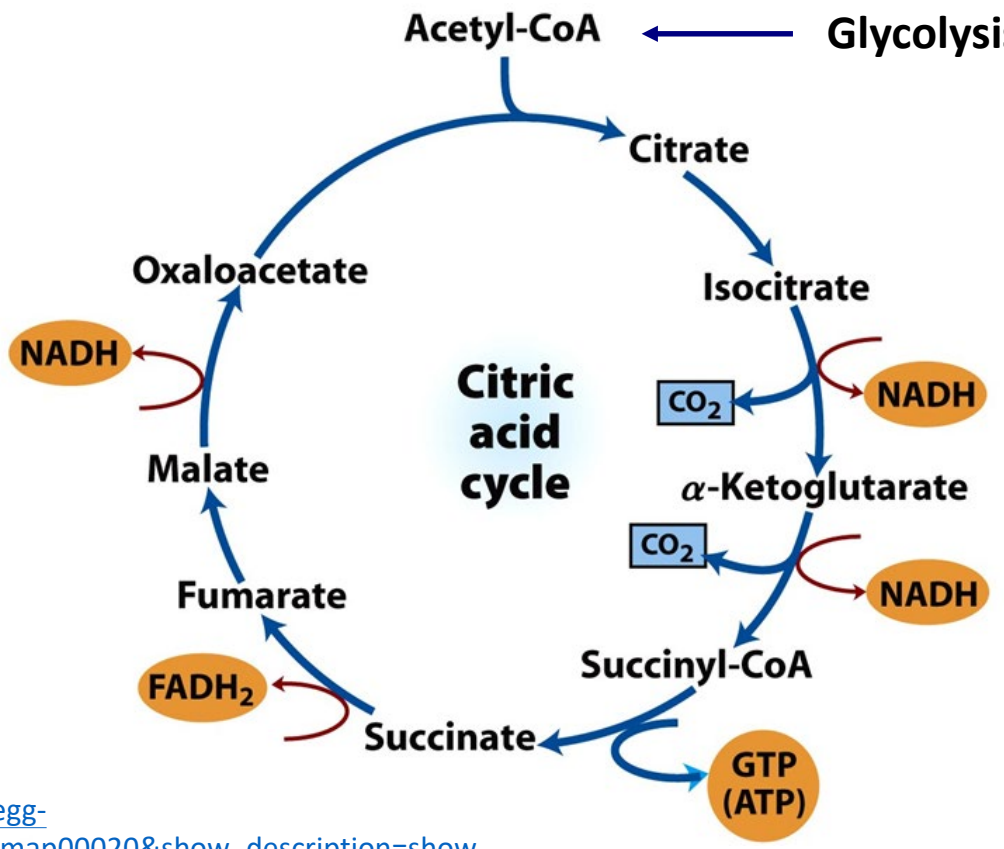
5.10 Environmental adaptation
- Circadian rhythm
- Circadian entrainment
- Circadian rhythm - fly
- Circadian rhythm - plant
- Plant-pathogen interaction

Citrate cycle (TCA cycle) - Reference pathway (KO)

Pathway description

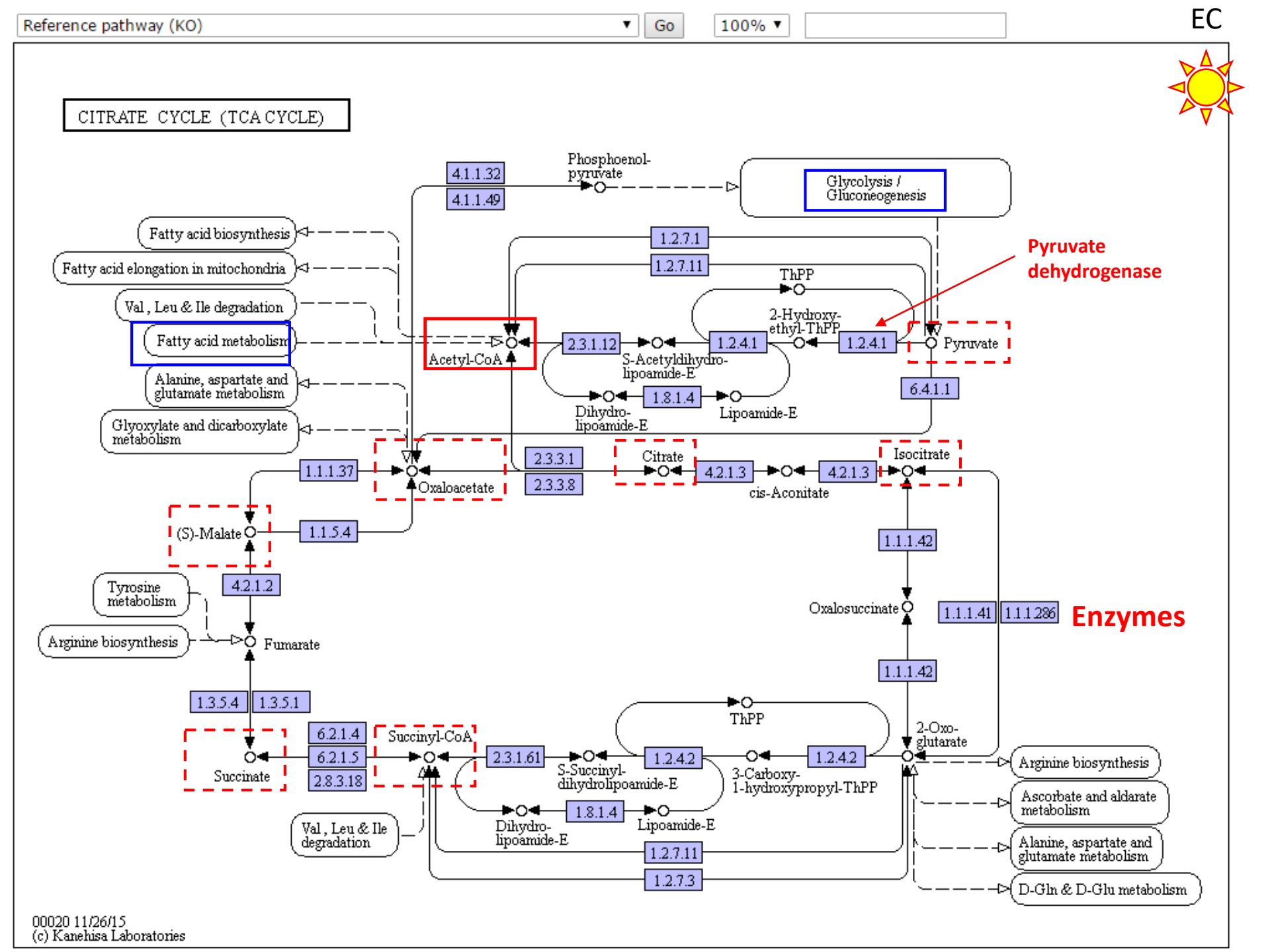
[Pathway menu | Organism menu | Pathway entry | Download KGML | Hide description | User data mapping]

The citrate cycle (TCA cycle, Krebs cycle) is an important aerobic pathway for the final steps of the oxidation of carbohydrates and fatty acids. The cycle starts with acetyl-CoA, the activated form of acetate, derived from glycolysis and pyruvate oxidation for carbohydrates and from beta oxidation of fatty acids. The two-carbon acetyl group in acetyl-CoA is transferred to the four-carbon compound of oxaloacetate to form the six-carbon compound of citrate. In a series of reactions two carbons in citrate are oxidized to CO2 and the reaction pathway supplies NADH for use in the oxidative phosphorylation and other metabolic processes. The pathway also supplies important precursor metabolites including 2-oxoglutarate. At the end of the cycle the remaining four-carbon part is transformed back to oxaloacetate. According to the genome sequence data, many organisms seem to lack genes for the full cycle [MD:M00009], but contain genes for specific segments [MD:M00010 M00011].



http://www.genome.jp/kegg-bin/show_pathway?map=map00020&show_description=show

From Lehninger Biochem


Reference pathway (KO)
Go
100%
EC
CITRATE CYCLE (TCA CYCLE)
Pyruvate dehydrogenase
Enzymes
Phosphoenol-pyruvate
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val , Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
Acetyl-CoA
ThPP
2-Hydroxy-ethyl-ThPP
S-Acetyldihydro-lipoamide-E
Dihydro-lipoamide-E
Lipoamide-E
Pyruvate
Oxaloacetate
Citrate
cis-Aconitate
Isocitrate
(S)-Malate
Tyrosine metabolism
Arginine biosynthesis
Fumarate
Oxalosuccinate
2-Oxo-glutarate
Succinyl-CoA
Succinate
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Arginine biosynthesis
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
D-Gln & D-Glu metabolism
Val , Leu & Ile degradation
4.1.1.32
4.1.1.49
1.2.7.1
1.2.7.11
2.3.1.12
1.2.4.1
1.2.4.1
6.4.1.1
1.8.1.4
2.3.3.1
2.3.3.8
4.2.1.3
4.2.1.3
1.1.1.37
1.1.5.4
4.2.1.2
1.1.1.42
1.1.1.41
1.1.1.286
1.1.1.42
1.3.5.4
1.3.5.1
6.2.1.4
6.2.1.5
2.8.3.18
2.3.1.61
1.2.4.2
1.2.4.2
1.8.1.4
1.2.7.11
1.2.7.3
00020 11/26/15
(c) Kanehisa Laboratories

EC Number

Group	Reaction catalyzed	Typical reaction	Enzyme example(s) with trivial name
EC 1 ***Oxidoreductases***	To catalyze oxidation/reduction reactions; transfer of H and O atoms or electrons from one substance to another	AH + B → A + BH (**reduced**) A + O → AO (**oxidized**)	Dehydrogenase, oxidase
EC 2 ***Transferases***	Transfer of a functional group from one substance to another. The group may be methyl-, acyl-, amino- or phosphate group	AB + C → A + BC	Transaminase, kinase
EC 3 ***Hydrolases***	Formation of two products from a substrate by hydrolysis	AB + H_2O → AOH + BH	Lipase, amylase, peptidase
EC 4 ***Lyases***	Non-hydrolytic addition or removal of groups from substrates. C-C, C-N, C-O or C-S bonds may be cleaved	RCOCOOH → RCOH + CO_2 or [X-A-B-Y] → [A=B + X-Y]	Decarboxylase
EC 5 ***Isomerases***	Intramolecule rearrangement, i.e. isomerization changes within a single molecule	ABC → BCA	Isomerase, mutase
EC 6 ***Ligases***	Join together two molecules by synthesis of new C-O, C-S, C-N or C-C bonds with simultaneous breakdown of ATP	X + Y+ ATP → XY + ADP + Pi	Synthetase

- *EC 3* enzymes are hydrolases (enzymes that use water to break up some other molecule)
- *EC 3.4* are hydrolases that act on peptide bonds
- *EC 3.4.11* are those hydrolases that cleave off the amino-terminal amino acid from a polypeptide
- *EC 3.4.11.4* are those that cleave off the amino-terminal end from a tripeptide

https://en.wikipedia.org/wiki/Enzyme_Commission_number

Pyruvate dehydrogenase

KEGG ORTHOLOGY: K00163

Help

Entry	K00163 KO
Name	aceE
Definition	pyruvate dehydrogenase E1 component [EC:1.2.4.1]
Pathway	ko00010 Glycolysis / Gluconeogenesis ko00020 Citrate cycle (TCA cycle) ko00620 Pyruvate metabolism ko01200 Carbon metabolism
Module	M00307 Pyruvate oxidation, pyruvate => acetyl-CoA
Brite	KEGG Orthology (KO) [BR:ko00001] Metabolism Overview 01200 Carbon metabolism K00163 aceE; pyruvate dehydrogenase E1 component Carbohydrate metabolism 00010 Glycolysis / Gluconeogenesis K00163 aceE; pyruvate dehydrogenase E1 component 00020 Citrate cycle (TCA cycle) K00163 aceE; pyruvate dehydrogenase E1 component 00620 Pyruvate metabolism K00163 aceE; pyruvate dehydrogenase E1 component KEGG modules [BR:ko00002] Pathway module Carbohydrate and lipid metabolism Central carbohydrate metabolism M00307 Pyruvate oxidation, pyruvate => acetyl-CoA K00163 aceE; pyruvate dehydrogenase E1 component Enzymes [BR:ko01000] 1. Oxidoreductases 1.2 Acting on the aldehyde or oxo group of donors 1.2.4 With a disulfide as acceptor 1.2.4.1 pyruvate dehydrogenase (acetyl-transferring) K00163 aceE: vate dehydrogenase E1 component BRITE hierarchy
Other DBs	RN: R00014 R01699 R03270 COG: COG2609 GO: 0004739
Genes	ECO: b0114(aceE) ECJ: JW0110(aceE) ECD: ECDH10B_0094(aceE)

All links

Ontology (5)
- KEGG BRITE (3)
- GO (1)
- COG (1)

Pathway (17)
- KEGG PATHWAY (16)
- KEGG MODULE (1)

Chemical reaction (8)
- KEGG ENZYME (1)
- KEGG REACTION (3)
- KEGG RCLASS (4)

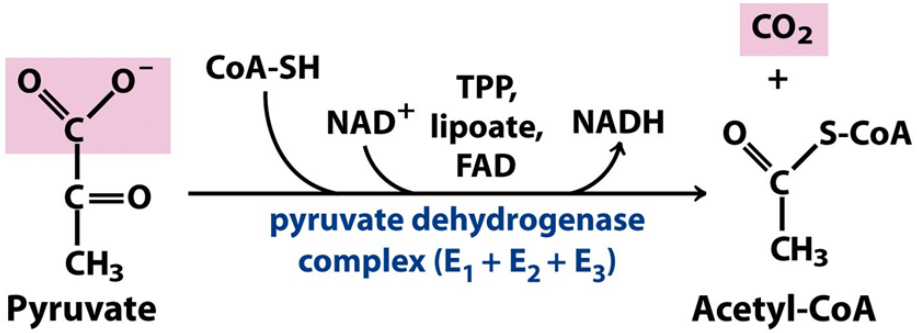
Gene (81681)
- KEGG GENES (1923)
- KEGG MGENES (65791)
- RefGene (13945)
- EGENES (1)
- OC (21)

Protein sequence (1329)
- UniProt (1317)
- SWISS-PROT (12)

Literature (1)
- PubMed (1)

All databases (83041)

Download RDF



BRITE hierarchies of pyruvate dehydrogenase

Search BRITE hierarchies

```
KEGG Orthology (KO) [BR:ko00001]
  Metabolism
    Overview
      01200 Carbon metabolism [PATH:ko01200]
        K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]
    Carbohydrate metabolism
      00010 Glycolysis / Gluconeogenesis [PATH:ko00010]
        K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]
      00020 Citrate cycle (TCA cycle) [PATH:ko00020]
        K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]
      00620 Pyruvate metabolism [PATH:ko00620]
        K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]

KEGG modules [BR:ko00002]
  Pathway module
    Carbohydrate and lipid metabolism
      Central carbohydrate metabolism
        M00307  Pyruvate oxidation, pyruvate => acetyl-CoA [PATH:map01200 map00010 map00020 map00620]
          K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]

Enzymes [BR:ko01000]
  1. Oxidoreductases
    1.2  Acting on the aldehyde or oxo group of donors
      1.2.4  With a disulfide as acceptor
    → 1.2.4.1  pyruvate dehydrogenase (acetyl-transferring)
        K00163  aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1]
```

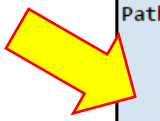
Enzyme ID: 1.2.4.1

KEGG

ENZYME: 1.2.4.1

Help

Entry	EC 1.2.4.1 Enzyme
Name	pyruvate dehydrogenase (acetyl-transferring); MtPDC (mitochondrial pyruvate dehydogenase complex); pyruvate decarboxylase; pyruvate dehydrogenase; pyruvate dehydrogenase (lipoamide); pyruvate dehydrogenase complex; pyruvate:lipoamide 2-oxidoreductase (decarboxylating and acceptor-acetylating); pyruvic acid dehydrogenase; pyruvic dehydrogenase
Class	Oxidoreductases; Acting on the aldehyde or oxo group of donors; With a disulfide as acceptor BRITE hierarchy
Sysname	pyruvate:[dihydrolipoyllysine-residue acetyltransferase]-lipoyllysine 2-oxidoreductase (decarboxylating, acceptor-acetylating)
Reaction(IUBMB)	pyruvate + [dihydrolipoyllysine-residue acetyltransferase] lipoyllysine = [dihydrolipoyllysine-residue acetyltransferase] S-acetyldihydrolipoyllysine + CO2 [RN:R01699]
Reaction(KEGG)	R01699; (other) R00014 R00209 R03270 Reaction
Substrate	pyruvate [CPD:C00022]; [dihydrolipoyllysine-residue acetyltransferase] lipoyllysine
Product	[dihydrolipoyllysine-residue acetyltransferase] S-acetyldihydrolipoyllysine [CPD:C16255]; CO2 [CPD:C00011]
Comment	Contains thiamine diphosphate. It is a component (in multiple copies) of the multienzyme pyruvate dehydrogenase complex in which it is bound to a core of molecules of EC 2.3.1.12, dihydrolipoyllysine-residue acetyltransferase, which also binds multiple copies of EC 1.8.1.4, dihydrolipoyl dehydrogenase. It does not act on free lipoamide or lipoyllysine, but only on the lipoyllysine residue in EC 2.3.1.12.
History	EC 1.2.4.1 created 1961, modified 2003
Pathway	ec00010 Glycolysis / Gluconeogenesis ec00020 Citrate cycle (TCA cycle) ec00620 Pyruvate metabolism ec01100 Metabolic pathways ec01110 Biosynthesis of secondary metabolites ec01120 Microbial metabolism in diverse environments ec01130 Biosynthesis of antibiotics



All links

Pathway (15)
- KEGG PATHWAY (14)
- KEGG MODULE (1)

Chemical substance (11)
- KEGG COMPOUND (11)

Chemical reaction (19)
- KEGG REACTION (4)
- KEGG RPAIR (9)
- KEGG RCLASS (6)

Gene (193582)
- KEGG ORTHOLOGY (4)
- KEGG GENES (8877)
- KEGG DGENES (109)
- KEGG MGENES (149161)
- RefGene (34240)
- EGENES (1191)

Protein sequence (14378)
- UniProt (11654)
- SWISS-PROT (117)
- RefSeq(pep) (2444)
- PDBSTR (144)
- PMD (19)

DNA sequence (10713)
- RefSeq(nuc) (6415)
- GenBank (2234)
- EMBL (2064)

3D Structure (24)
- PDB (24)

Protein domain (4)
- InterPro (4)

Literature (3)
- PubMed (3)

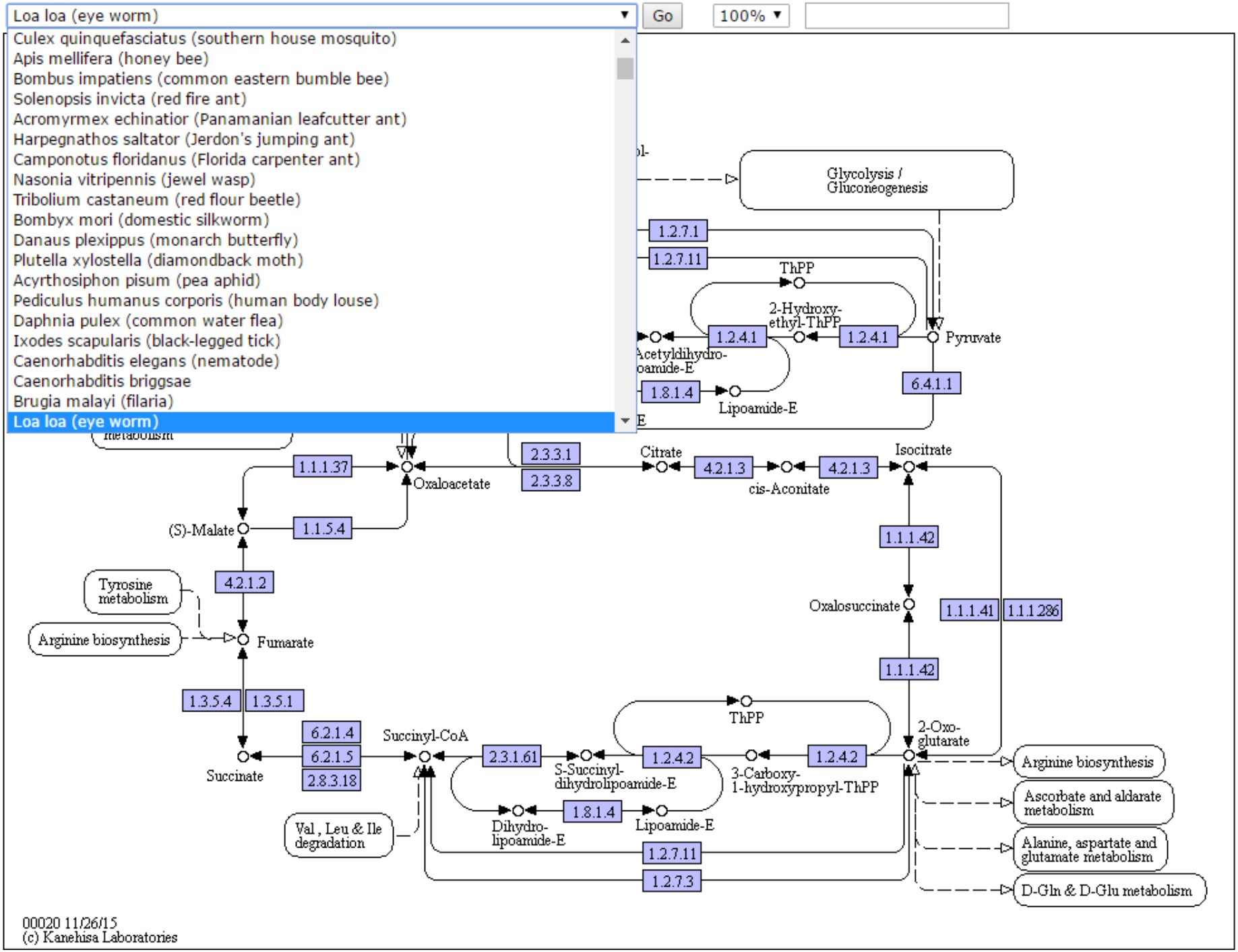
All databases (218749)

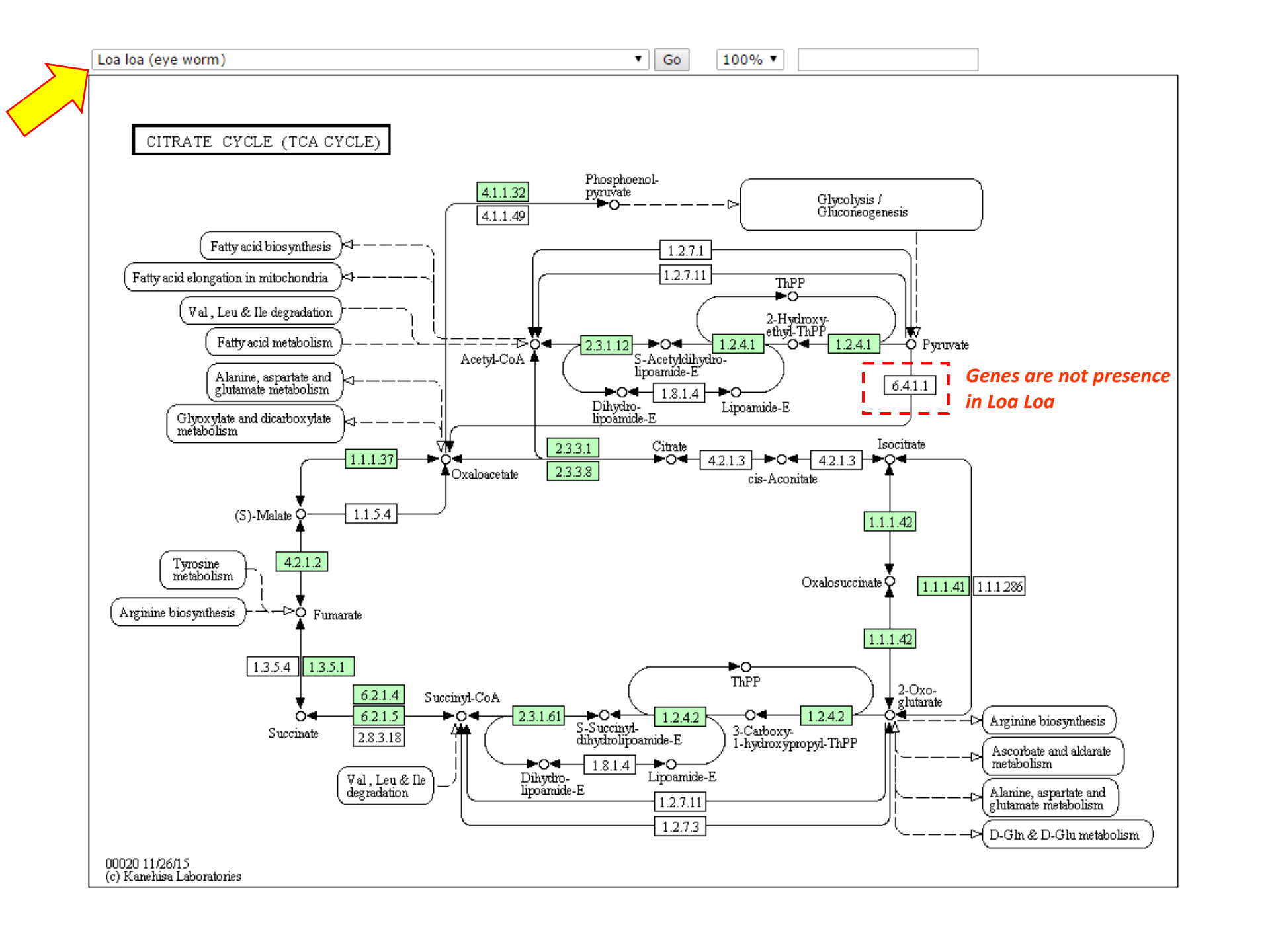
Download RDF

Exercise:

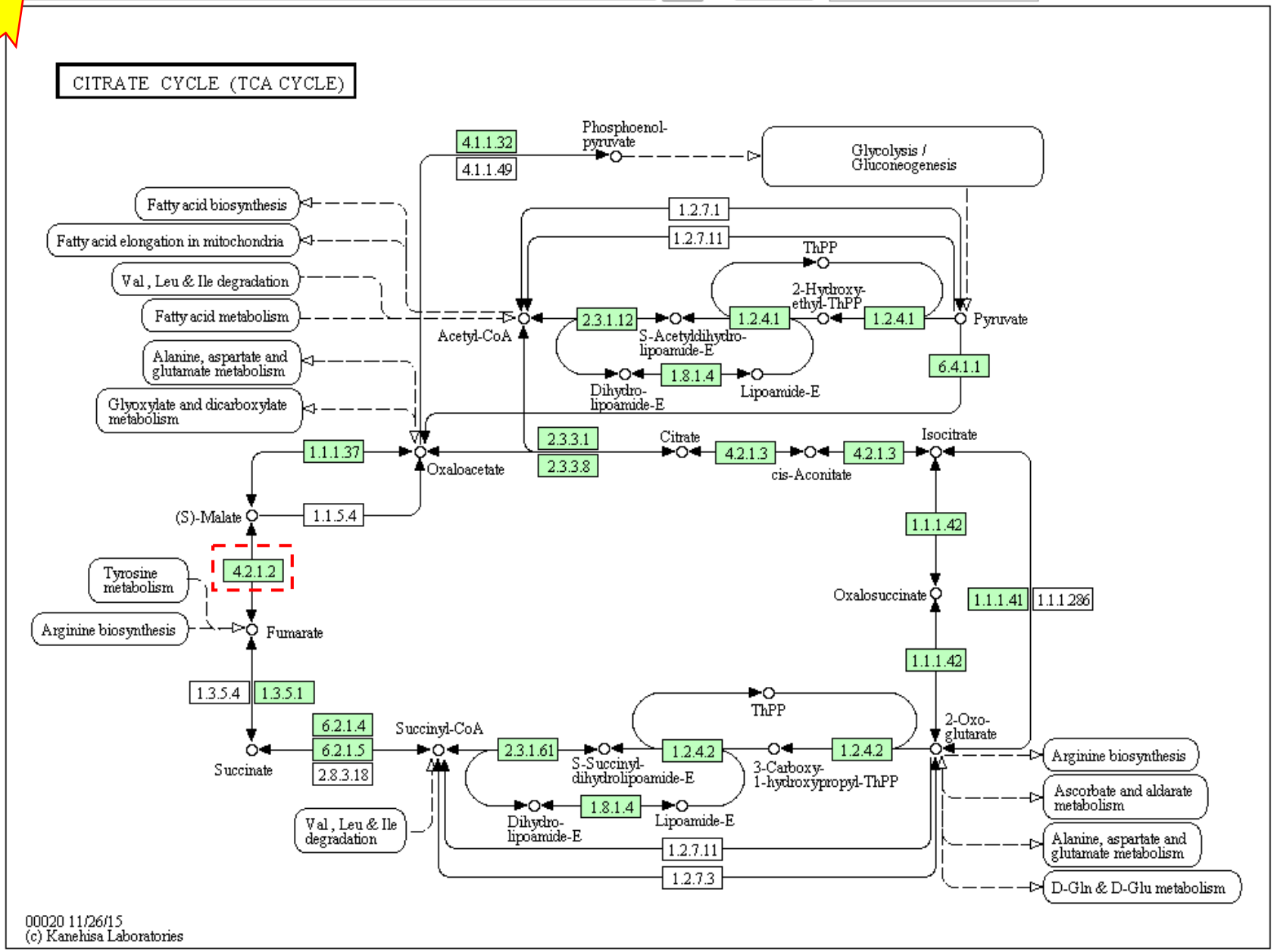
What is enzyme: EC 6.4.1.2?

How many pathways are EC 6.4.1.2 involved in?


Loa loa (eye worm)
Go
100%
Culex quinquefasciatus (southern house mosquito)
Apis mellifera (honey bee)
Bombus impatiens (common eastern bumble bee)
Solenopsis invicta (red fire ant)
Acromyrmex echinatior (Panamanian leafcutter ant)
Harpegnathos saltator (Jerdon's jumping ant)
Camponotus floridanus (Florida carpenter ant)
Nasonia vitripennis (jewel wasp)
Tribolium castaneum (red flour beetle)
Bombyx mori (domestic silkworm)
Danaus plexippus (monarch butterfly)
Plutella xylostella (diamondback moth)
Acyrthosiphon pisum (pea aphid)
Pediculus humanus corporis (human body louse)
Daphnia pulex (common water flea)
Ixodes scapularis (black-legged tick)
Caenorhabditis elegans (nematode)
Caenorhabditis briggsae
Brugia malayi (filaria)
Loa loa (eye worm)
Glycolysis / Gluconeogenesis
1.2.7.1
1.2.7.11
ThPP
2-Hydroxy-ethyl-ThPP
1.2.4.1
1.2.4.1
Pyruvate
Acetyldihydro-oamide-E
1.8.1.4
Lipoamide-E
6.4.1.1
2.3.3.1
2.3.3.8
Citrate
4.2.1.3
4.2.1.3
Isocitrate
cis-Aconitate
1.1.1.37
Oxaloacetate
(S)-Malate
1.1.5.4
1.1.1.42
Tyrosine metabolism
4.2.1.2
Oxalosuccinate
1.1.1.41
1.1.1.286
Arginine biosynthesis
Fumarate
1.1.1.42
1.3.5.4
1.3.5.1
ThPP
2-Oxo-glutarate
6.2.1.4
6.2.1.5
2.8.3.18
Succinyl-CoA
2.3.1.61
1.2.4.2
1.2.4.2
Succinate
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Arginine biosynthesis
Ascorbate and aldarate metabolism
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Val , Leu & Ile degradation
1.2.7.11
Alanine, aspartate and glutamate metabolism
1.2.7.3
D-Gln & D-Glu metabolism
00020 11/26/15
(c) Kanehisa Laboratories


Loa loa (eye worm)
Go
100%
CITRATE CYCLE (TCA CYCLE)
4.1.1.32
4.1.1.49
Phosphoenol-pyruvate
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val , Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
1.2.7.11
ThPP
2-Hydroxy-ethyl-ThPP
2.3.1.12
1.2.4.1
1.2.4.1
Pyruvate
Acetyl-CoA
S-Acetyldihydro-lipoamide-E
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
6.4.1.1
Genes are not presence in Loa Loa
1.1.1.37
2.3.3.1
2.3.3.8
Citrate
4.2.1.3
4.2.1.3
Isocitrate
Oxaloacetate
cis-Aconitate
(S)-Malate
1.1.5.4
1.1.1.42
Tyrosine metabolism
4.2.1.2
Oxalosuccinate
1.1.1.41
1.1.1.286
Arginine biosynthesis
Fumarate
1.1.1.42
1.3.5.4
1.3.5.1
ThPP
6.2.1.4
6.2.1.5
2.8.3.18
Succinyl-CoA
2.3.1.61
1.2.4.2
1.2.4.2
2-Oxo-glutarate
Succinate
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Arginine biosynthesis
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
D-Gln & D-Glu metabolism
Val , Leu & Ile degradation
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
1.2.7.11
1.2.7.3
00020 11/26/15
(c) Kanehisa Laboratories

Homo sapiens (human) Go 100%



KEGG

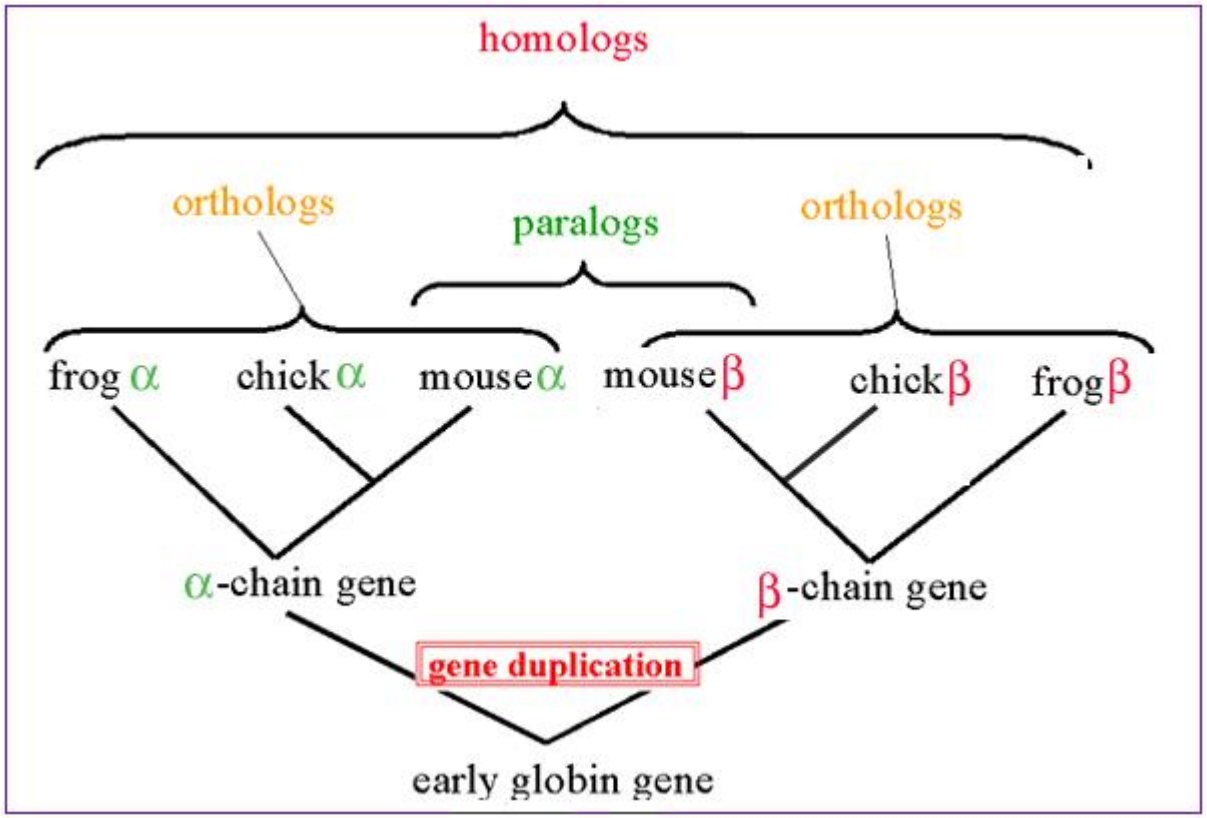
Homo sapiens (human): 2271

Help

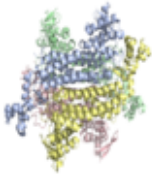
Entry	2271 CDS T01001
Gene name	FH, FMRD, HLRCC, LRCC, MCL, MCUL1
Definition	(RefSeq) fumarate hydratase (EC:4.2.1.2)
KO	K01679 fumarate hydratase, class II [EC:4.2.1.2]
Organism	hsa Homo sapiens (human)
Pathway	hsa00020 Citrate cycle (TCA cycle) hsa00620 Pyruvate metabolism hsa01100 Metabolic pathways hsa01200 Carbon metabolism hsa05200 Pathways in cancer hsa05211 Renal cell carcinoma
Module	hsa_M00009 Citrate cycle (TCA cycle, Krebs cycle) hsa_M00011 Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate
Disease	H00021 Renal cell carcinoma H00804 Multiple cutaneous and uterine leiomyomata H01022 Diseases of the tricarboxylic acid cycle
Brite	KEGG Orthology (KO) [BR:hsa00001] Metabolism Overview 01200 Carbon metabolism 2271 (FH) Carbohydrate metabolism 00020 Citrate cycle (TCA cycle) 2271 (FH) 00620 Pyruvate metabolism 2271 (FH) Human Diseases Cancers 05200 Pathways in cancer 2271 (FH) 05211 Renal cell carcinoma 2271 (FH) Enzymes [BR:hsa01000] 4. Lyases 4.2 Carbon-oxygen lyases 4.2.1 Hydro-lyases 4.2.1.2 fumarate hydratase 2271 (FH) BRITE hierarchy
SSDB	Ortholog Paralog GFIT

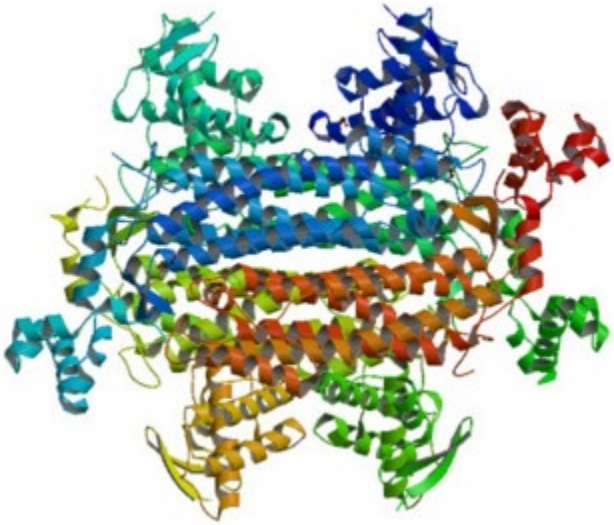
Orthologs/paralogs,
conserved gene clusters,
and sequence motifs

GFIT:
Gene Function Identification Tool


homologs
orthologs
paralogs
orthologs
frog α
chick α
mouse α
mouse β
chick β
frog β
α-chain gene
β-chain gene
gene duplication
early globin gene

hsa:5621 (253 a.a.)							K05634	PRNP	[tr:Q53YK7_HUMAN sp:PRIO_HUMAN] PRNP; prion pr
DBHit(Length)	Over-lap	Ident. (%)	Score	margin	Para-log	Cont.	KO	Orth	Current Annotation
> ptr:458076(253)	253	99.2	1844	(1627)	501		K05634	PRNP	[sp:PRIO_PANTR tr:A5A6I7_PANTR] PRNP; prion prot
> mcc:717859(253)	253	96.4	1809	(1605)	425		K05634	PRNP	[sp:PRIO_MACMU] PRNP; prion protein ; K05634 prio
> bta:281427(264)	261	88.9	1706	(1505)	384		K05634	PRNP	[sp:PRIO_BOVIN] PRNP; prion protein (p27-30) (Creu Gerstmann-Strausler-Scheinker syndrome, fatal fami protein
> ssc:494014(257)	254	89.0	1690	(1481)	135		K05634	PRNP	[sp:PRIO_PIG tr:Q5PSB2_PIG] PRNP; prion protein ;
> rno:24686(254)	254	89.4	1676	(1470)	393		K05634	PRNP	[sp:PRIO_RAT] Prnp; prion protein ; K05634 prion pr
> ecb:100065904(255)	253	88.9	1675	(1469)	300		K05634	PRNP	[tr:B5B2H1_HORSE] PRNP; prion protein ; K05634 pr
> mmu:19122(254)	255	89.0	1675	(1460)	424		K05634	PRNP	[sp:PRIO_MOUSE tr:Q4FJQ7_MOUSE] Prnp; prion pro
> cfa:485783(257)	254	87.0	1662	(1444)	463		K05634	PRNP	[tr:O46593_CANFA] PRNP; prion protein (p27-30) (C Gerstmann-Strausler-Scheinker syndrome, fatal fami protein
> mdo:554189(269)	261	71.6	1295	(1072)	557		K05634	PRNP	PRNP; prion protein ; K05634 prion protein
> oaa:100086593(183)	186	67.7	858	(656)	347		K05634	PRNP	similar to prion protein ; K05634 prion protein
> tgu:100231829(269)	254	40.9	534	(340)	213		K05634	PRNP	prion protein ; K05634 prion protein
> gga:396452(267)	268	39.2	531	(324)	345		K05634	PRNP	[sp:PRIO_CHICK] PRNP; prion protein (p27-30) (Creu Gerstmann-Strausler-Scheinker syndrome, fatal fami protein
> xla:444620(235)	255	33.7	471	(29)	120		K05634	PRNP	[tr:Q6GLY6_XENLA] MGC84114 protein ; K05634 prio
> xtr:100158502(224)	257	33.5	461	(235)	88		K05634	PRNP	[tr:B2GU34_XENTR tr:A2BDH3_XENTR] hypothetical p prion protein
> cbr:CBG05352(337)	125	42.4	291	(40)	322		K12741	HNRNPA1_3	Hypothetical protein CBG05352 ; K12741 heterogen A1/3
> scl:sce8934(440)	122	41.0	273	(35)	314				[tr:A9G922_SORC5] hypothetical protein
> cel:Y43F8C.20(193)	146	36.3	261	(3)	360				[tr:Q7YWQ8_CAEEL] grsp-1; Glycine Rich Secreted P
> bja:bll5579(139)	116	41.4	254	(58)	27				[tr:Q89IQ5_BRAJA] hypothetical protein
> apt:APA01_26700(231)	167	36.5	253	(39)	9				hypothetical protein
> dya:Dyak_GE13461(147)	139	38.1	253	(17)	399				[tr:B4P0S7_DROYA] GE13461 gene product from tran

Motif	Pfam: Lyase_1 FumaraseC_C YadA_anchor Motif
Other DBs	NCBI-ProteinID: NP_000134 NCBI-GeneID: 2271 OMIM: 136850 HGNC: 3700 HPRD: 00652 Ensembl: ENSG00000091483 Vega: OTTHUMG00000039597 UniProt: P07954
Structure	PDB: 3E04 Thumbnail  Jmol
Position	1q42.1
AA seq	510 aa AA seq DB search MYRALRLLARSRPLVRAPAAALASAPGLGGAAVPSFWPPNAARMASQNSFRIEYDTFGEL KVPNDKYYGAQTVRSTMNFKIGGVTERMPTPVIKAFGILKRAAAEVNQDYGLDPKIANAI MKAADEVAEGKLNDHFPLVVWQTGSGTQTNMNVNEVISNRAIEMLGGELGSKIPVHPNDH VNKSQSSNDTFPTAMHIAAAIEVHEVLLPGLQKLHDALDAKSKEFAQIIKIGRTHTQDAV PLTLGQEFSGYVQQVKYAMTRIKAAMPRIYELAAGGTAVGTGLNTRIGFAEKVAAKVAAL TGLPFVTAPNKFEALAAHDALVELSGAMNTTACSLMKIANDIRFLGSGPRSGLGELILPE NEPGSSIMPGKVNPTQCEAMTMVAAQVMGNHVAVTVGGSNGHFELNVFKPMMIKNVLHSA RLLGDASVSFTENCVVGIQANTERINKLMNESLMLVTALNPHIGYDKAAKIAKTAHKNGS TLKETAIELGYLTAEQFDEWVKPKDMLGPK
NT seq	1533 nt NT seq atgtaccgagcacttcggctcctcgcgcgctcgcgtcccctcgtgcgggctccagccgca gccttagcttcggctcccggcttgggtggcgcggccgtgccctcgttttggcctccgaac gcggctcgaatggcaagccaaaattccttccggatagaatatgataccttggtgaacta aaggtgccaaatgataagtattatggcgcccagaccgtgagatctacgatgaactttaag attggaggtgtgacagaacgcatgccaaccccagttattaaagcttttggcatcttgaag cgagcggccgctgaagtaaaccaggattatggtcttgatccaaagattgctaatgcaata



Exercise:

What is enzyme: EC 6.4.1.2?

How many pathways are EC 6.4.1.2 involved in?

Does mouse have this enzyme?

Is there any 3D structure published?

How many orthologs & paralogs?

Show the protein sequence.

Drug development pathways

7. Drug Development

7.1 Chronology: Antiinfectives
- Penicillins
- Cephalosporins - parenteral agents
- Cephalosporins - oral agents
- Aminoglycosides
- Tetracyclines
- Macrolides and ketolides
- Quinolones
- Rifamycins
- Antifungal agents
- Antiviral agents
- Anti-HIV agents

7.2 Chronology: Antineoplastics
- Antineoplastics - alkylating agents
- Antineoplastics - antimetabolic agents
- Antineoplastics - agents from natural products
- Antineoplastics - hormones
- Antineoplastics - protein kinases inhibitors

7.3 Chronology: Nervous system agents
- Hypnotics
- Anxiolytics
- Anticonvulsants
- Local analgesics
- Opioid analgesics
- Antipsychotics
- Antipsychotics - phenothiazines
- Antipsychotics - butyrophenones
- Antidepressants
- Agents for Alzheimer-type dementia
- Antiparkinsonian agents

7.4 Chronology: Other drugs
- Sulfonamide derivatives - overview
- Sulfonamide derivatives - sulfa drugs
- Sulfonamide derivatives - diuretics
- Sulfonamide derivatives - hypoglycemic agents
- Antiarrhythmic drugs
- Antiulcer drugs
- Immunosuppressive agents
- Osteoporosis drugs
- Antimigraines
- Antithrombosis agents
- Antirheumatics - DMARDs and biological agents
- Antidiabetics
- Antidyslipidemic agents
- Antiglaucoma agents

7.5 Target-based classification: G protein-coupled receptors
- Cholinergic and anticholinergic drugs
- alpha-Adrenergic receptor agonists/antagonists
- beta-Adrenergic receptor agonists/antagonists
- Dopamine receptor agonists/antagonists
- Histamine H1 receptor antagonists
- Histamine H2/H3 receptor agonists/antagonists
- Serotonin receptor agonists/antagonists
- Eicosanoid receptor agonists/antagonists
- Opioid receptor agonists/antagonists
- Angiotensin receptor and endothelin receptor antagonists

7.6 Target-based classification: Nuclear receptors
- Glucocorticoid and mineralocorticoid receptor agonists/antagonists
- Progesterone, androgen and estrogen receptor agonists/antagonists
- Retinoic acid receptor (RAR) and retinoid X receptor (RXR) agonists/antagonists
- Peroxisome proliferator-activated receptor (PPAR) agonists

7.7 Target-based classification: Ion channels
- Nicotinic cholinergic receptor antagonists
- GABA-A receptor agonists/antagonists
- Calcium channel blocking drugs
- Sodium channel blocking drugs
- Potassium channel blocking and opening drugs
- N-Metyl-D-aspartic acid receptor antagonists

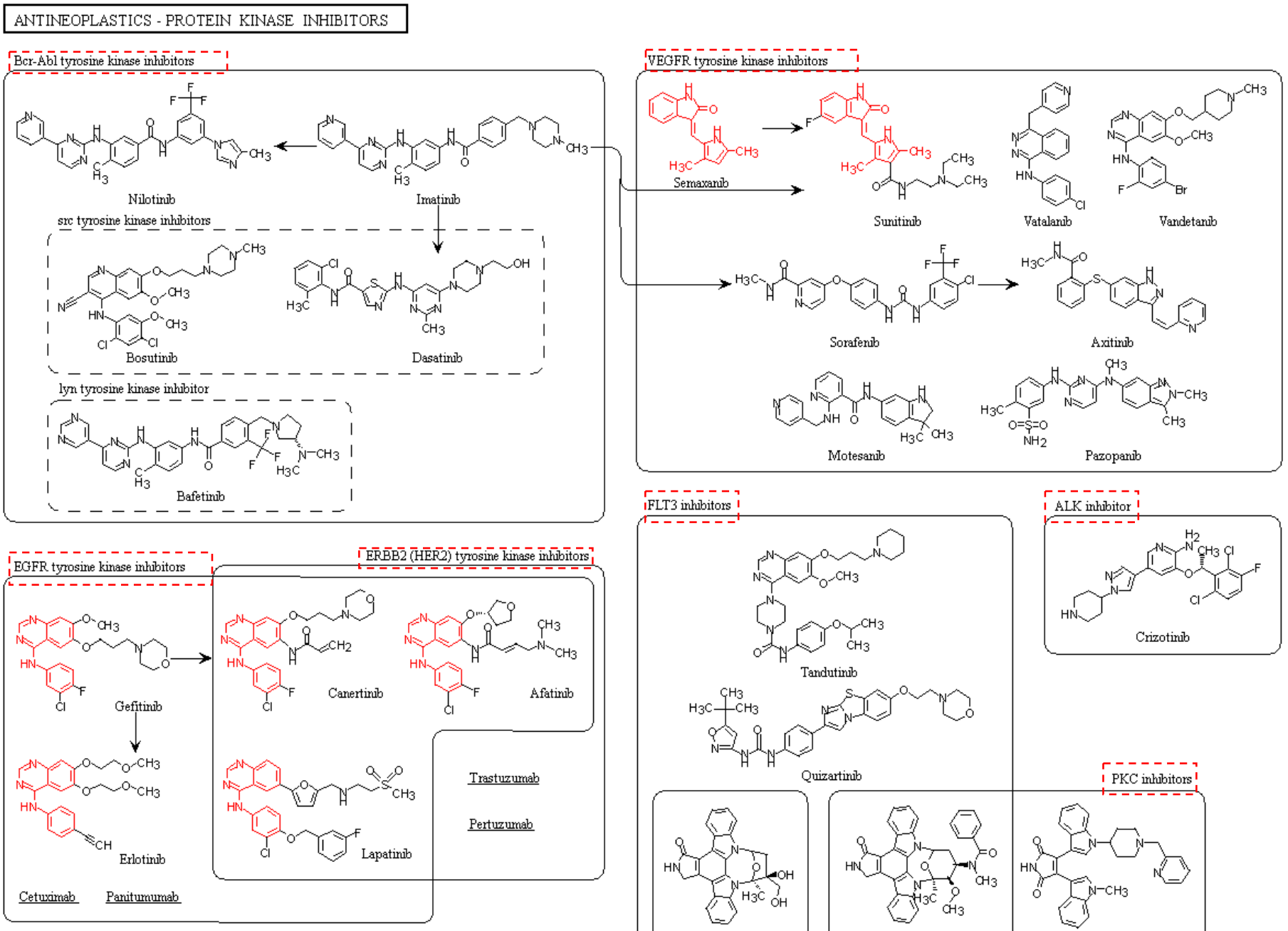
7.8 Target-based classification: Transporters
- Ion transporter inhibitors
- Neurotransmitter transporter inhibitors

7.9 Target-based classification: Enzymes
- Catecholamine transferase inhibitors
- Cyclooxygenase inhibitors
- HMG-CoA reductase inhibitors
- Renin-angiotensin system inhibitors
- HIV protease inhibitors

7.10 Structure-based classification
- Quinolines
- Eicosanoids
- Prostaglandins

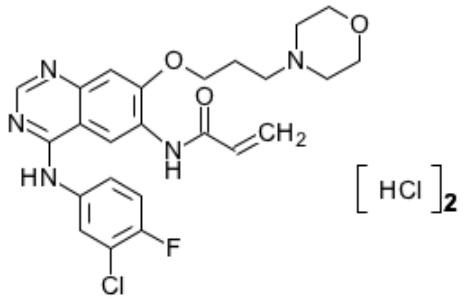
7.11 Skeleton-based classification
- Benzoic acid family
- 1,2-Diphenyl substitution family
- Naphthalene family
- Benzodiazepine family

Antineoplastics: protein kinase inhibitors



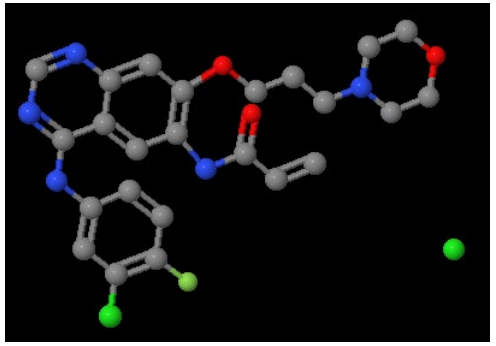
KEGG DRUG: D03350

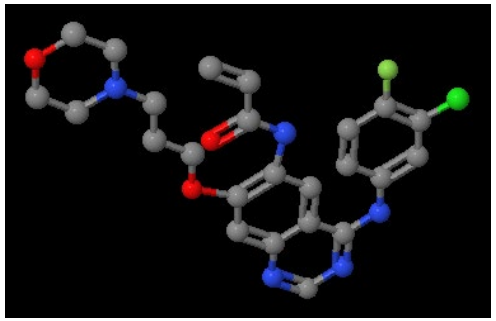
Help

Entry	D03350 Drug
Name	Canertinib dihydrochloride (USAN)
Formula	C24H25ClFN5O3. 2HCl
Mass	557.1164
Structure	 D03350 Mol file KCF file DB search Jmol KegDraw
Target	EGFR tyrosine kihase inhibitor [HSA:1956] [PATH:hsa04010] [PATH:hsa04012] [PATH:hsa04020] [PATH:hsa04060] [PATH:hsa04320] [PATH:hsa04510] [PATH:hsa04520] [PATH:hsa04540] [PATH:hsa04810] [PATH:hsa04912] [PATH:hsa05223]; ERBB2 tyrosine kinase inhibitor [HSA:2064] [PATH:hsa04012] [PATH:hsa04020] [PATH:hsa04320] [PATH:hsa04510] [PATH:hsa04520] [PATH:hsa05223]; ERBB3 tyrosine kinase inhibitor [HSA:2065] [PATH:hsa04012] [PATH:hsa04020]; ERBB4 tyrosine kinase inhibitor [HSA:2066] [PATH:hsa04012] [PATH:hsa04320]
Activity	Treatment of epithelial tumors

Remark	BRITE hierarchy
Pathway	PATH: map07045 Antineoplastics - protein kinases inhibitors
Other DBs	CAS: 289499-45-2 PubChem: 17397500 LigandBox: D03350
LinkDB	All DBs
KCF data	Show

Java script to see and rotate the 3D structure





Search pathway by gene name



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 | PATHWAY | BRITE | MODULE | DISEASE | DRUG | KO | GENOME | GENES | LIGAND | DBGET

Select prefix: map | Organism

Enter keywords: CDK2 **Search by Gene Name** | Go | Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps, change history, and last updates) representing our knowledge on the molecular interaction and reaction networks for:

0. Global Map
1. Metabolism
Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
2. Genetic Information Processing
3. Environmental Information Processing
4. Cellular Processes
5. Organismal Systems
6. Human Diseases

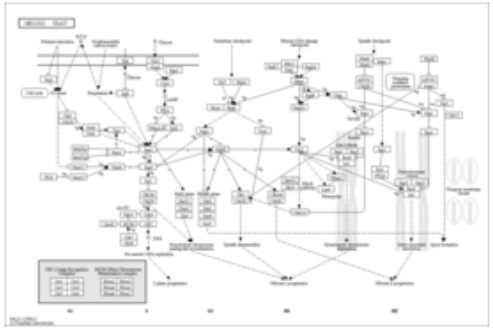
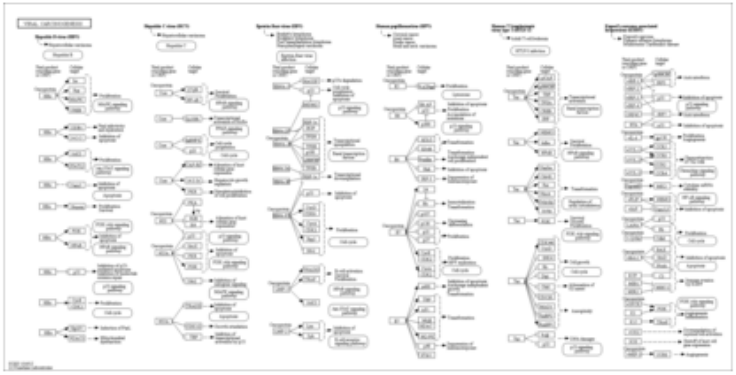
and also on the structure relationships (KEGG drug structure maps) in:

7. Drug Development

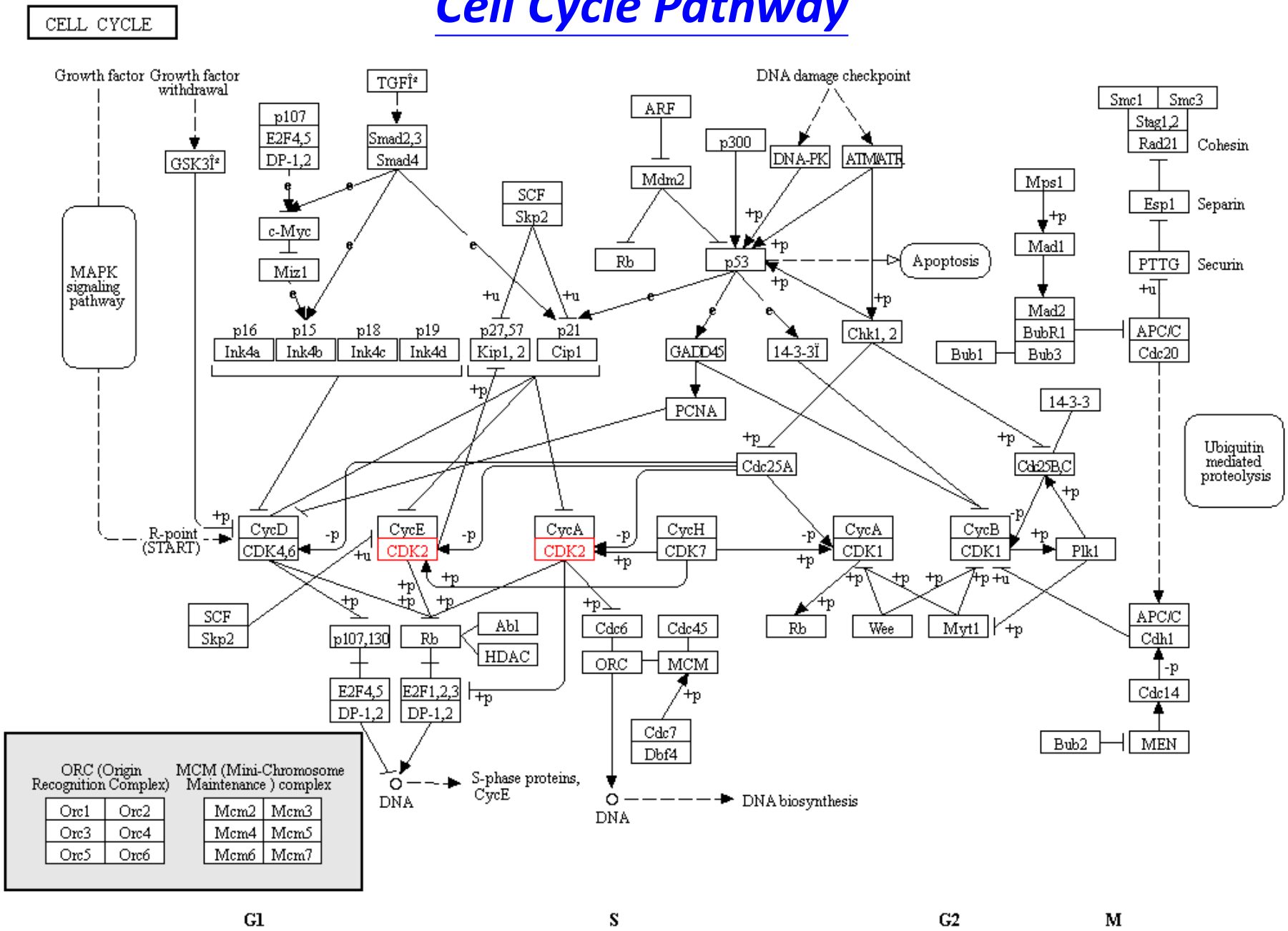
Pathway Mapping

CDK2 related pathways

Items : 1 - 15 of 15

Entry	Thumbnail Image	Name	Description	Object	Legend
Object ID map04113		Meiosis - yeast	...enes, a transcriptional activator, Ndt80, and a CDK2 homolog, Ime2, are required for the transcripti...	C00031 (D-Glucose) C00031 (D-Glucose) C00288 (HCO3-) C00575 (3',5'-Cyclic AMP) K02210 (MCM7) K02542 ...	Mcm7 Mcm6 Mcm5 Mcm4 Mcm3 Mcm2 Orc6 Orc5 Orc4 Orc3 Orc2 Orc1 MEIOSIS - YEAST MCM (Mini-Chromosome M...
map05203		Viral carcinogenesis	There is a strong association between viruses and the development of human malignancies. We now know...	...GR3) 944566 944566 K04451 (TP53) 944566 K02206 (CDK2) 1491938 1491938 1491938 1491938 K02089 (CDK4),...	... JAKS STAT3,5 PI3K NFκB Egr2/3 HBx HBx p53 HBx CDK2 Tax Tax Tax Tax CDK4/6 INK4 Dlg1 Polβ p53 MAD1...
map04110		Cell cycle	Mitotic cell cycle progression is accomplished through a reproducible sequence of events, DNA replic...	...5868 (CCNB) K06627 (CCNA) K02202 (CDK7) K02206 (CDK2) K02206 (CDK2) K02089 (CDK4), K02091 (CDK6) K06...	...,C Cdc45 Cdc6 Cdc25A GSK3β CDK1 CycB CycA CDK7 CDK2 CDK2 CDK4,6 CycH CycA CycE CycD MAPK signaling...

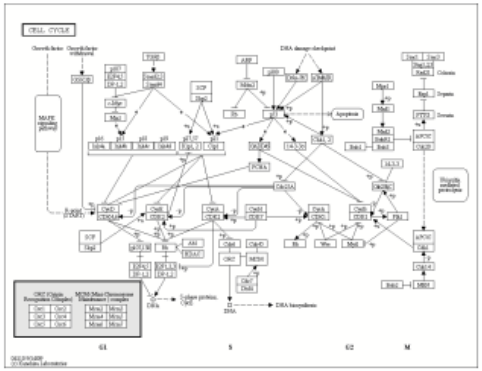
Cell Cycle Pathway



04110 4/5/16
(c) Kanehisa Laboratories

KEGG PATHWAY: map04110

Help

Entry	map04110 Pathway
Name	Cell cycle
Description	Mitotic cell cycle progression is accomplished through a reproducible sequence of events, DNA replication (S phase) and mitosis (M phase) separated temporally by gaps known as G1 and G2 phases. Cyclin-dependent kinases (CDKs) are key regulatory enzymes, each consisting of a catalytic CDK subunit and an activating cyclin subunit. CDKs regulate the cell's progression through the phases of the cell cycle by modulating the activity of key substrates. Downstream targets of CDKs include transcription factor E2F and its regulator Rb. Precise activation and inactivation of CDKs at specific points in the cell cycle are required for orderly cell division. Cyclin-CDK inhibitors (CKIs), such as p16Ink4a, p15Ink4b, p27Kip1, and p21Cip1, are involved in the negative regulation of CDK activities, thus providing a pathway through which the cell cycle is negatively regulated. Eukaryotic cells respond to DNA damage by activating signaling pathways that promote cell cycle arrest and DNA repair. In response to DNA damage, the checkpoint kinase ATM phosphorylates and activates Chk2, which in turn directly phosphorylates and activates p53 tumor suppressor protein. p53 and its transcriptional targets play an important role in both G1 and G2 checkpoints. ATR-Chk1-mediated protein degradation of Cdc25A protein phosphatase is also a mechanism conferring intra-S-phase checkpoint activation.
Class	Cellular Processes; Cell Growth and Death BRITE hierarchy
Pathw	map04110 Cell cycle  All organisms Ortholog table
Module	M00621 origin recognition complex M00622 MCM complex M00625 pre-replication complex M00643 ATR-ATRIP complex

Description

BRITE

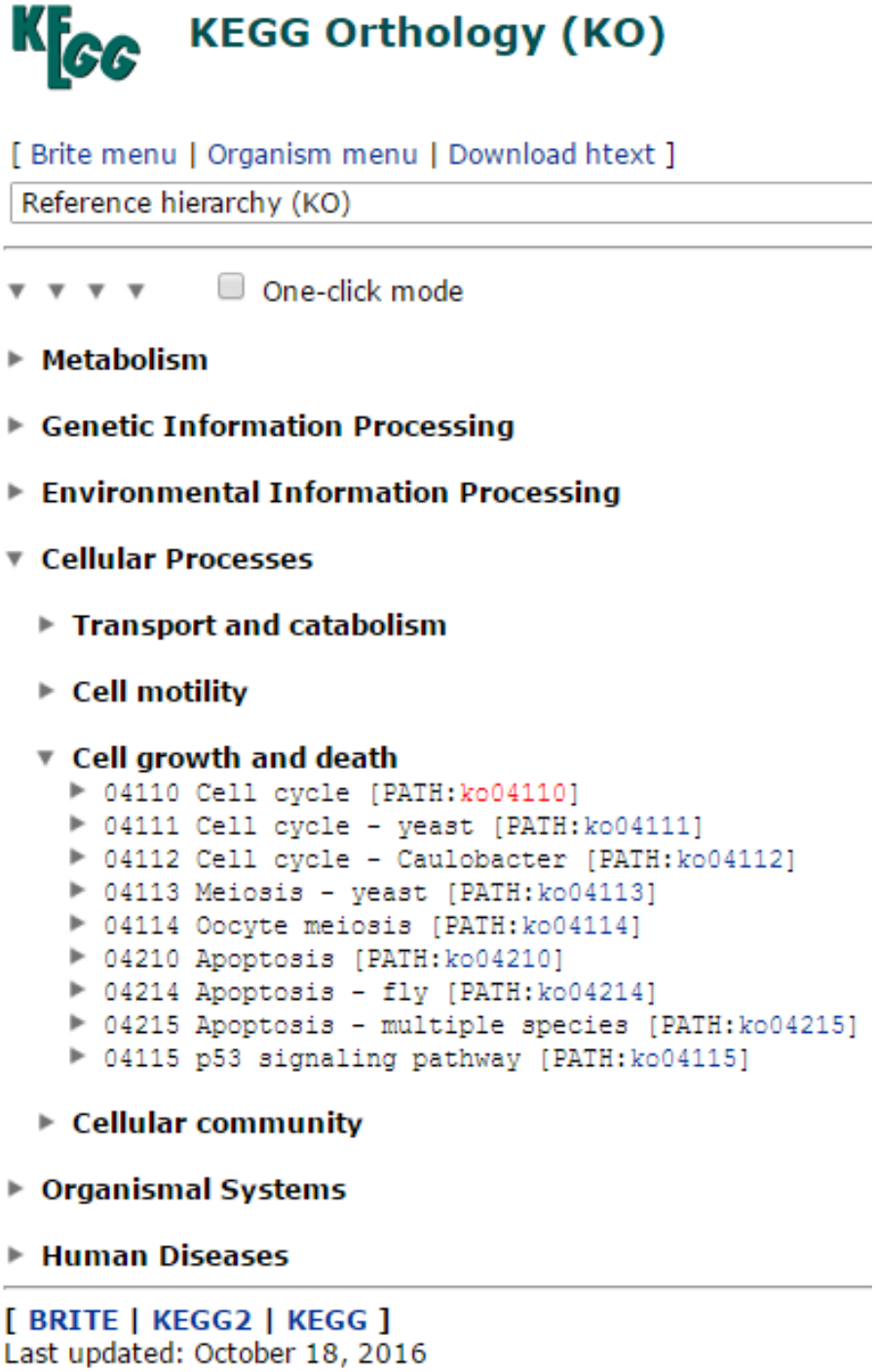
PATHWAY

Ortholog

Module

Pathway modules are subnetworks that are tighter functional units compared to pathway entries

Cell cycle in BRITE



Ortholog table of cell cycle

Ortholog table

PATHWAY: 04110

Eukaryotes ▼ Page: 1 ▼

Genes

Grp	Genus	Organism	K04503 (CCND1)[80]	K10151 (CCND2)[90]	K10152 (CCND3) [60]	K02089 (CDK4)[63]	K02091 (CDK6)[82]	K06618 (RB1)[75]
E.An	Homo	hsa	595	894	896	1019	1021	5925
E.Ani	Pan	ptr	747442	747123	462689	452025	746968	452715
E.Ani	Pan	pps	100995605	100990850	100972595	100976294	100981258	100973837
E.Ani	Gorilla	ggo	101130447 101144860	101139749	101130628	101147065	101132317	101133237 101134294 101134653
E.Ani	Pongo	pon	100171625	100457978	100452570	100443314	100454924	100462155
E.Ani	Nomascus	nle	100605736	100585770	100580170	100594967	100602542	100588629
E.Ani	Macaca	mcc	574320	710379	695215	715650	701822	704871
E.Ani	Macaca	mcf	102135329	102119572	101865999	101925733	102140874	102145859
E.Ani	Chlorocebus	csab	103246489	103218410	103221504	103238589	103226555	103214406
E.Ani	Rhinopithecus	rro	104677168	104658979	104667309	104682344	104682750	104662760

KEGG Mapper – Color Pathway



KEGG Mapper

A suite of KEGG mapping tools

- About KEGG Mapper

Search Pathway
Search&Color Pathway
Color Pathway
Color Pathway WebGL

Search Brite
Search&Color Brite
Join Brite
Join Brite Table

Search Module
Search&Color Module

Search Disease *New!*

Reconstruct Pathway
Reconstruct Brite
Reconstruct Module
Map Taxonomy

Annotate Sequence
BlastKOALA

KEGG Atlas
KEGG

About KEGG Mapper

KEGG Mapper is a collection of tools for KEGG mapping: KEGG pathway mapping, BRITE mapping, and MODULE mapping. Two pathway mapping tools, "Search Pathway" and "Color Pathway", were made available from the beginning of the KEGG project. The naming of these tools was somewhat misleading since both involved searching and coloring procedures. The latter is now called "Search&Color Pathway" and a new "Color Pathway" is introduced for coloring of a selected pathway map (without searching).

Pathway mapping tools

Search Pathway - basic pathway mapping tool

- Search against all pathway maps in one category (ko, org, etc.)
- Multiple search objects may be specified in one line
- Mapped objects are marked in red

Search&Color Pathway - pathway mapping tool with coloring options

- Search against all pathway maps in one category (ko, org, etc.)
- Each search object is specified in one line together with color attributes (background color and foreground color)
- Mapped objects can be marked in any color
- NCBI and UniProt identifiers may be used

Color Pathway - selected pathway map coloring tool

- One selected pathway map is colored in multiple ways
- Each object is specified in one line together with color attributes, i.e., multiple background and foreground colors in multiple columns
- Results shown as multiple snapshots of image maps
- When numerical values are given as attributes, they are displayed by gradation or by using WebGL graphics

http://www.kegg.jp/kegg/mapper.html

KEGG Color Codes

KEGG Color Codes

Database category (see: KEGG overview)

KEGG top	#f5e05f	#ffefaf
KEGG PATHWAY KEGG BRITE KEGG MODULE	#005050	#e0f0e0
KEGG ORTHOLOGY	#663300	#f2e9b6
KEGG GENES KEGG GENOME	#660033	#f0e0e4
KEGG LIGAND	#003366	#d4e4f0
KEGG MEDICUS	#440066	#e4ddf8
GenomeNet	#000066	#ccccff

Functional category (used in global pathway maps and genome maps)

Carbohydrate metabolism	#0000ee
Energy metabolism	#9933cc
Lipid metabolism	#009999
Nucleotide metabolism	#ff0000
Amino acid metabolism	#ff9933
Metabolism of other amino acids	#ff6600
Glycan biosynthesis and metabolism	#3399ff
Metabolism of cofactors and vitamins	#ff6699
Metabolism of terpenoids and polyketides	#00cc33
Biosynthesis of other secondary metabolites	#cc3366
Xenobiotics biodegradation and metabolism	#ccaa99
Enzyme families	#ccffff
Genetic information processing	#ffcccc
Environmental information processing	#ffff00
Cellular processes	#99cc66
Organismal systems	#99cc66
Human diseases	#99cc66

KEGG PATHWAY

Organism-specific pathway	#bfffbf	#66cc66
Reference pathway (KO, EC, Reaction)	#bfbfff	#6666cc
KEGG module	#ffbfbf	#cc6666
Reaction module	#00ccff	#0099ff
Disease gene	#ffcfff	#ff99ff
Drug target	#cfefff	#66cccc
Disease gene & drug target	#ffcfff	#6666cc

Pathway mapping of two organisms (global/overview maps)

Organism 1	#00cc33
Organism 2	#ff3366
Organisms 1 and 2	#3366ff

Pathway mapping of multiple organisms (regular maps)

Organism 1	#bfffbf
Organism 2	#ffbbcc
Organism 3	#bbccff
Organism 4	#cfffcf
Organism 5	#ffcfef
Organism 6	#cfefff
Organism 7	#dfefcf
Organism 8	#ffefcc
Organism 9	#dfccff
Organism 10	#dfdfcc

Ortholog table and SSDB gene cluster table

Cluster 1	#ffefcc
Cluster 2	#ffcfef
Cluster 3	#dfefcf
Cluster 4	#cfefff
Cluster 5	#dfcccc
Cluster 6	#dfccff
Cluster 7	#ffcccc
Cluster 8	#bbccff
Cluster 9	#dfdfcc
Cluster 10	yellow
Cluster 11	pink

Last updated: August 5, 2016

KEGG **Biosynthesis of amino acids**

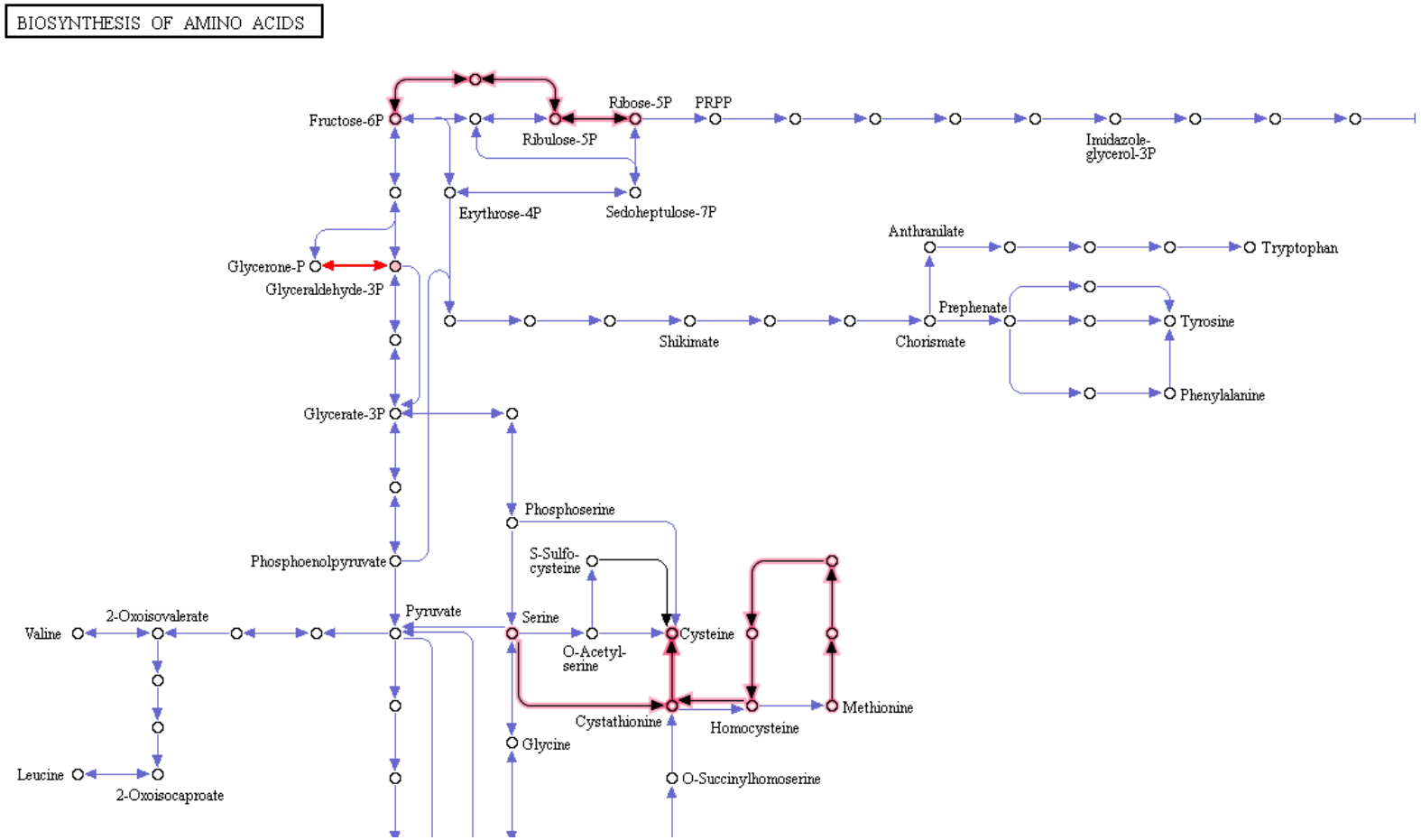
[Pathway menu | Organism menu | Pathway entry | Hide module list | Show description | User data mapping | Image (png) file]

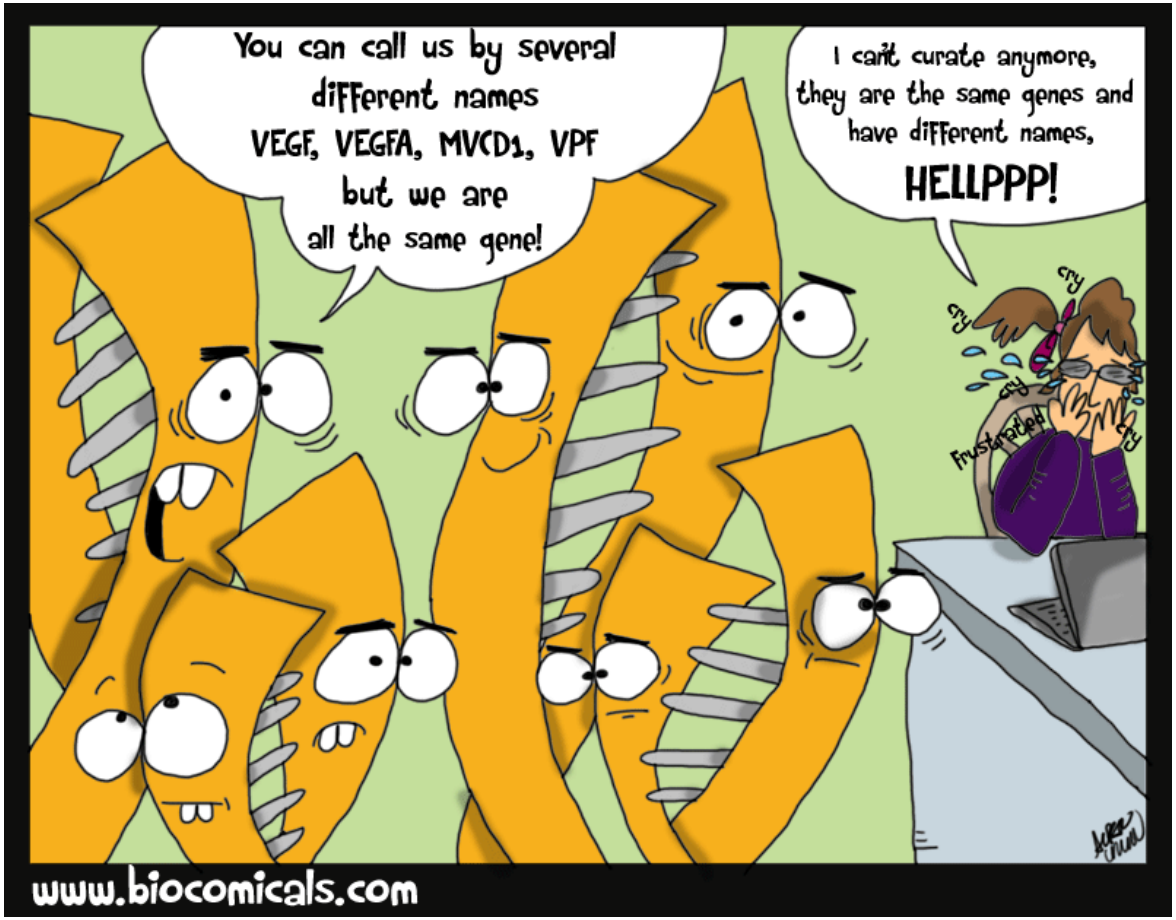
Reference pathway (KO) ▾ Go 100% ▾

KEGG module

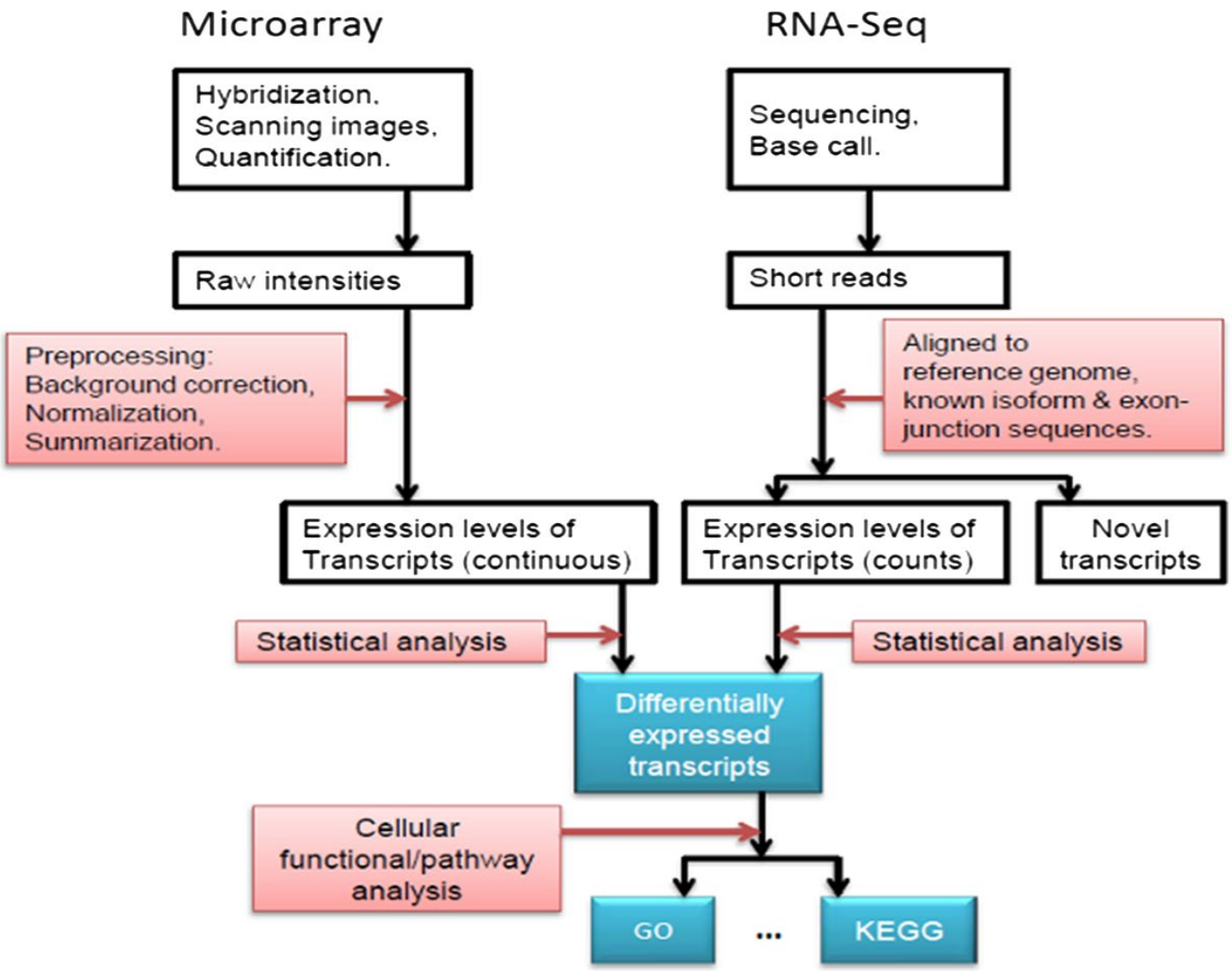
Carbohydrate and lipid metab
- Central carbohydrate me
  - M00002 Glycolysis, co
  - M00010 Citrate cycle,
  - M00007 Pentose phos
  - M00580 Pentose phos
  - M00005 PRPP biosyntl

Nucleotide and amino acid me
- Serine and threonine me
  - M00020 Serine biosyn
  - M00018 Threonine bic
- Cysteine and methionine
  - M00021 Cysteine bios
  - M00338 Cysteine bios
  - M00609 Cysteine bios
  - M00017 Methionine bi
- Branched-chain amino a
  - M00019 Valine/isoleuc
  - M00535 Isoleucine bic
  - M00570 Isoleucine bic
  - M00432 Leucine biosy
- Lysine metabolism
  - M00016 Lysine biosyn
  - M00525 Lysine biosyn
  - M00526 Lysine biosyn
  - M00527 Lysine biosyn
  - M00030 Lysine biosyn
  - M00433 Lysine biosyn
  - M00031 Lysine biosyn
- Arginine and proline met
  - M00015 Proline biosyr
  - M00028 Ornithine bios
  - M00763 Ornithine bios
  - M00029 Urea cycle
- Histidine metabolism
  - M00026 Histidine bios




You can call us by several different names VEGF, VEGFA, MVCD1, VPF but we are all the same gene!
I can't curate anymore, they are the same genes and have different names,
HELLPPP!
cry
cry
cry
cry
cry
Frustrated!
www.biocomicals.com

Overview of analysis workflow for microarray and RNA-seq transcriptional profiling



Fang Z et.al. Cell Biosci. 2012;2:26

Before Analysis

- ✓ Normalization
- ✓ Background adjustment
- ✓ Quality control (garbage in, garbage out)

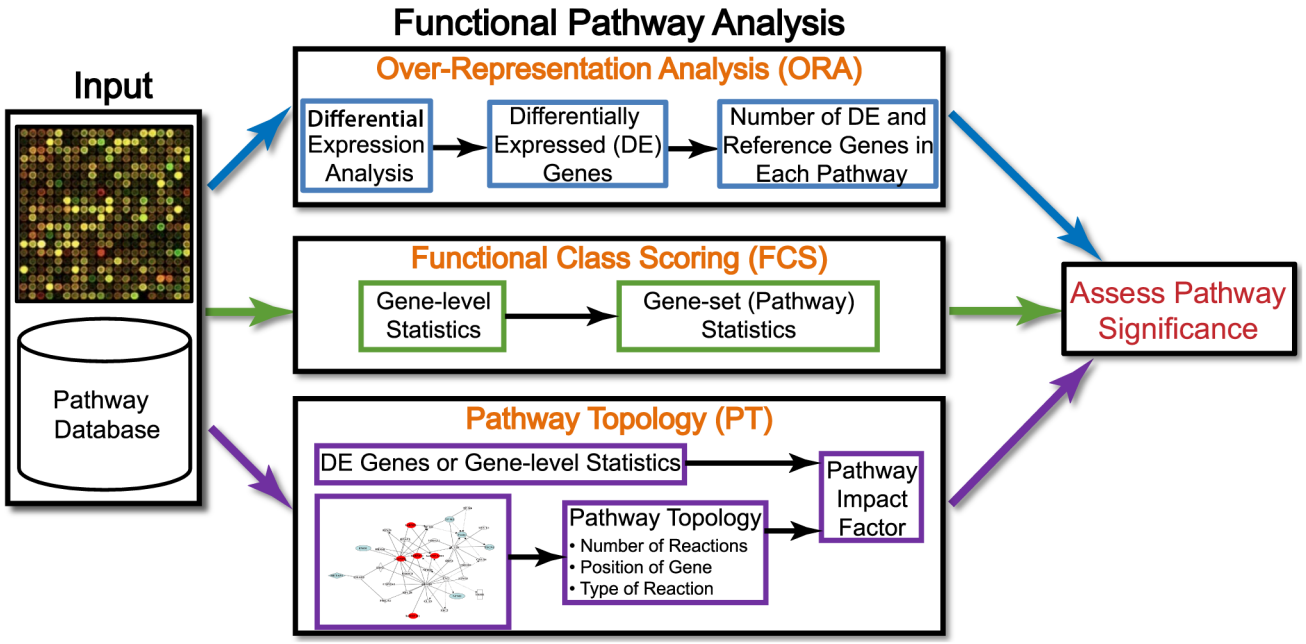
- ✓ Use statistics that will increase signal and reduce noise specifically for your experiment
- ✓ Gene list size
- ✓ Make sure your gene IDs are compatible with software

to Gary Bader

Recommendations

- For proteins and genes
  - (doesn't consider splice forms)
- Map everything to Entrez Gene IDs or Official Gene Symbols using a spreadsheet
- If 100% coverage desired, manually curate missing mappings using multiple resources
- Be careful of Excel auto conversions – especially when pasting large gene lists!
  - Remember to format cells as 'text' before pasting

to Gary Bader



Overview of existing pathway analysis methods using gene expression data as an example

PLoS Comput Biol. 2012;8(2):e1002375

Name	Availability	Reference
ORA tools		
Onto-Express	Web (http://vortex.cs.wayne.edu)	[4,5]
GenMAPP	Standalone (http://www.genmapp.org)	[11,71]
GoMiner	Standalone, Web (http://discover.nci.nih.gov/gominer)	[72,73]
FatiGO	Web (http://babelomics.bioinfo.cipf.es)	[74]
GOstat	Web (http://gostat.wehi.edu.au)	[7]
FuncAssociate	Web (http://llama.mshri.on.ca/funcassociate/)	[6]
GOToolBox	Web (http://genome.crg.es/GOToolBox/)	[10]
GeneMerge	Standalone, Web (http://genemerge.cbcb.umd.edu/)	[9]
GOEAST	Web (http://omicslab.genetics.ac.cn/GOEAST/)	[75]
ClueGO	Standalone (http://www.ici.upmc.fr/cluego/)	[76]
FunSpec	Web (http://funspec.med.utoronto.ca/)	[77]
GARBAN	Web	[78]
GO:TermFinder	Standalone (http://search.cpan.org/dist/GO-TermFinder/)	[8]
WebGestalt	Web (http://bioinfo.vanderbilt.edu/webgestalt/)	[79]
agriGO	Web (http://bioinfo.cau.edu.cn/agriGO/)	[80]
GOFFA	Standalone, Web (http://edkb.fda.gov/webstart/arraytrack/)	[81]
WEGO	Web (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)	[82]
FCS tools		
GSEA	Standalone (http://www.broadinstitute.org/gsea/)	[21,29]
sigPathway	Standalone (BioConductor)	[22]
Category	Standalone (BioConductor)	[24]
SAFE	Standalone (BioConductor)	[30]
GlobalTest	Standalone (BioConductor)	[15]
PCOT2	Standalone (BioConductor)	[17]
SAM-GS	Standalone (http://www.ualberta.ca/~yyasui/software.html)	[83]
Catmap	Standalone (http://bioinfo.thep.lu.se/catmap.html)	[84]
T-profiler	Web (http://www.t-profiler.org)	[85]
FunCluster	Standalone (http://corneliu.henegar.info/FunCluster.htm)	[86]
GeneTrail	Web (http://genetrail.bioinf.uni-sb.de)	[87]
GAzer	Web	[88]
PT-based tools		
ScorePAGE	No implementation available	[37]
Pathway-Express	Web (http://vortex.cs.wayne.edu)	[38,39]
SPIA	Standalone (BioConductor)	[40]
NetGSA	No implementation available	[43]

doi:10.1371/journal.pcbi.1002375.t001

Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states

http://www.broadinstitute.org/gsea/index.jsp

GSEA
Gene Set Enrichment Analysis

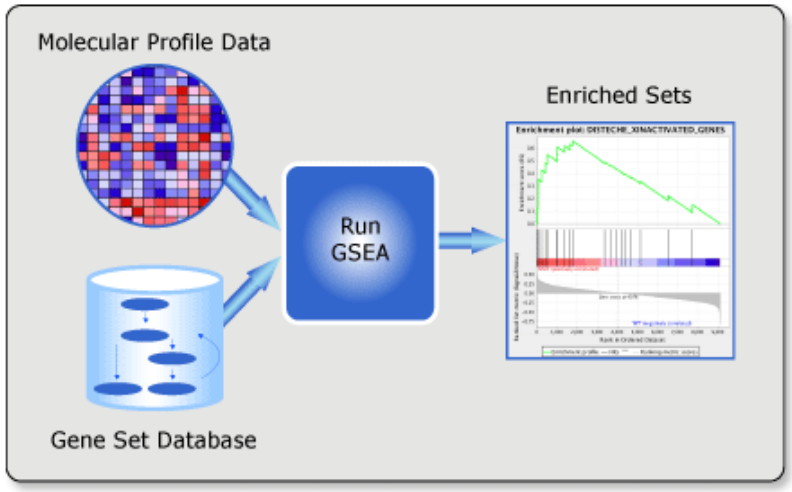
GSEA Home | Downloads | Molecular Signatures Database | Documentation | Contact

Overview

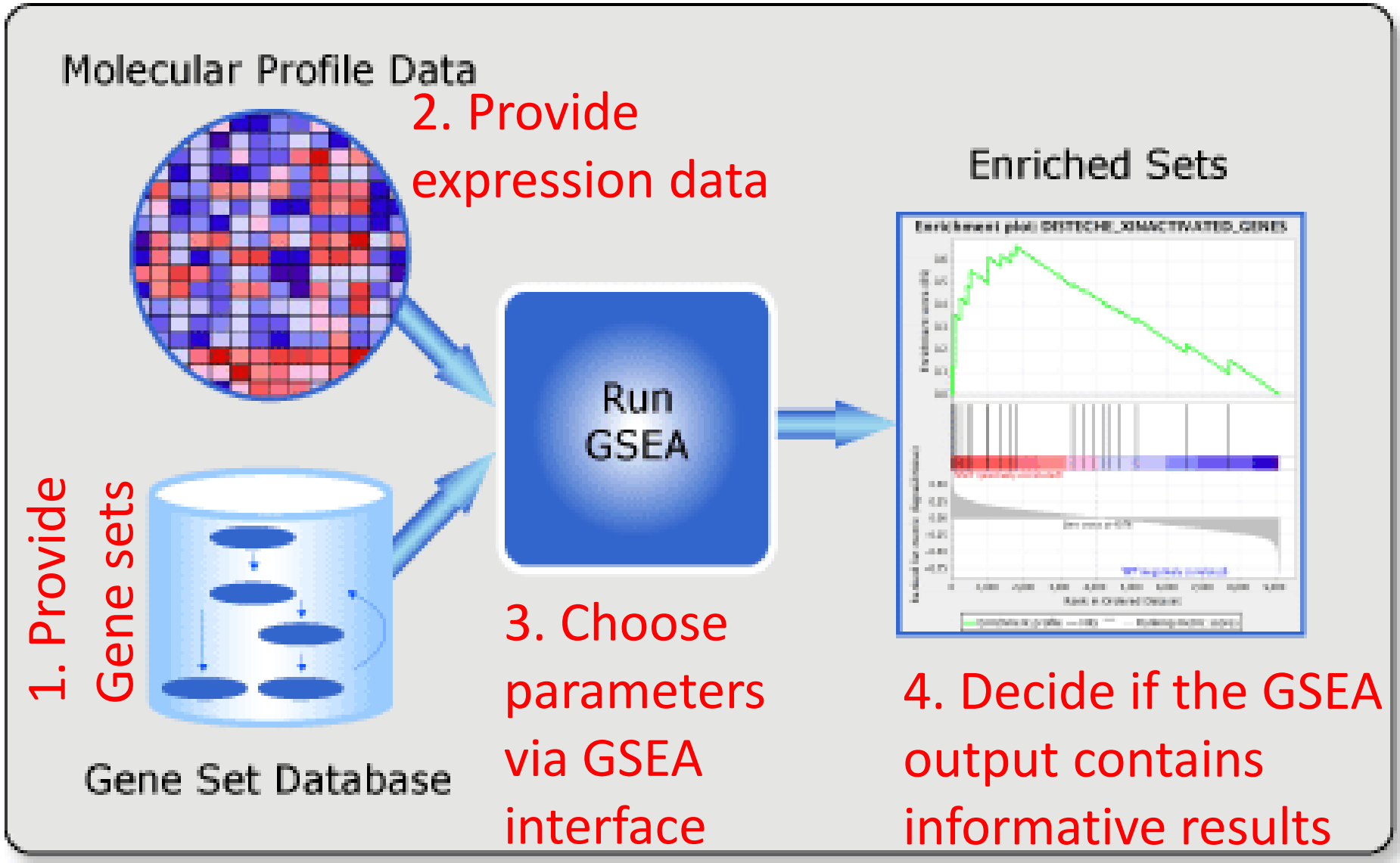
Gene Set Enrichment Analysis (GSEA) is a computational method that determines whether an a priori defined set of genes shows statistically
significant, concordant differences between two biological states (e.g. phenotypes).

From this web site, you can:

- **Download** the GSEA software and additional resources to analyze, annotate and interpret enrichment results.
- **Explore the Molecular Signatures Database (MSigDB)**, a collection of annotated gene sets for use with GSEA software.
- **View documentation** describing GSEA and MSigDB.



Running GSEA



Main Data Files for Input

- **Gene sets database files**
  - **GeneMatrix (filename.gmt) in local machine (download from the Broad MSigDB)**
- **Expression Data files (expression values or differential expression levels)**
  - **Gene Cluster Text file: filename.gct (full expression data)**
  - **Ranked list file format: filename.rnk (condensed differential expression data)**
  - **ExpRESsion (with P and A calls) file: filename.res (a format for Affymetrix data)**
- **Phenotype information files (categorical = specify the group for each sample)**
  - **Need a .cls file if using full expression data (.gct)**
  - **Categorical (e.g. tumor vs normal) class file format: filename.cls**
  - **Continuous (e.g. clinical data or a gene profile) file format: filename.cls**

Information available from the Broad Institute GSEA documentation

Preparing Data Files for GSEA

When you use GSEA, you supply four data files: an expression dataset file, phenotype labels file, gene sets file, and chip annotations file. The following table lists each data file and its valid file formats. All files are tab-delimited ASCII text files; they can be created and edited using any text editor.

For descriptions and examples of each file format, see GSEA file formats. For more information about each data file, click the data file link in the following table.

Data File	Content	Format	Source
Expression dataset	Contains features (genes or probes), samples, and an expression value for each feature in each sample. Expression data can come from any source (Affymetrix, Stanford cDNA, and so on).	res, gct, pcl, or txt	You create the file.
Phenotype labels	Contains phenotype labels and associates each sample with a phenotype.	cls	You create the file or have GSEA create it for you.
Gene sets	Contains one or more gene sets. For each gene set, gives the gene set name and list of features (genes or probes) in that gene set.	gmx or gmt	You use the files on the Broad ftp site, export gene sets from the Molecular Signature Database (MSigDb) or create your own gene sets file.
Chip annotations	Lists each probe on a DNA chip and its matching HUGO gene symbol. Optional for the gene set enrichment analysis.	Chip	You use the files on the Broad ftp site, download the files from the GSEA web site, or create your own chip file.

The expression dataset, gene sets, and chip annotation files all contain lists of features (genes or probes) to be analyzed. It is critical that you use the same feature (gene or probe) identifiers across all of the data files. For more information, see Consistent Feature Identifiers Across Data Files.

Consistent Feature Identifiers Across Data Files

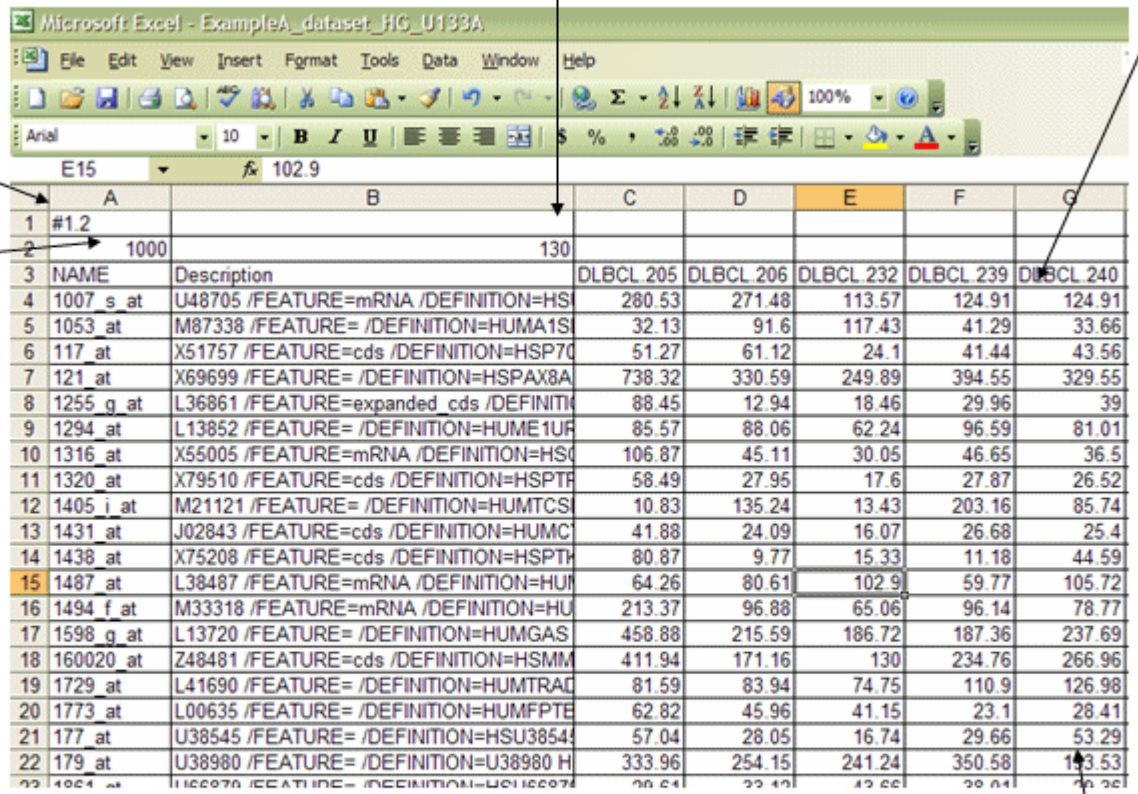
Typically, the feature identifiers in your expression dataset are the probe identifiers for the DNA chip array used to produce the data. For example, an expression dataset produced using the HG_U133A chip contains HG_U133A probe identifiers and an expression dataset produced using the HG_U95Av2 chip contains HG_U95Av2 probe identifiers. When using GSEA, it is critical that your expression dataset, gene sets, and chip annotation files all use the same feature identifiers.

Typically, you use one of two approaches to ensure that you are using consistent feature identifiers across files: you collapse your probe sets into genes and use the HUGO gene symbols as your consistent feature identifiers or you use the probe identifiers from your expression dataset as your consistent feature identifiers.

- **HUGO gene symbols.** To use HUGO gene symbols as your consistent feature identifiers:

From http://www.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html

GCT: Gene Cluster Text file format (*.gct)



	A	B	C	D	E	F	G
1	#1.2						
2	1000	130					
3	NAME	Description	DLBCL 205	DLBCL 206	DLBCL 232	DLBCL 239	DLBCL 240
4	1007_s_at	U48705 /FEATURE=mRNA /DEFINITION=HS	280.53	271.48	113.57	124.91	124.91
5	1053_at	M87338 /FEATURE= /DEFINITION=HUMA1S	32.13	91.6	117.43	41.29	33.66
6	117_at	X51757 /FEATURE=cds /DEFINITION=HSP7	51.27	61.12	24.1	41.44	43.56
7	121_at	X69699 /FEATURE= /DEFINITION=HSPAX8A	738.32	330.59	249.89	394.55	329.55
8	1255_g_at	L36861 /FEATURE=expanded_cds /DEFINITI	88.45	12.94	18.46	29.96	39
9	1294_at	L13852 /FEATURE= /DEFINITION=HUME1UR	85.57	88.06	62.24	96.59	81.01
10	1316_at	X55005 /FEATURE=mRNA /DEFINITION=HS	106.87	45.11	30.05	46.65	36.5
11	1320_at	X79510 /FEATURE=cds /DEFINITION=HSPTP	58.49	27.95	17.6	27.87	26.52
12	1405_i_at	M21121 /FEATURE= /DEFINITION=HUMTCS	10.83	135.24	13.43	203.16	85.74
13	1431_at	J02843 /FEATURE=cds /DEFINITION=HUMC	41.88	24.09	16.07	26.68	25.4
14	1438_at	X75208 /FEATURE=cds /DEFINITION=HSPTH	80.87	9.77	15.33	11.18	44.59
15	1487_at	L38487 /FEATURE=mRNA /DEFINITION=HUI	64.26	80.61	102.9	59.77	105.72
16	1494_f_at	M33318 /FEATURE=mRNA /DEFINITION=HU	213.37	96.88	65.06	96.14	78.77
17	1598_g_at	L13720 /FEATURE= /DEFINITION=HUMGAS	458.88	215.59	186.72	187.36	237.69
18	160020_at	Z48481 /FEATURE=cds /DEFINITION=HSMM	411.94	171.16	130	234.76	266.96
19	1729_at	L41690 /FEATURE= /DEFINITION=HUMTRAD	81.59	83.94	74.75	110.9	126.98
20	1773_at	L00635 /FEATURE= /DEFINITION=HUMFPTE	62.82	45.96	41.15	23.1	28.41
21	177_at	U38545 /FEATURE= /DEFINITION=HSU3854	57.04	28.05	16.74	29.66	53.29
22	179_at	U38980 /FEATURE= /DEFINITION=U38980 H	333.96	254.15	241.24	350.58	193.53

Always "#1.2"

The # of rows (i.e probe sets)

# of samples

Third column onwards are sample names. These must be UNIQUE

Data starts on line 4

Column 1: Row identifiers. Typically probe set ids or clone ids. These must be UNIQUE

Column 2: Row descriptions. Ignored by the program – can be dummy values (e.g. "na")

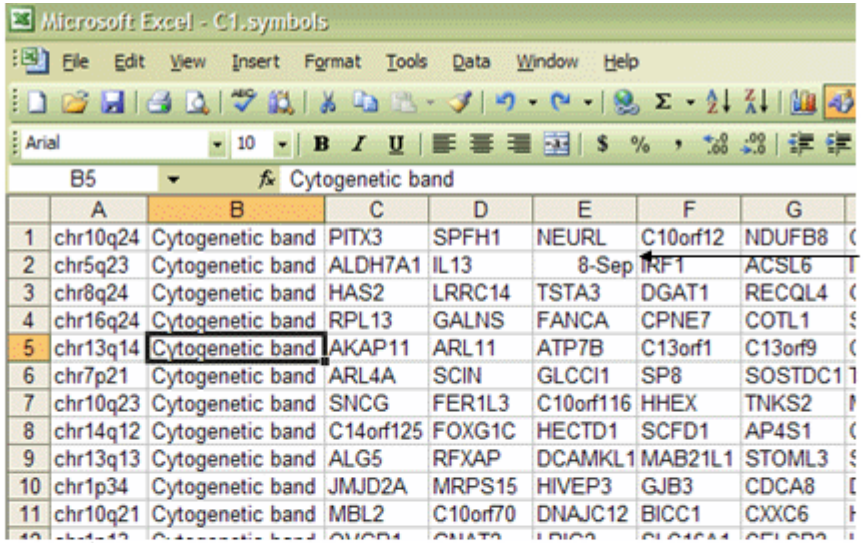
Each column contains expression values from 1 sample. Missing values are allowed (leave empty).

If editing, in Excel, make sure to save your data as "tab delimited text"

http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats

GMT: Gene Matrix Transposed file format (*.gmt)

Each row represents one gene set



	A	B	C	D	E	F	G
1	chr10q24	Cytogenetic band	PITX3	SPFH1	NEURL	C10orf12	NDUFB8
2	chr5q23	Cytogenetic band	ALDH7A1	IL13	8-Sep	IRF1	ACSL6
3	chr8q24	Cytogenetic band	HAS2	LRRC14	TSTA3	DGAT1	RECQL4
4	chr16q24	Cytogenetic band	RPL13	GALNS	FANCA	CPNE7	COTL1
5	chr13q14	Cytogenetic band	AKAP11	ARL11	ATP7B	C13orf1	C13orf9
6	chr7p21	Cytogenetic band	ARL4A	SCIN	GLCCI1	SP8	SOSTDC1
7	chr10q23	Cytogenetic band	SNCG	FER1L3	C10orf116	HHEX	TNKS2
8	chr14q12	Cytogenetic band	C14orf125	FOXG1C	HECTD1	SCFD1	AP4S1
9	chr13q13	Cytogenetic band	ALG5	RFXAP	DCAMKL1	MAB21L1	STOML3
10	chr1p34	Cytogenetic band	JMJD2A	MRPS15	HIVEP3	GJB3	CDCA8
11	chr10q21	Cytogenetic band	MBL2	C10orf70	DNAJC12	BICC1	CXXC6

If editing in excel, watch out for its tendency to auto-format gene sets (SEP8 becomes 8-Sep)

First column are gene set names. Duplicates are not allowed

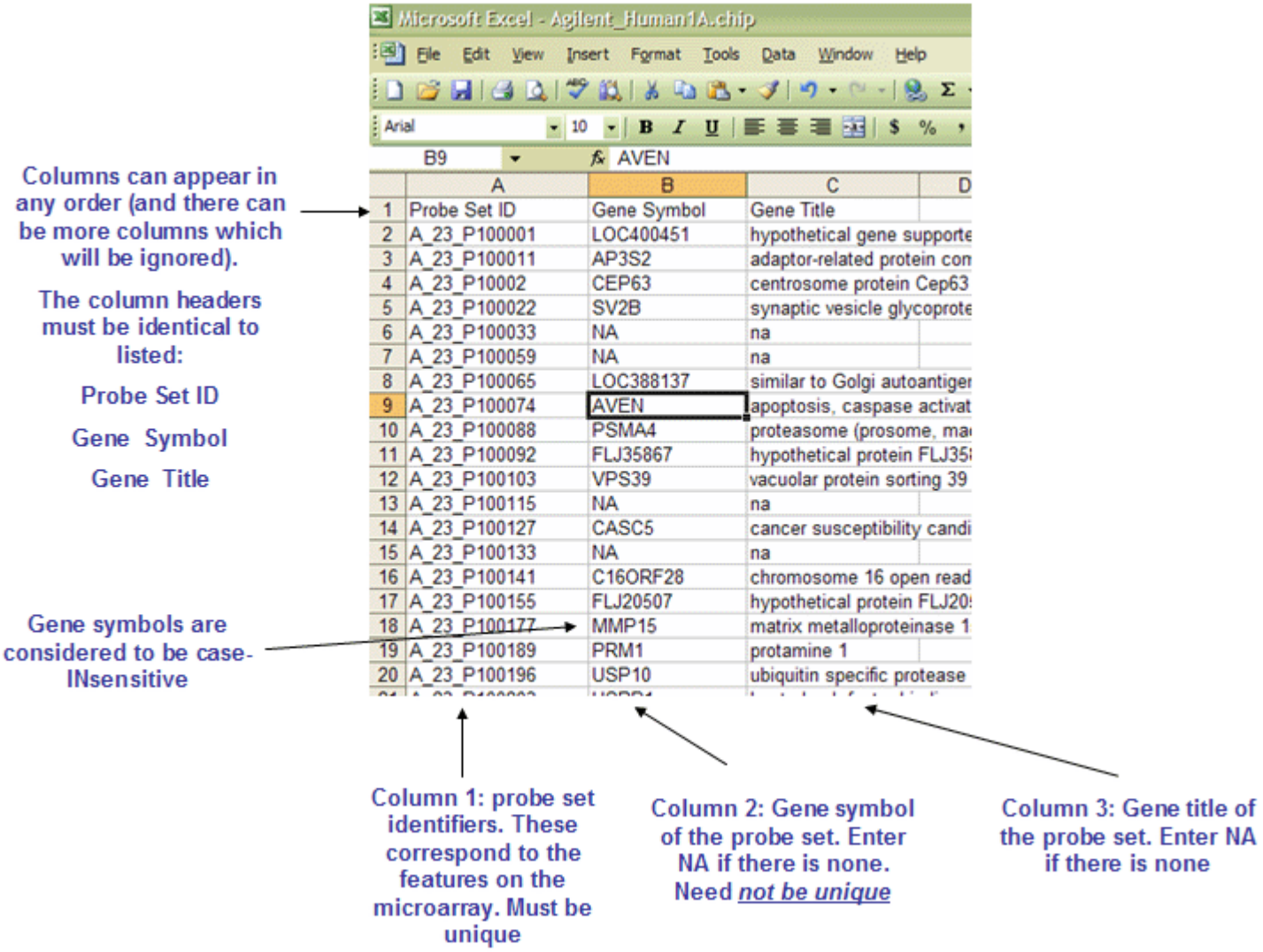
Second column contains a brief description. Its optional – you can fill in a dummy field (e.g. "na")

Unequal lengths (i.e # of genes) *is* allowed

GMT format is convenient to store large databases of gene sets. For a handful of sets (<256) the gmx format offers greater excel-editability

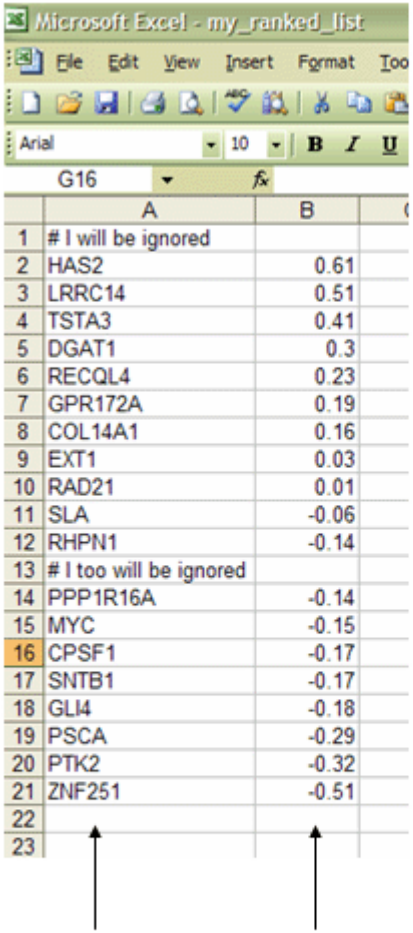
CHIP: Chip file format (*.chip)

The CHIP file contains **annotation** about a microarray. It should list the features (i.e probe sets) used in the microarray along with their mapping to gene symbols (when available). While this file is not used directly in the GSEA algorithm, it is used to annotate the output results and may also be used to collapse each probe set in the expression dataset to a single gene vector.



RNK: Ranked list file format (*.rnk)

The RNK file contains a single, rank ordered gene list (*not* gene set) in a simple newline-delimited text format. It is used when you have a pre-ordered ranked list that you want to analyze with GSEA. For instance, you might have used your favorite **tTest-like statistic** to produce a ranked ordered gene list from your dataset which you now want to test for enrichment. Order of lines does not matter. It is important, however, that the second column will have numeric values - they will be used to rank order genes by GSEA.

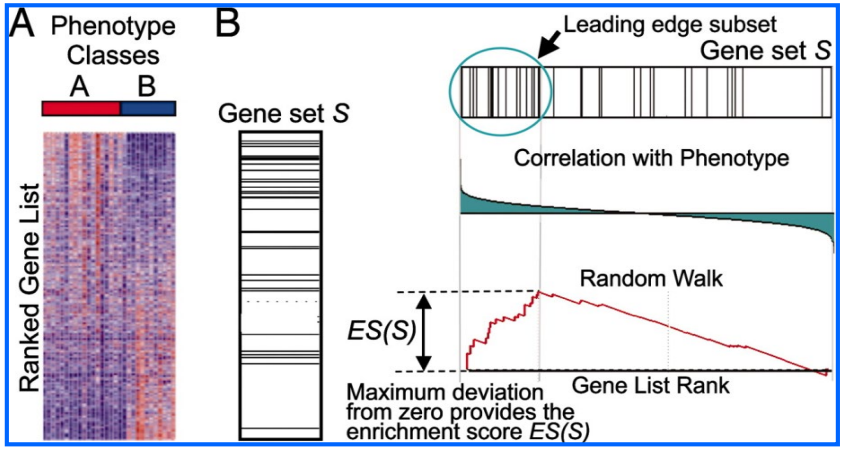


Column 1:contains feature identifiers (i.e gene symbols, affy probe sets ids etc)

Column 2:contains the weight (i.e class-difference metric) for this feature. The list need not be sorted.

GSEA algorithm

- Calculate the enrichment score of each gene set
  - Move down the list and increment a counter
    - If the current gene is a member of the set, add quantity whose magnitude depends on the gene's correlation with the class variable
    - For non-members subtract this quantity
    - The enrichment score (ES) is the maximum deviation from 0
- Estimate the significance level of the gene set enrichment values
  - Repeatedly permute the class labels (1000 times) and then calculate the enrichment values of the gene sets
  - The significance level of each gene set is determined by this null distribution
    - For example, if the enrichment value of a gene set is greater than all but 5 of the enrichments values from the permutation, then its p-value is 0.005.



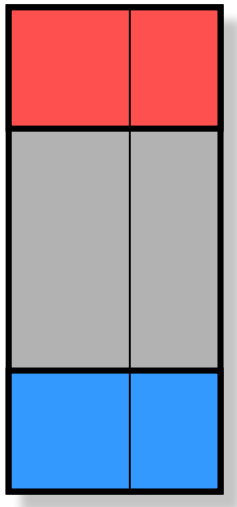
Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50

GSEA Enrichment tests

ENRICHMENT TEST

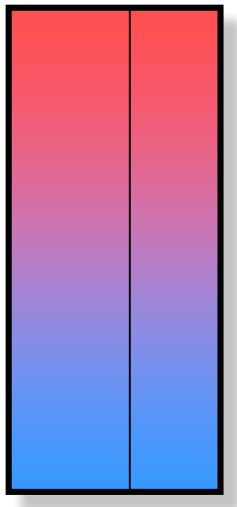
Gene list
Fisher's Exact Test, Binomial and Chi-squared.

Ranked list (semi-quantitative)
Minimum hypergeometric test, **GSEA**, Wilcoxon ranksum, Mann-Whitney U, Kolmogorov-Smirnov



UP

DOWN



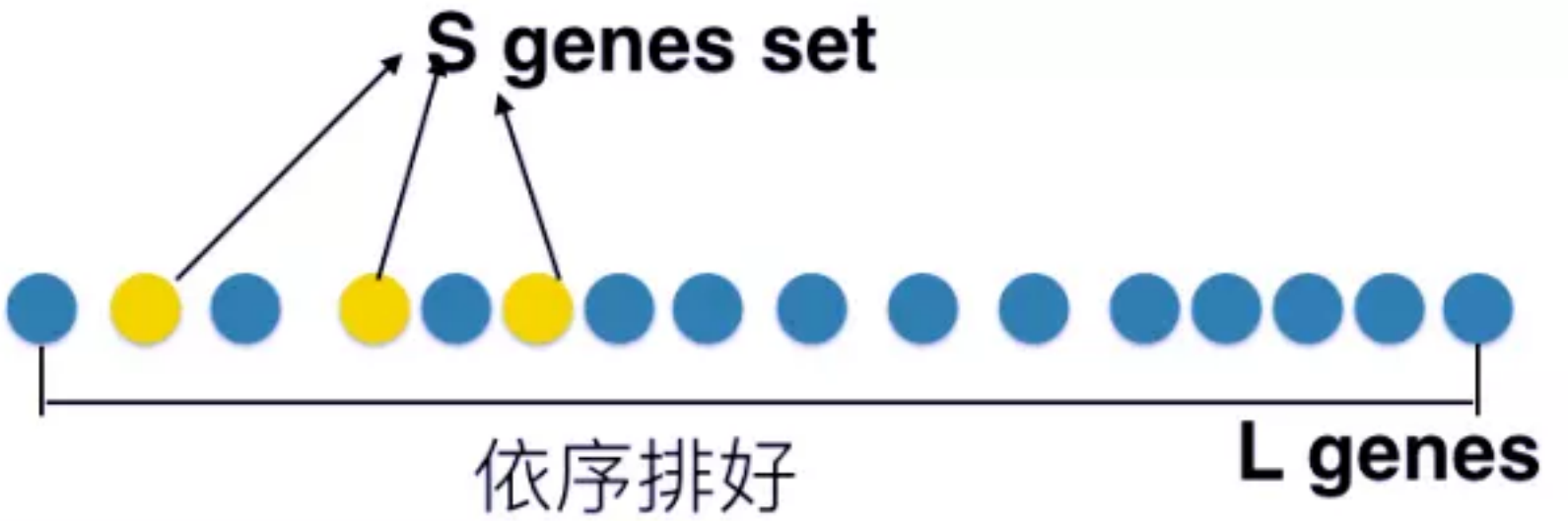
UP

DOWN

Quaid Morris

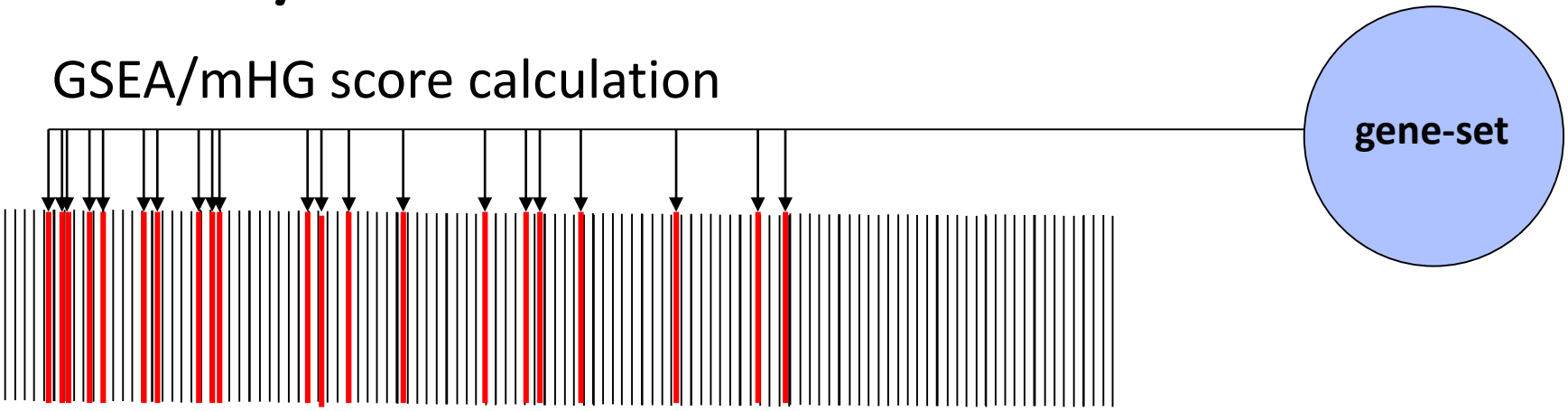
to Quaid Morris

GSEA原理



GSEA/mHG: Method

GSEA/mHG score calculation

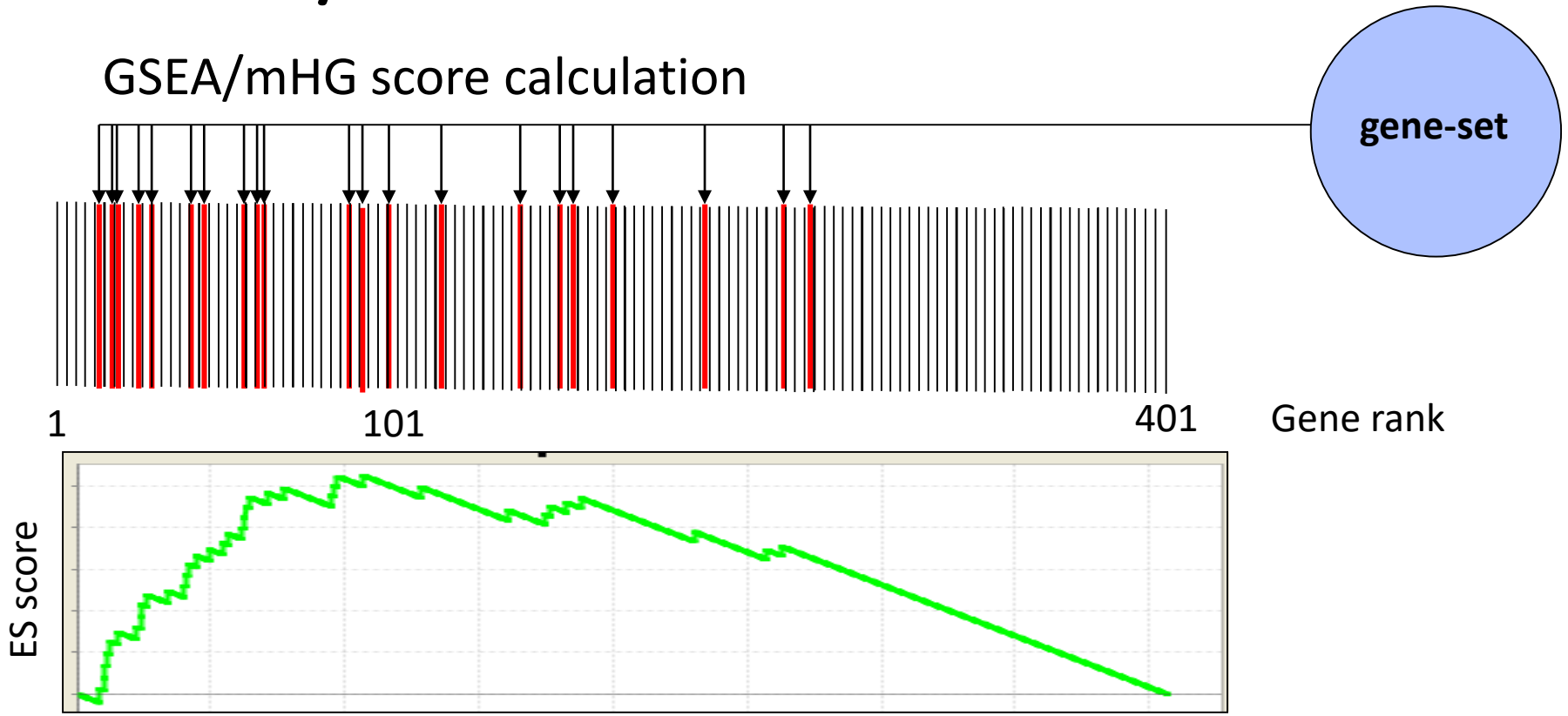


Where are the gene-set genes located in the ranked list?
Is there distribution random, or is there an enrichment in either end?

to Quaid Morris

GSEA/mHG: Method

GSEA/mHG score calculation



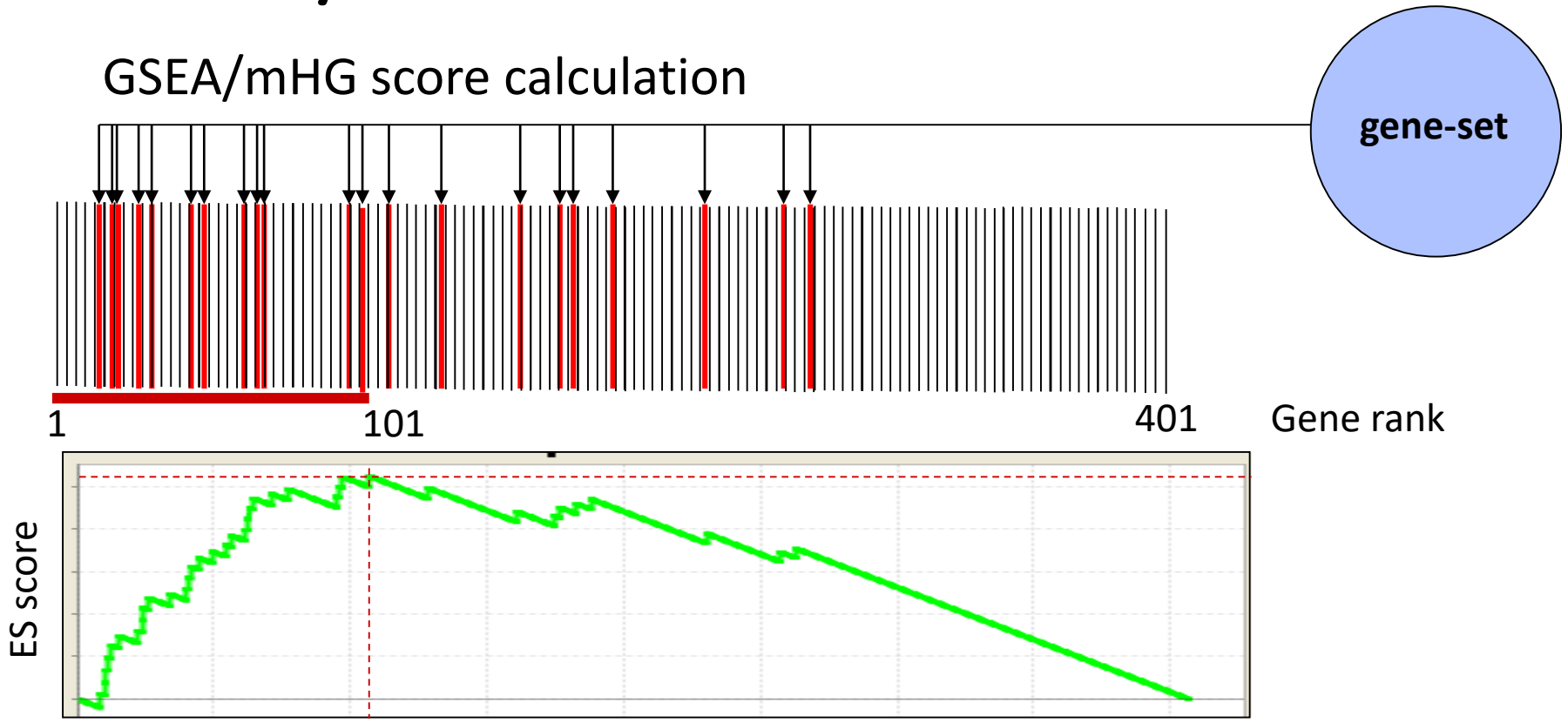
Every present gene (thick red vertical bar) gives a positive contribution,
Every absent gene (black vertical bar) gives a negative contribution

Warning: the alignment here between bars and plot is a little off

to Quaid Morris

GSEA/mHG: Method

GSEA/mHG score calculation



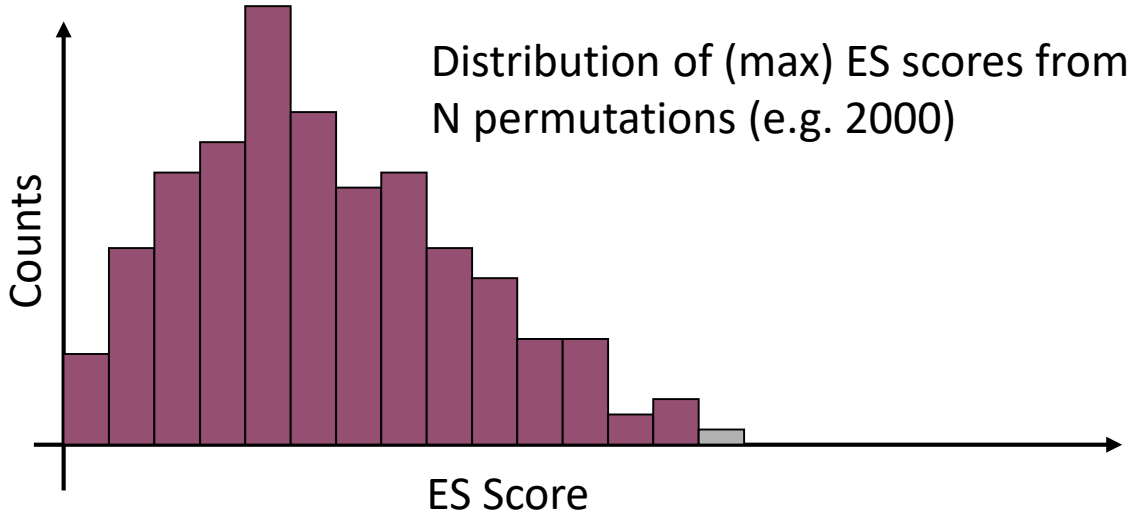
1. Maximum (or minimum) ES score is the final **ES score** for the gene set
2. Can define “leading edge subset” as all those genes ranked as least as high as the enriched set.

to Quaid Morris

Permutation-based P-values

Empirical p-value estimation (for every gene-set)

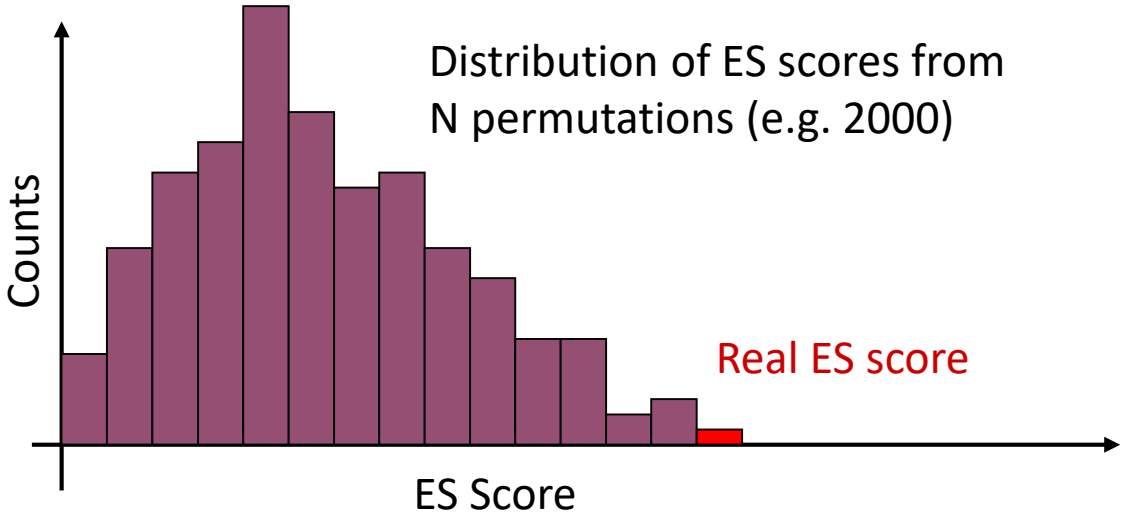
1. Generate null-hypothesis distribution from randomized data (see permutation settings)



to Quaid Morris

Permutation-based P-values

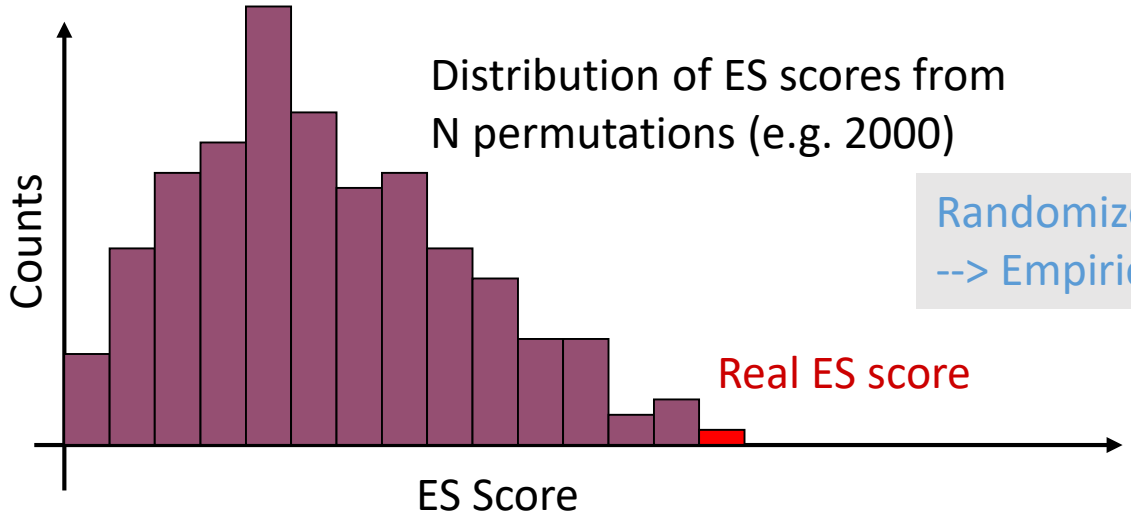
Estimate empirical p-value by comparing observed max ES score to null-hypothesis distribution from randomized data (for every gene-set)



to Quaid Morris

Permutation-based P-values

Estimate empirical p-value by comparing observed max ES score to null-hypothesis distribution from randomized data (for every gene-set)

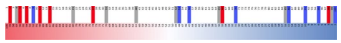


Randomized with ES score ≥ real: 4 / 2,000
--> Empirical p-value = 0.002

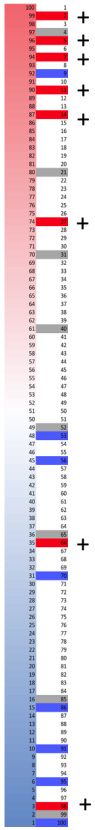
to Quaid Morris

More GSEA examples

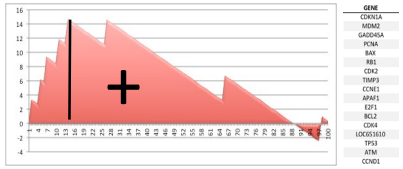
Ranked gene list



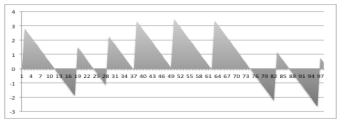
UP in treated

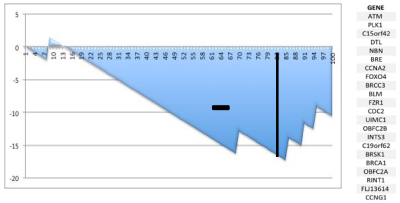


DOWN in treated



Enrichment Score





Enrichment Score

Gene-set 1 (n=17)
Enriched in treated

Gene-set 2
Not enriched

Gene-set 3 (n=22)
Depleted in treated

to Quaid Morris

Molecular Signatures Database v6.2



MSigDB
Molecular Signatures Database

Overview

The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. From this web site, you can

- **Search** for gene sets by keyword.
- **Browse** gene sets by name or collection.
- **Examine** a gene set and its annotations. See, for example, the GO_NOTCH_SIGNALING_PATHWAY gene set page.
- **Download** gene sets.
- **Investigate** gene sets:
  - **Compute overlaps** between your gene set and gene sets in MSigDB.
  - **Categorize** members of a gene set by gene families.
  - **View the expression profile** of a gene set in any of the three provided public expression compendia.

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** for each human chromosome and cytogenetic band.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3 **motif gene sets** based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.

C4 **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

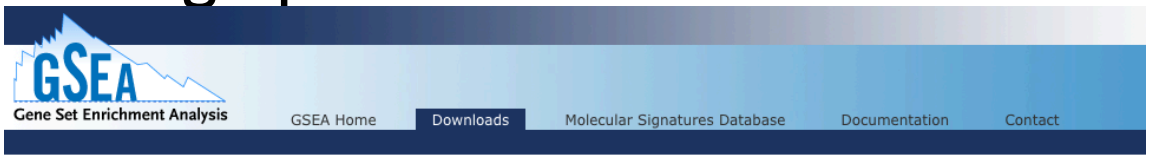
C5 **GO gene sets** consist of genes annotated by the same GO terms.

C6 **oncogenic signatures** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic signatures** defined directly from microarray gene expression data from immunologic studies.

- **H** (hallmark gene sets, 50 gene sets)
- **C1** (positional gene sets, 326 gene sets)
  - by chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
- **C2** (curated gene sets, 4762 gene sets)
  - **CGP** (chemical and genetic perturbations, 3433 gene sets)
  - **CP** (Canonical pathways, 1329 gene sets)
  - **CP:BIOCARTA** (BioCarta gene sets, 217 gene sets)
  - **CP:KEGG** (KEGG gene sets, 186 gene sets)
  - **CP:REACTOME** (Reactome gene sets, 674 gene sets)
- **C3** (motif gene sets, 836 gene sets)
  - **MIR** (microRNA targets, 221 gene sets)
  - **TFT** (transcription factor targets, 615 gene sets)
- **C4** (computational gene sets, 858 gene sets)
  - **CGN** (cancer gene neighborhoods, 427 gene sets)
  - **CM** (cancer modules, 431 gene sets)
- **C5** (GO gene sets, 5917 gene sets)
  - **BP** (GO biological process, 4436 gene sets)
  - **CC** (GO cellular component, 580 gene sets)
  - **MF** (GO molecular function, 901 gene sets)
- **C6** (oncogenic signatures, 189 gene sets)
- **C7** (immunologic signatures, 4872 gene sets)

Easy-to-use graphical user interface



Downloads

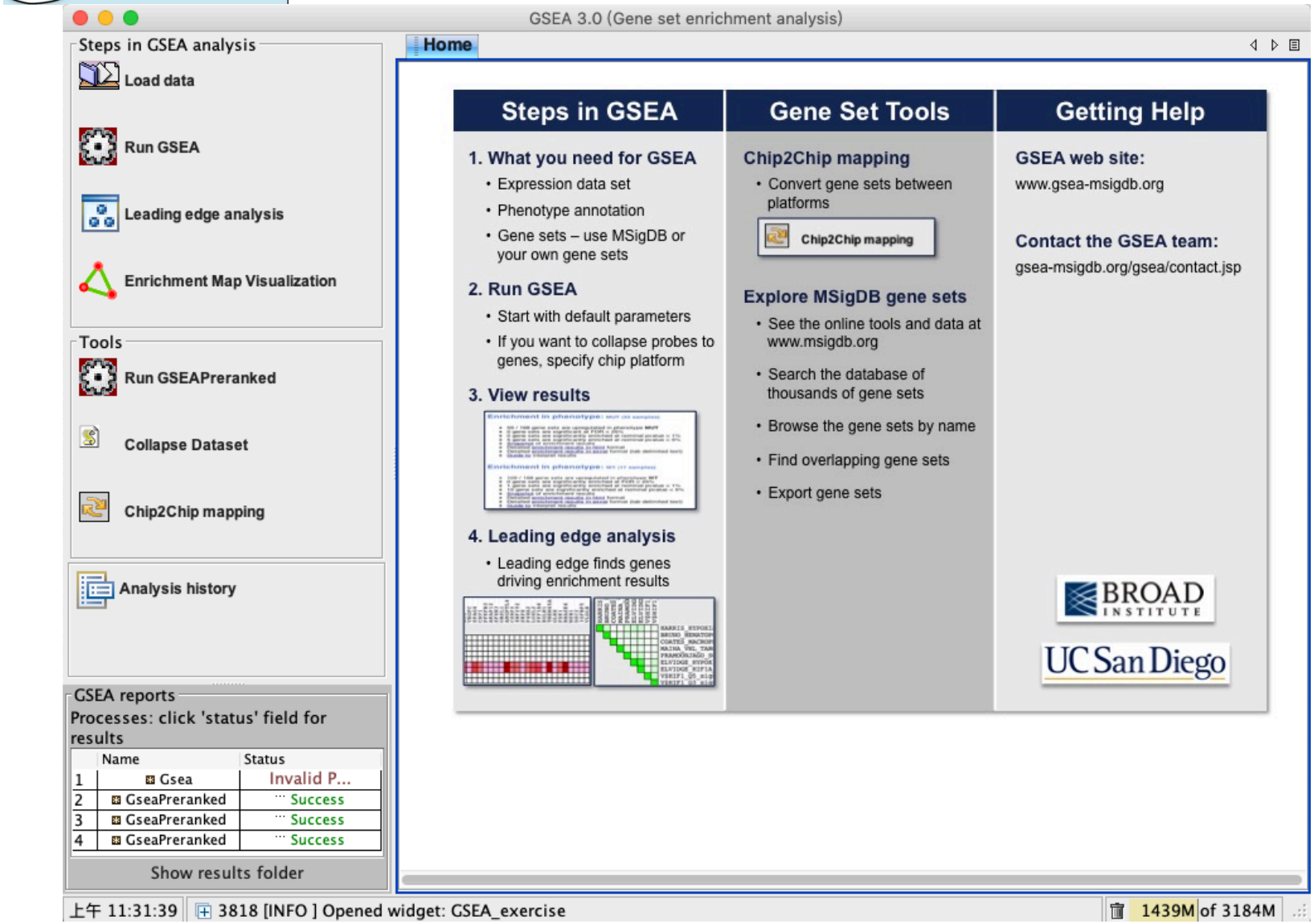
Software

There are several options for GSEA software. All options implement exactly the same algorithm. Usage recommendations and installation instructions are listed below. Current Java implementations of GSEA require Java 8.

See the license terms page for details about the license for the GSEA software and source code. Please note that the license terms vary for different versions of the software.

javaGSEA Desktop Application	▸ Easy-to-use graphical user interface. ▸ Runs on any desktop computer (Windows, macOS, Linux etc.) that supports Java 8. **Oracle Java is recommended as there are known issues when running with OpenJDK. Java 9 and higher are not supported at this time.** ▸ Produces richly annotated reports of enrichment results. ▸ This release is open source under a BSD-style license. The source is available on our GitHub repository. The changes are noted in the Release Notes. ▸ **We recommend using a memory configuration smaller than your computer's total memory.**	Launch with 1GB (for 32 or 64-bit Java) memory: Launch
javaGSEA Java Jar file	▸ Command line or offline usage. See our User Guide for details. ▸ Runs on any platform that supports Java 8. **Oracle Java is recommended as there are known issues when running with OpenJDK. Java 9 and higher are not supported at this time.** ▸ We recommend using the 'Launch' buttons above instead of this mode for most users.	download gsea-3.0.jar
BETA **MSigDB XML Browser Java Jar file**	▸ The current Beta version of the MSigDB XML Browser (formerly part of the GSEA Desktop). ▸ Please contact us with bugs or other feedback. We will aim to address problems as soon as possible in future Beta releases. ▸ Download and launch from the command line with 'java -jar MSigDB_XML_Browser-1.0_beta_4.jar', or double-click to launch. ▸ Runs on any platform that supports Java 8. **Oracle Java is recommended as there are known issues when running with OpenJDK. Java 9 and higher are not supported at this time.** ▸ This release is open source under a BSD-style license. The source is available on our GitHub repository.	**BETA** download MSigDB_XML_Browser-1.0_beta_4.jar
GenePattern GSEA Module	▸ Use GSEA from within GenePattern. ▸ Use GSEA in concert with a large suite of other analytics found in GenePattern (a powerful and flexible analysis platform developed at the Broad Institute).	GenePattern site

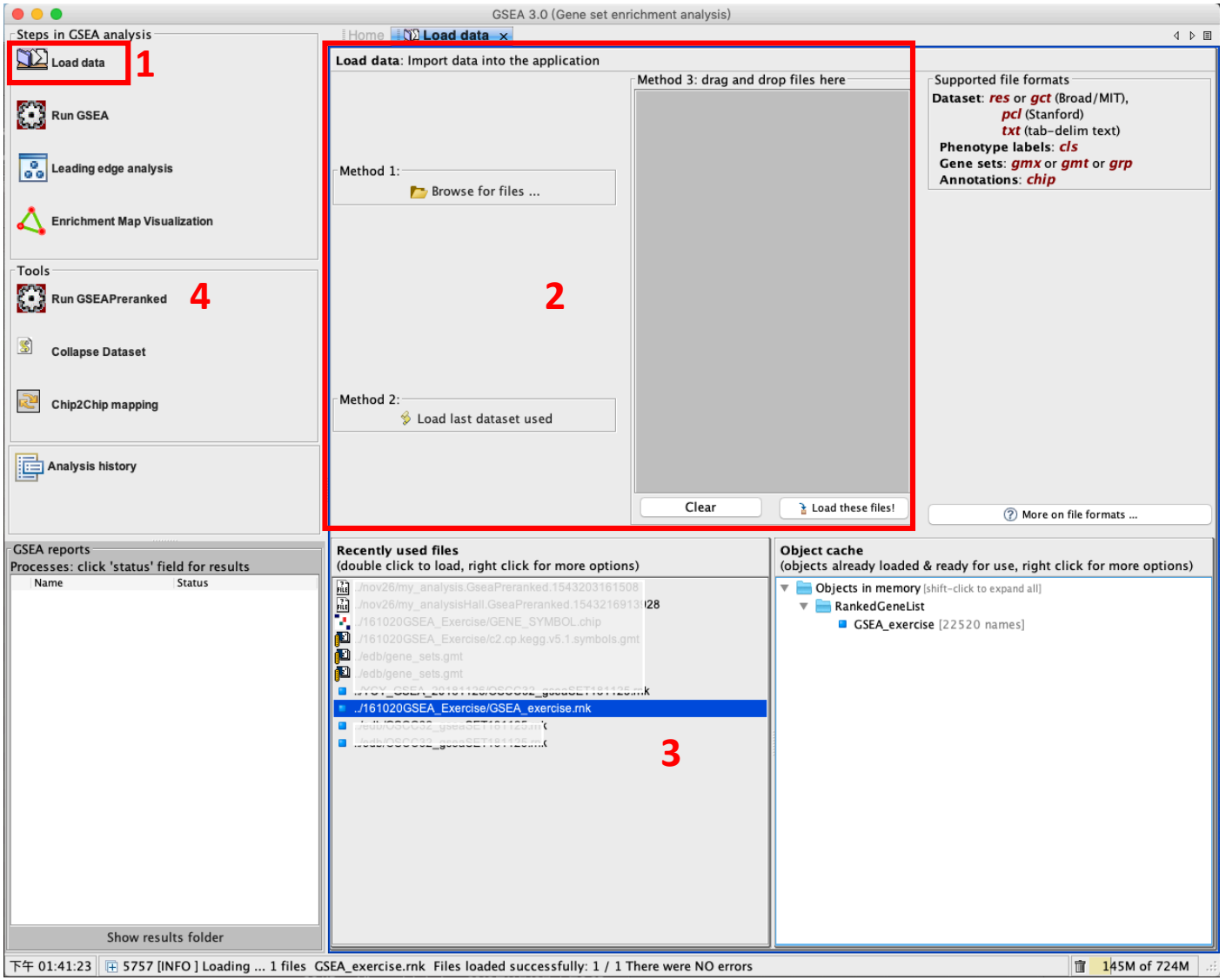
GSEA desktop application



Load data

	A	B
1	ZNF853	5.979982
2	HOXB-AS1	4.522795
3	MIR10A	4.372896
4	MIR196A2	4.163029
5	TRHDE	3.961958
6	HOXB7	3.888059
7	STMN2	3.256591
8	MAB21L1	3.237871
9	XIST	3.106976
10	HOXB3	2.956016
11	HOXC10	2.880392
12	RBP1	2.834098
13	PAX3	2.812718
14	NPTX1	2.811849
15	G0S2	2.758431
16	PLCXD1	2.562242
17	FLT1	2.517767
18	SCN9A	2.502511
19	HOXB6	2.432267
20	FRMD3	2.426308
21	SLC7A14	2.425239
22	HOXB4	2.318901
23	MIR18A	2.220983
24	LAMB3	2.218843
25	MIR708	2.187673
26	MRAP2	2.185362
27	AQP1	2.180603
28	TLR4	2.159465
29	ADRB2	2.157506

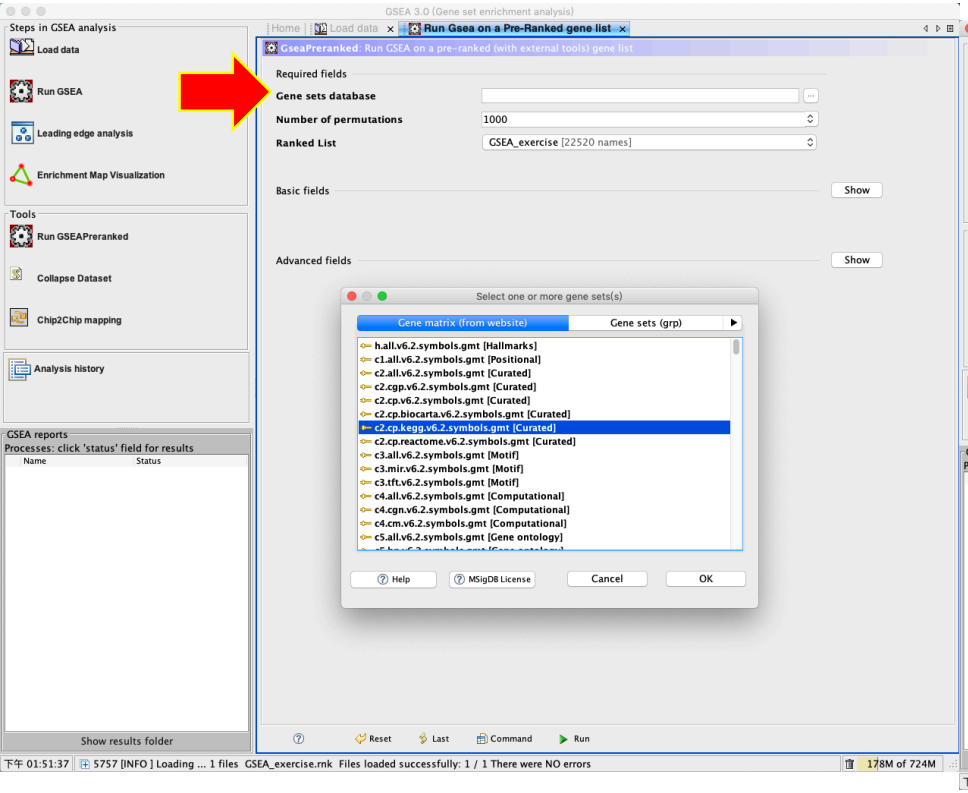
1. RPKM ratio
2. Sorting/unique
3. Save as .txt
4 ..txt → .rnk

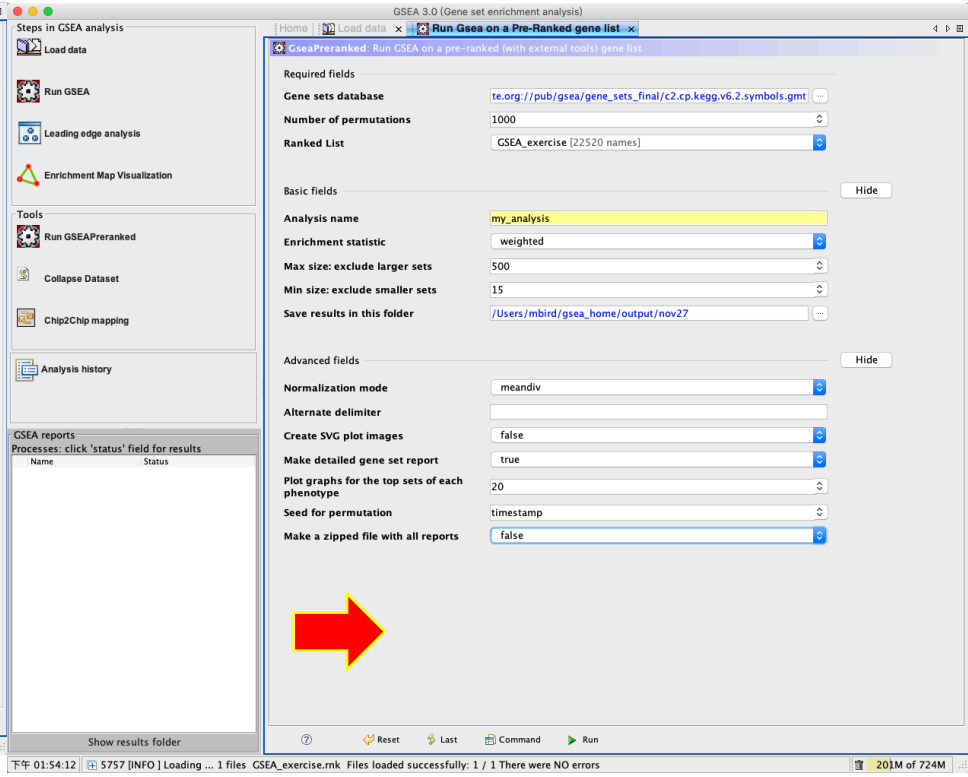


Parameters of pre-ranked GSEA I

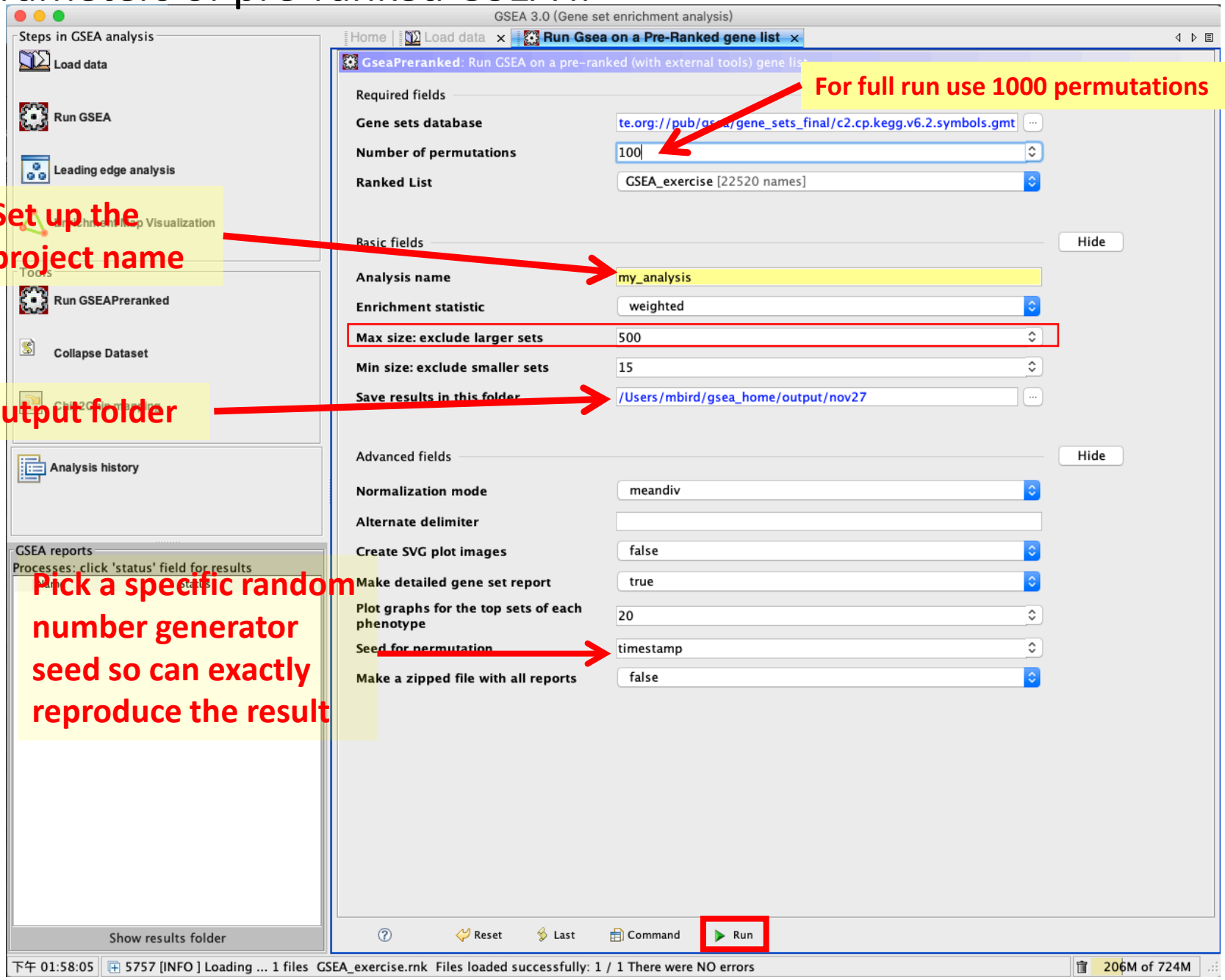
Gene sets database:
c2.cp.kegg.v6.2.symbols.gmt

Chip platform:
GENE_SYMBOL.chip





Parameters of pre-ranked GSEA II



Result→ index.html

GSEA Report for Dataset GSEA_exercise

Enrichment in phenotype: na

- 94 / 178 gene sets are upregulated in phenotype **na_pos**
- 11 gene sets are significant at FDR < 25%
- 9 gene sets are significantly enriched at nominal pvalue < 1%
- 15 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: na

- 84 / 178 gene sets are upregulated in phenotype **na_neg**
- 26 gene sets are significantly enriched at FDR < 25%
- 16 gene sets are significantly enriched at nominal pvalue < 1%
- 24 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Dataset details

- The dataset has 22520 features (genes)
- No probe set => gene symbol collapsing was requested, so all 22520 features were used

Gene set details

- Gene set size filters (min=15, max=500) resulted in filtering out 8 / 186 gene sets
- The remaining 178 gene sets were used in the analysis
- List of gene sets used and their sizes (restricted to features in the specified dataset)

Gene markers for the na_pos *versus* na_neg comparison

- The dataset has 22520 features (genes)
- Detailed rank ordered gene list for all features in the dataset

Global statistics and plots

- Plot of p-values *vs.* NES
- Global ES histogram

Other

- Parameters used for this analysis

Comments

- Timestamp used as the random seed: 1543298510119

Report: my_analysis.GseaPreranked.1543298501162.rpt by user: mbird
xtools.gsea.GseaPreranked [Tue, Nov 27, '18 2 PM 1]
Website: www.gsea-msigdb.org/gsea Questions & Suggestions: Contact page

Data of pre-ranked GSEA

Data summary

Table: GSEA Results Summary

Dataset	GSEA_exercise
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	KEGG_CELL_CYCLE
Enrichment Score (ES)	0.5274492
Normalized Enrichment Score (NES)	1.2457196
Nominal p-value	0.06557377
FDR q-value	0.6398135
FWER p-Value	1.0

Data of pre-ranked GSEA
snapshot of enrichment results

Table: GSEA Results Summary

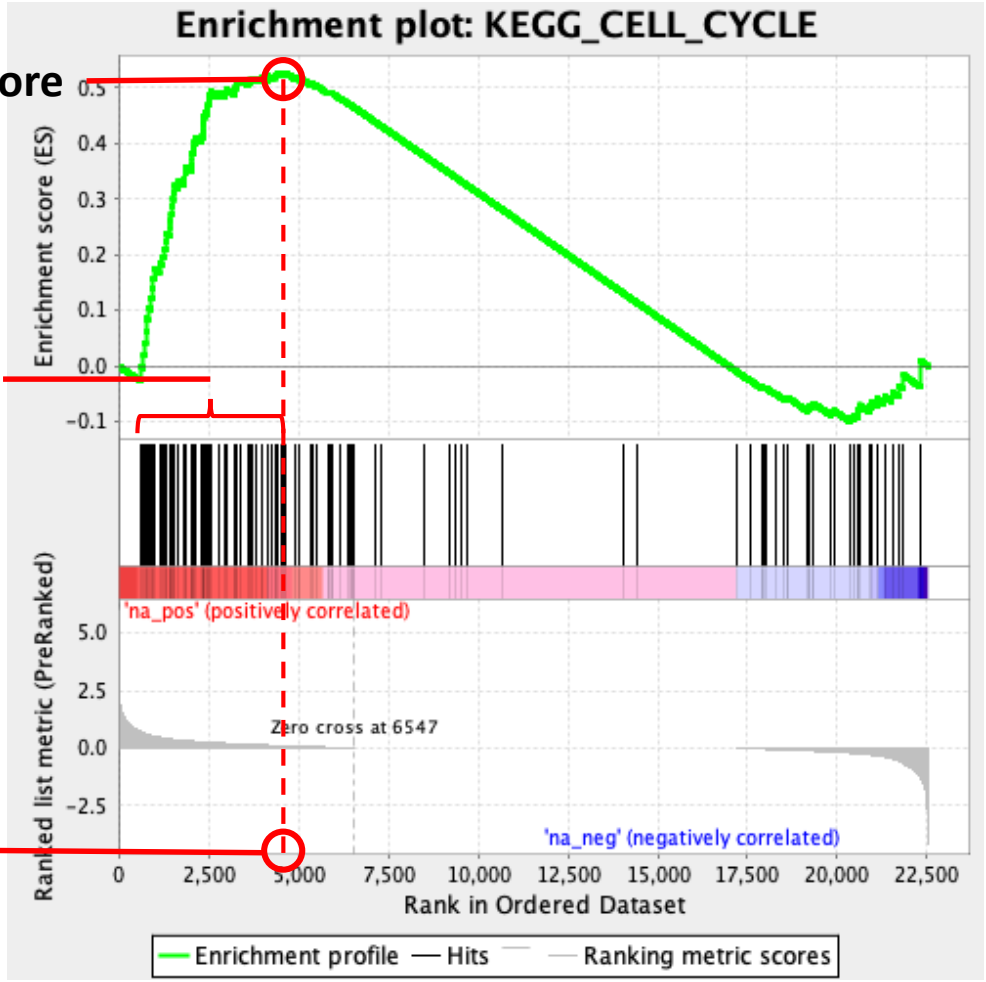
Dataset	GSEA_exercise
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	KEGG_CELL_CYCLE
Enrichment Score (ES)	0.5274492
Normalized Enrichment Score (NES)	1.2457196
Nominal p-value	0.06557377
FDR q-value	0.6398135
FWER p-Value	1.0

Snapshot of enrichment results

Enrichment score

Leading edge subset

Rank at max



Data of pre-ranked GSEA

Detailed enrichment results in html format

Table: Gene sets enriched in phenotype na [plain text format]

	GS follow link to MSigDB	GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	KEGG_TASTE_TRANSDUCTION	Details ...	49	0.82	1.79	0.000	0.008	0.010	2311	tags=10%, list=10%, signal=11%
2	KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS	Details ...	134	0.69	1.70	0.000	0.013	0.030	2760	tags=43%, list=12%, signal=49%
3	KEGG_OOCYTE_MEIOSIS	Details ...	110	0.65	1.65	0.000	0.030	0.100	3651	tags=44%, list=16%, signal=52%
4	KEGG_CALCIUM_SIGNALING_PATHWAY	Details ...	175	0.64	1.62	0.000	0.032	0.120	2560	tags=19%, list=11%, signal=22%
5	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	Details ...	268	0.59	1.57	0.000	0.071	0.310	1369	tags=6%, list=6%, signal=6%
6	KEGG_DNA_REPLICATION	Details ...	36	0.76	1.52	0.018	0.117	0.490	3859	tags=69%, list=17%, signal=84%
7	KEGG_HOMOLOGOUS_RECOMBINATION	Details ...	28	0.73	1.48	0.000	0.179	0.680	4297	tags=71%, list=19%, signal=88%
8	KEGG_TIGHT_JUNCTION	Details ...	131	0.60	1.48	0.017	0.166	0.730	4411	tags=36%, list=20%, signal=44%
9	KEGG_PROXIMAL_TUBULE_BICARBONATE_RECLAMATION	Details ...	22	0.81	1.48	0.000	0.149	0.740	866	tags=23%, list=4%, signal=24%
10	KEGG_MISMATCH_REPAIR	Details ...	23	0.74	1.46	0.018	0.162	0.800	4574	tags=70%, list=20%, signal=87%
11	KEGG_PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	Details ...	75	0.63	1.45	0.000	0.162	0.870	4173	tags=39%, list=19%, signal=47%
12	KEGG_CARDIAC_MUSCLE_CONTRACTION	Details ...	71	0.63	1.40	0.036	0.263	0.970	2378	tags=21%, list=11%, signal=24%
13	KEGG_DORSO_VENTRAL_AXIS_FORMATION	Details ...	24	0.74	1.37	0.053	0.311	0.990	2718	tags=29%, list=12%, signal=33%
14	KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	Details ...	210	0.51	1.33	0.030	0.435	1.000	3496	tags=33%, list=16%, signal=39%
15	KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION	Details ...	85	0.55	1.31	0.000	0.482	1.000	3702	tags=45%, list=16%, signal=53%
16	KEGG_OLFACTORY_TRANSDUCTION	Details ...	339	0.45	1.27	0.039	0.612	1.000	1971	tags=1%, list=9%, signal=1%
17	KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	Details ...	41	0.58	1.26	0.070	0.594	1.000	2652	tags=24%, list=12%, signal=28%
18	KEGG_CELL_CYCLE	Details ...	124	0.53	1.25	0.066	0.640	1.000	4643	tags=53%, list=21%, signal=67%

- Nominal P –value
  - Is the GS significantly changed between two classes
- NES
  - ES / Mean(All permutated ES)
  - A comparable ES across gene sets

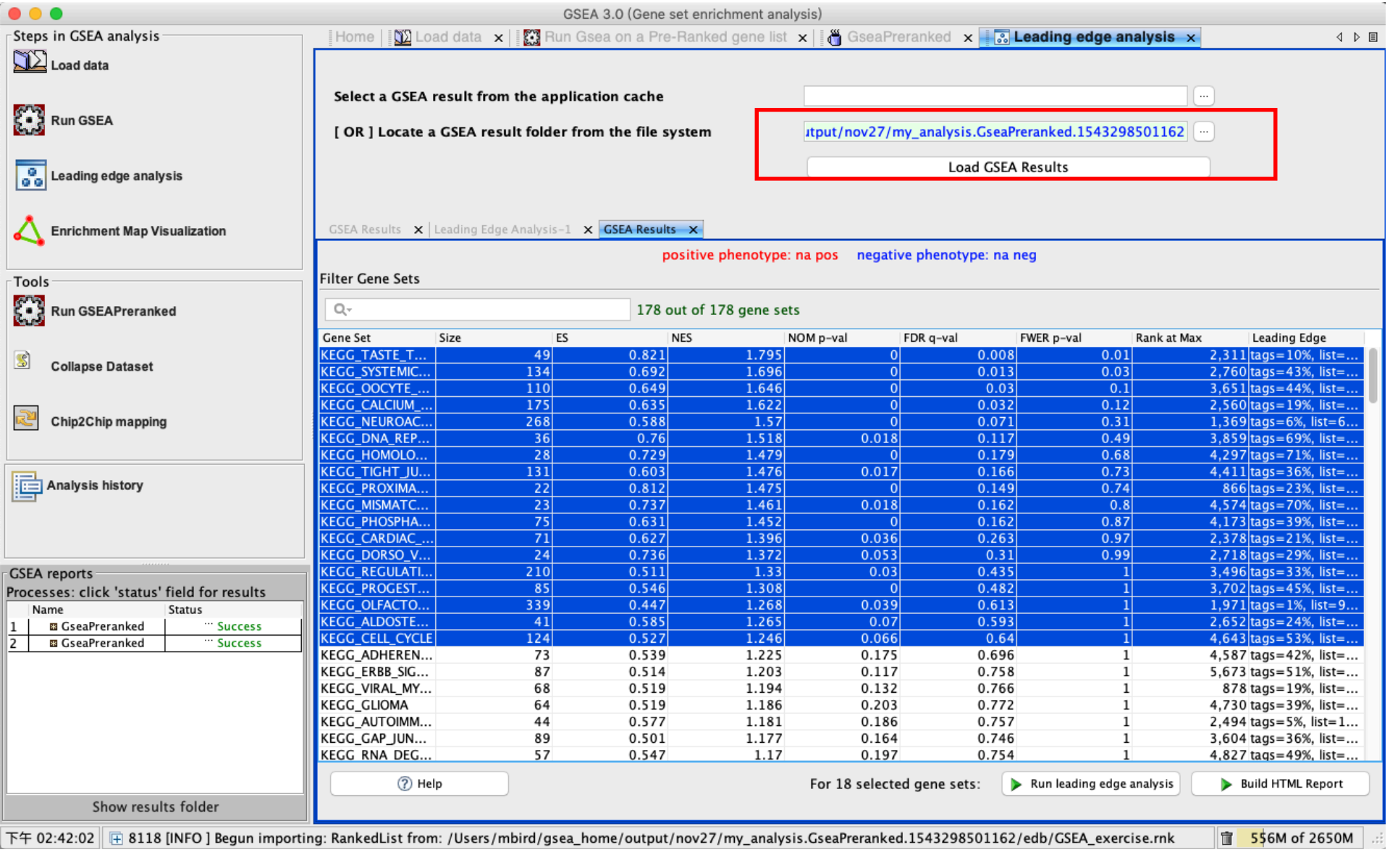
$$\mathrm{NES} = \frac{\text{actual ES}}{\textit{mean}(\text{ESs against all permutations of the dataset})}$$

- FDR or FWER
  - The estimated probability that a gene set with a given NES represents a false positive finding.

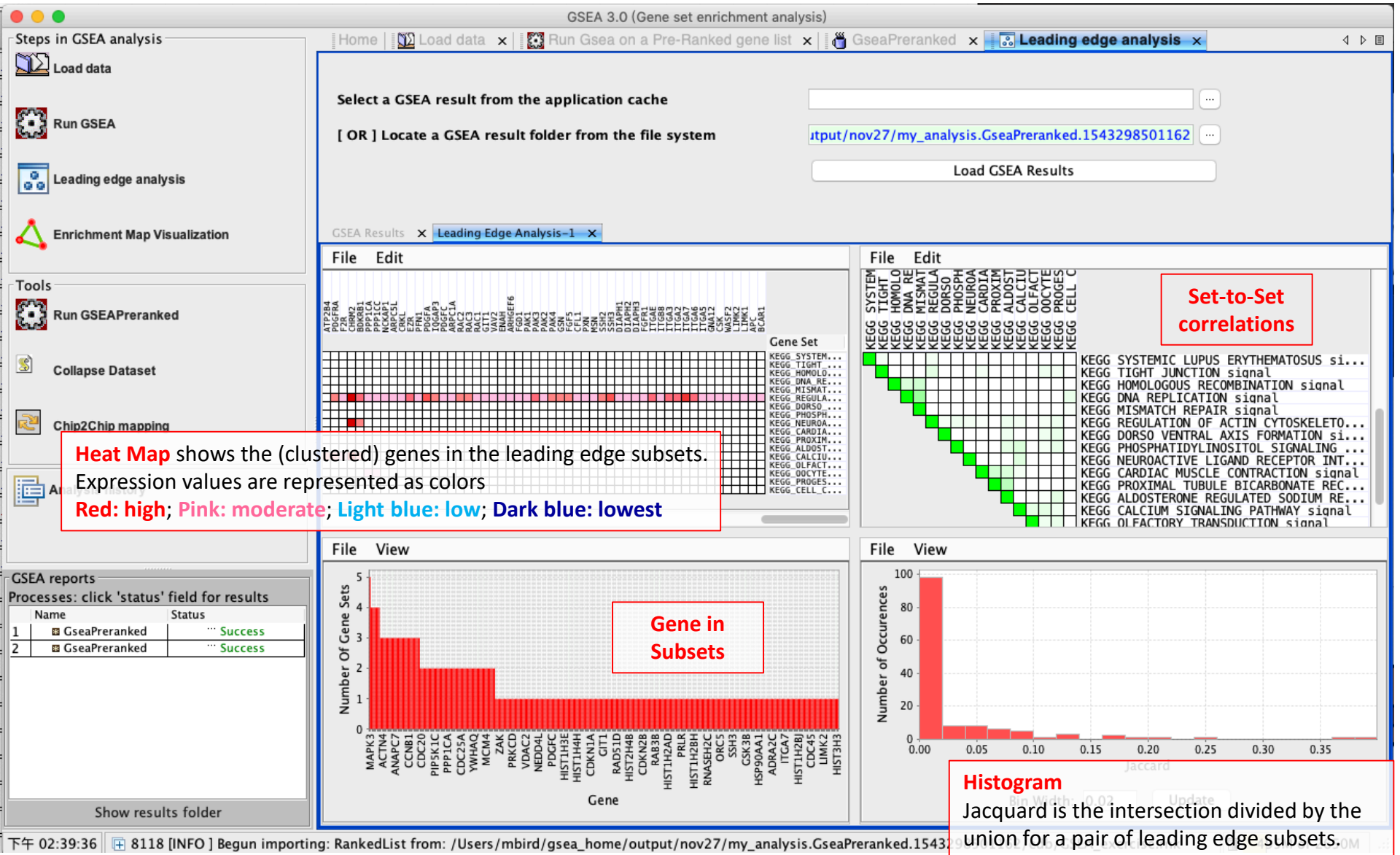
- Tags
  - % of gene hits before (for positive ES) or after (for negative ES) the peak in the running enrichment score.
  - % of genes contributing to the enrichment score.
- List
  - % of genes in the ranked gene list before (for positive ES) or after (for negative ES) the peak in the running enrichment score.
  - where in the list the enrichment score is attained.
- Signal
  - *N* is the number of genes in the list
  - *Nh* is the number of genes in the gene set
  - If the gene set is entirely within the first *Nh* positions in the list, then the signal strength is maximal or 100%. If the gene set is spread throughout the list, then the signal strength decreases towards 0%.

$$(\text{Tag }\%)(1-\text{Gene }\%)\left(\frac{N}{N-Nh}\right)$$

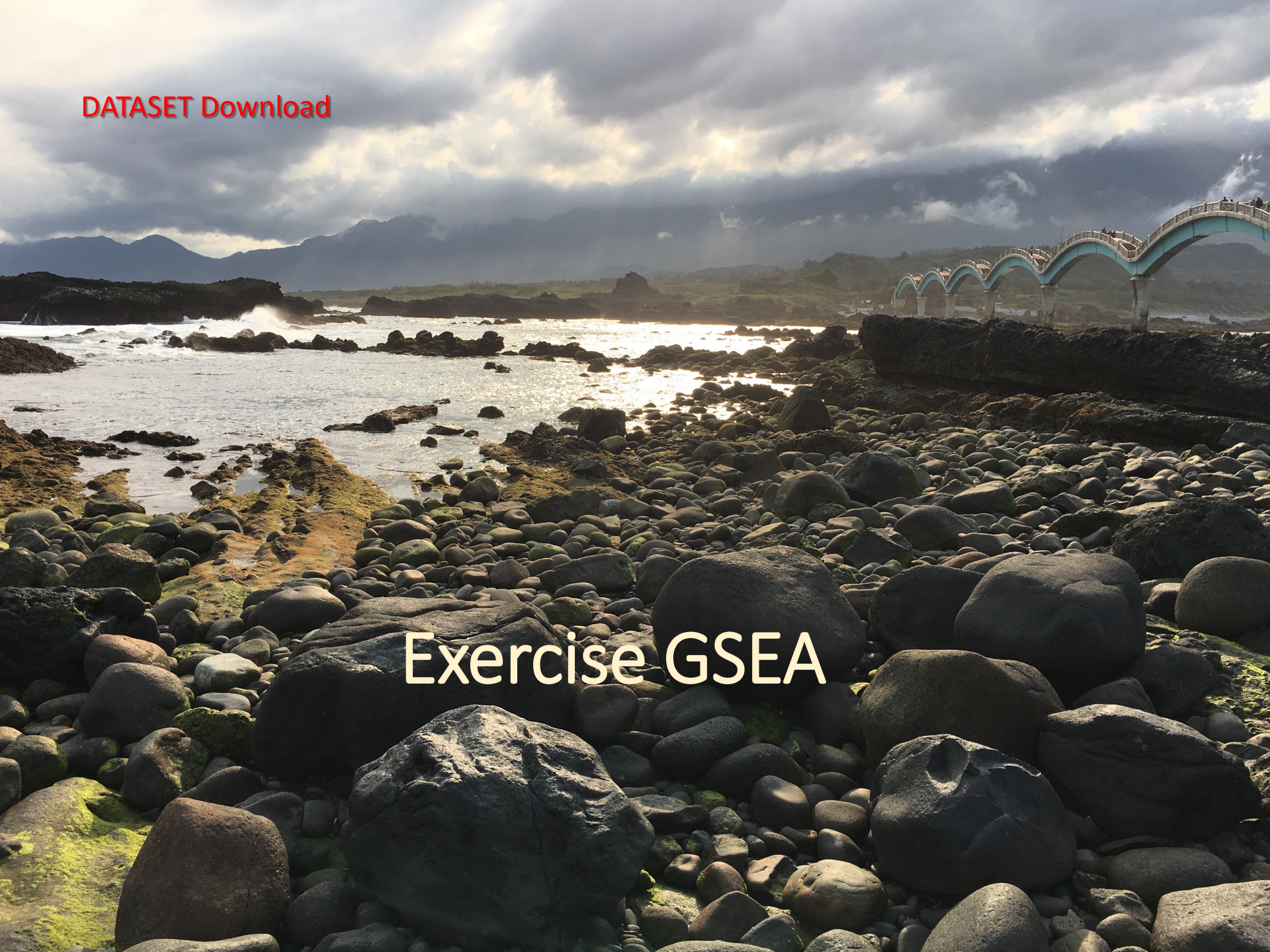
Leading Edge Analysis
Import GSEA results



Data of leading Edge Analysis



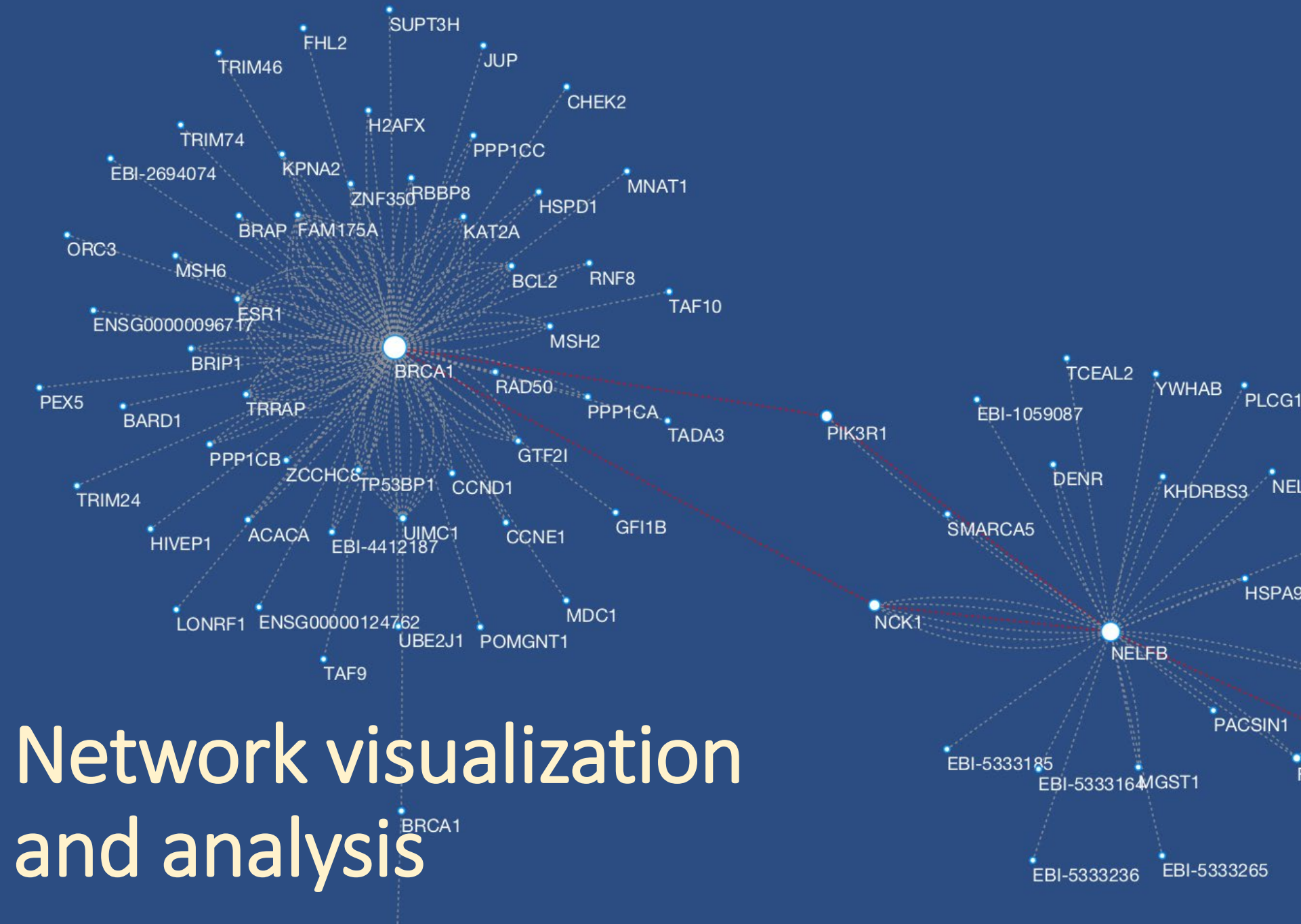
DATASET Download



Exercise GSEA

GSEA Exercise:

- 安裝與執行GSEA3.0
- 匯入Gene list檔案
- 執行Preranked tools (C2.KEGG)
- 輸出GSEA Report
- 比較 pos vs. neg enrichment table與產出各前五筆GSEA Enrichment plot
- Leading edge analysis 前五筆比較



Network visualization and analysis

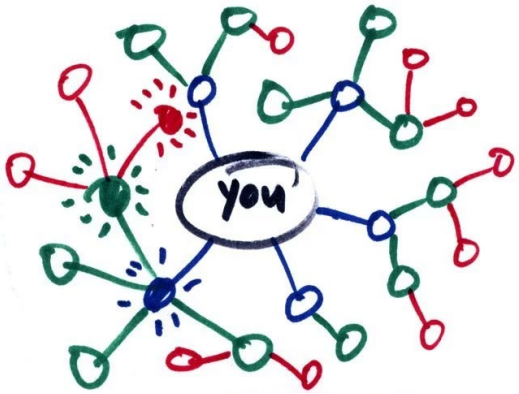
Networks

- Represent relationships
  - Physical, regulatory, genetic, functional interactions
- Useful for discovering relationships in large data sets
  - Better than tables in Excel
- Visualize multiple data types together
  - See interesting patterns
- Network analysis



to Gary Bader

Six Degrees of Separation



Picture thanks to bridgettribe.wordpress.com

- Everyone in the world is connected by at most six links
- Which path should we take?
- Shortest path by breadth first search
  - If two nodes are connected, will find the shortest path between them
- Are two proteins connected? If so, how?
- Biologically relevant?

to Gary Bader

Applications of Network Biology

- **Gene Function Prediction** –
  shows connections to sets of genes/proteins involved in same biological process
- **Detection of protein complexes/other modular structures** –
  discover modularity & higher order organization (motifs, feedback loops)
- **Network evolution** –
  biological process(es) conservation across species
- **Prediction of new interactions and functional associations** –
  Statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction
- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease.
- **Subnetwork-based diagnosis** – source of biomarkers for disease classification, identify interconnected genes whose aggregate expression levels are predictive of disease state
- **Subnetwork-based gene association** – map common pathway mechanisms affected by collection of genotypes

to Gary Bader

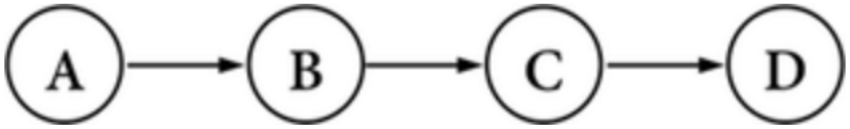
What's Missing?

- Dynamics
  - Pathways/networks represented as **static** processes
    - Difficult to represent a calcium wave or a feedback loop
  - More detailed mathematical representations exist that handle these e.g. Stoichiometric modeling, Kinetic modeling (VirtualCell, E-cell, …)
    - Need to accumulate or estimate comprehensive kinetic information
- Detail – atomic structures
- Context – cell type, developmental stage

to Gary Bader

What Have We Learned?

- Networks are useful for seeing relationships in large data sets
- Important to understand what the **nodes and edges** mean
- Important to define the biological question - know what you want to do with your gene list or network
- Many methods available for gene list and network analysis
  - Good to determine your question and search for a solution
  - Or get to know many methods and see how they can be applied to your data



to Gary Bader

Network Visualization and Analysis using Cytoscape

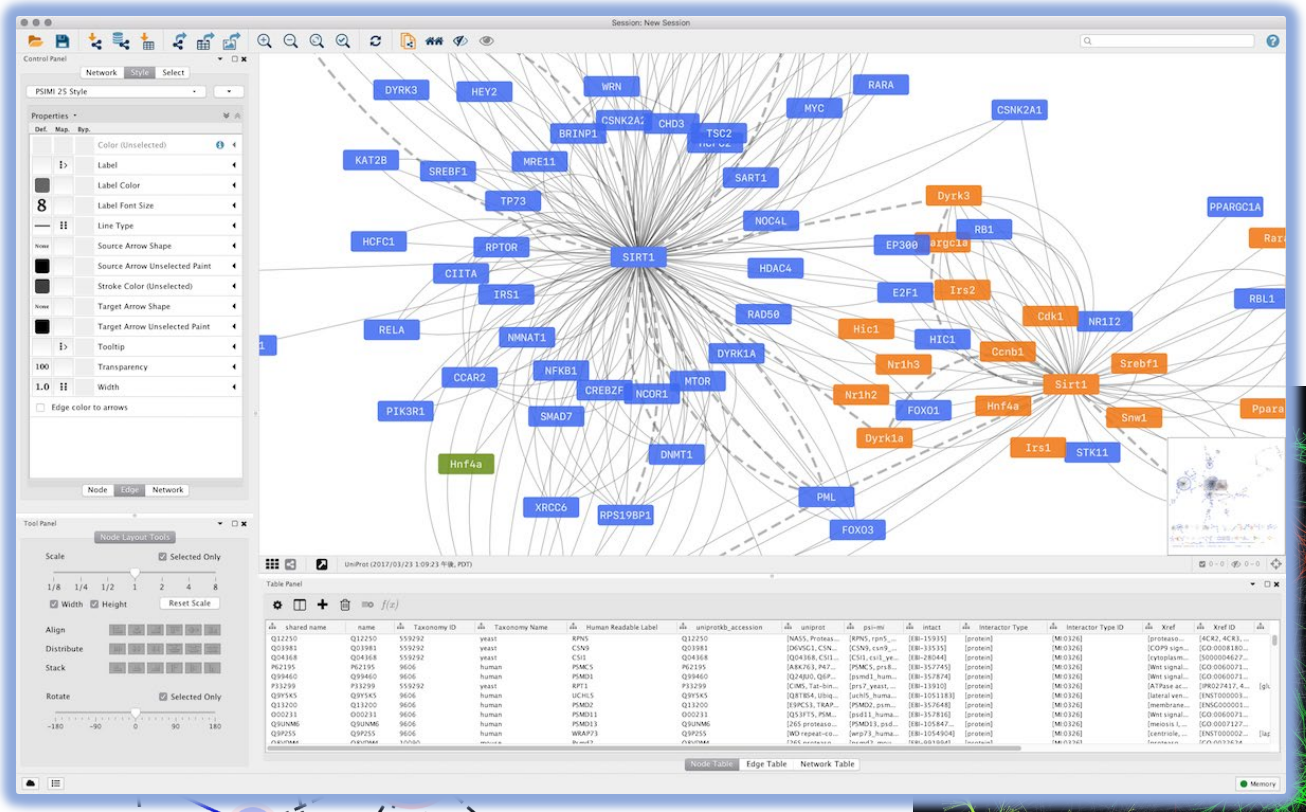
- Network visualization and analysis using Cytoscape software
- Cytoscape basics
- Cytoscape network analysis examples

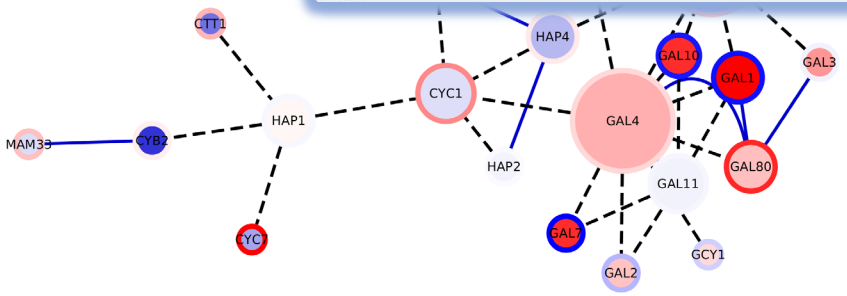
to Gary Bader

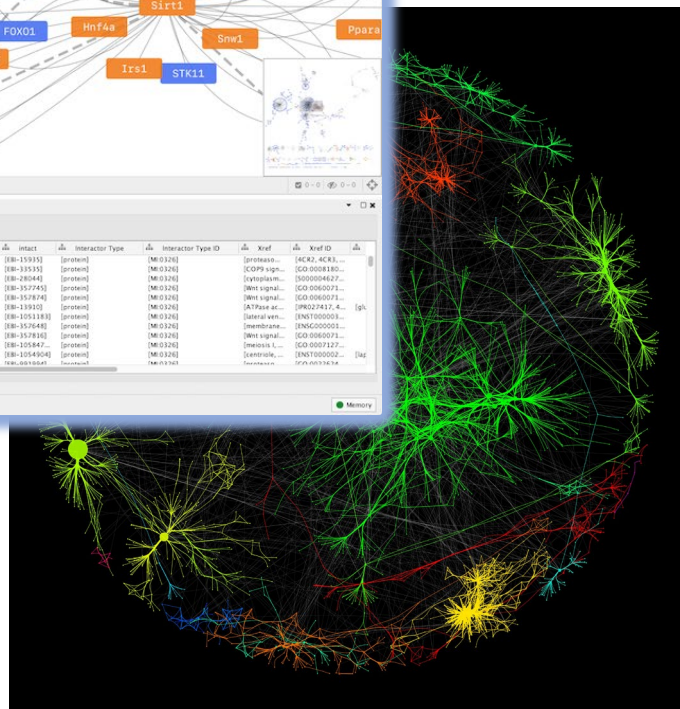
Java version: 1.8.0_162

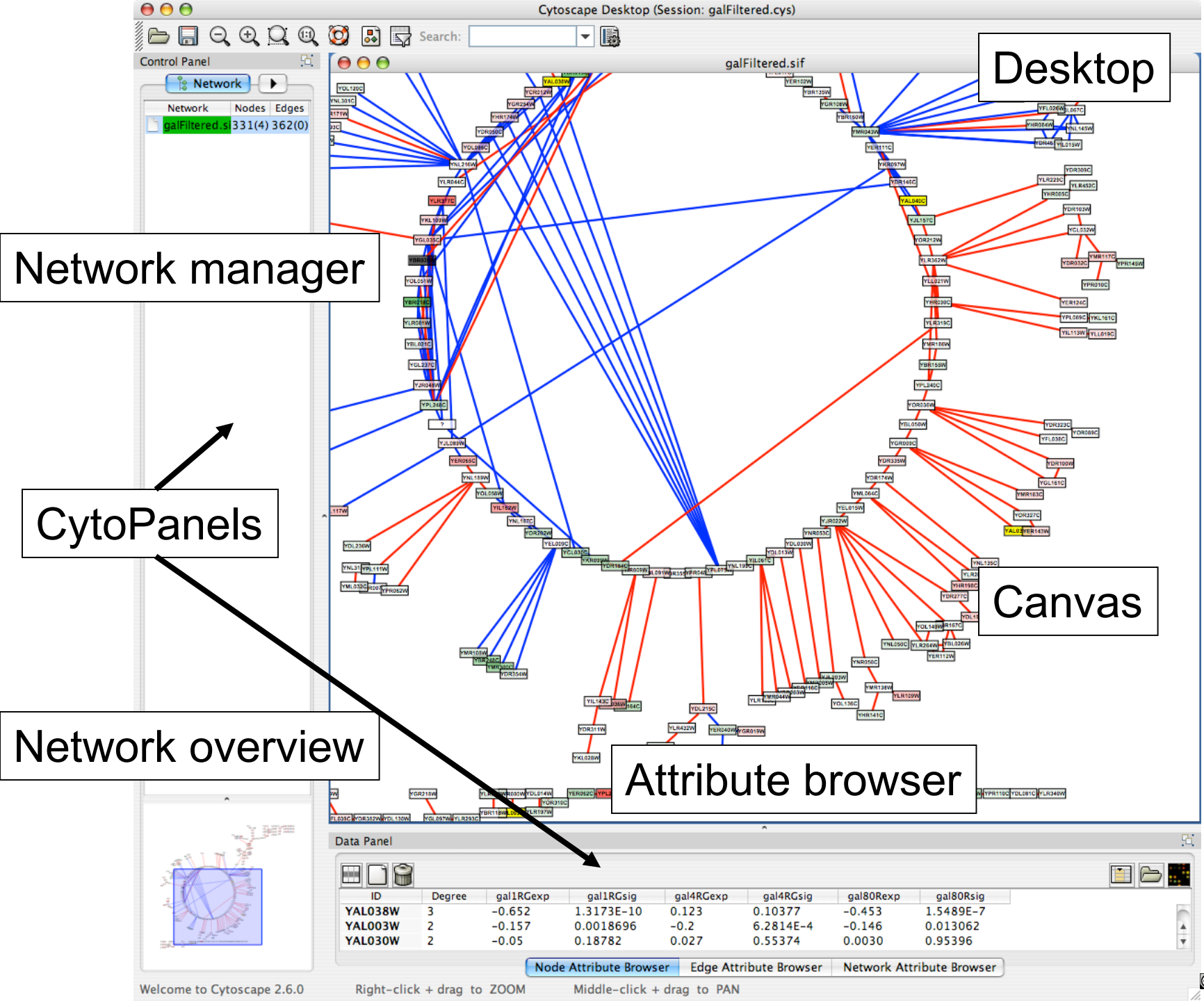
Cytoscape

Network Data Integration, Analysis, and Visualization in a Box

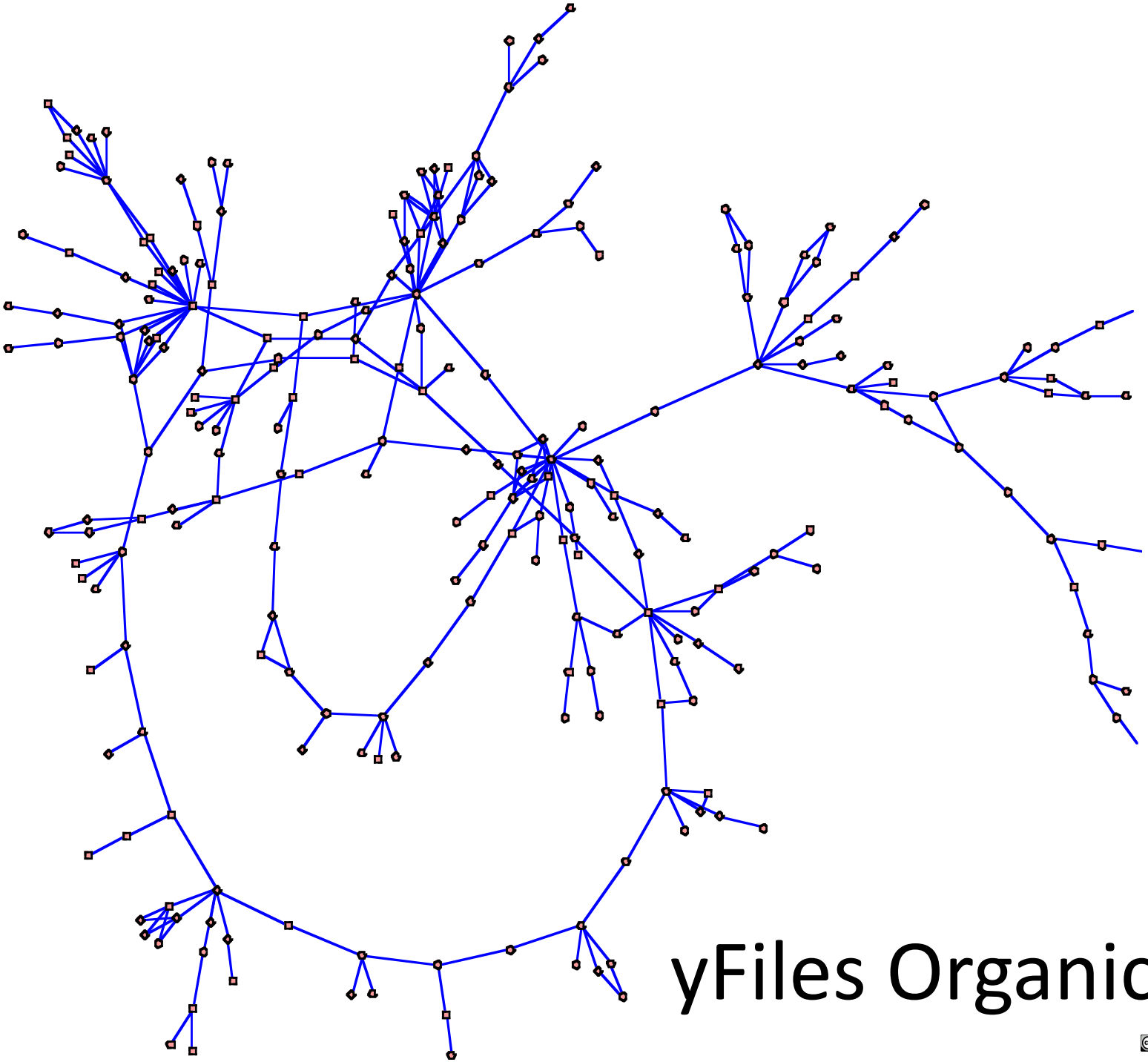




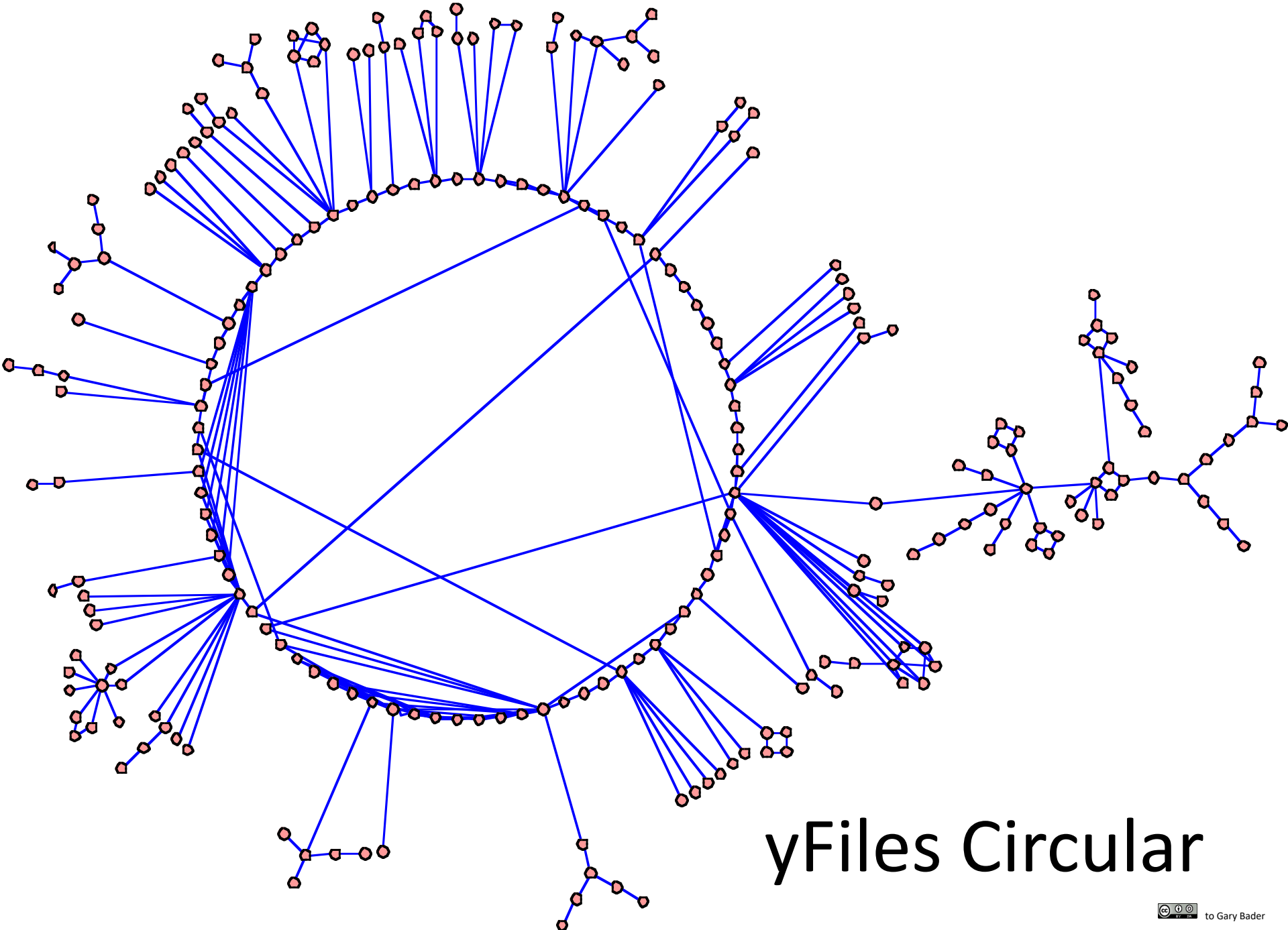




to Gary Bader



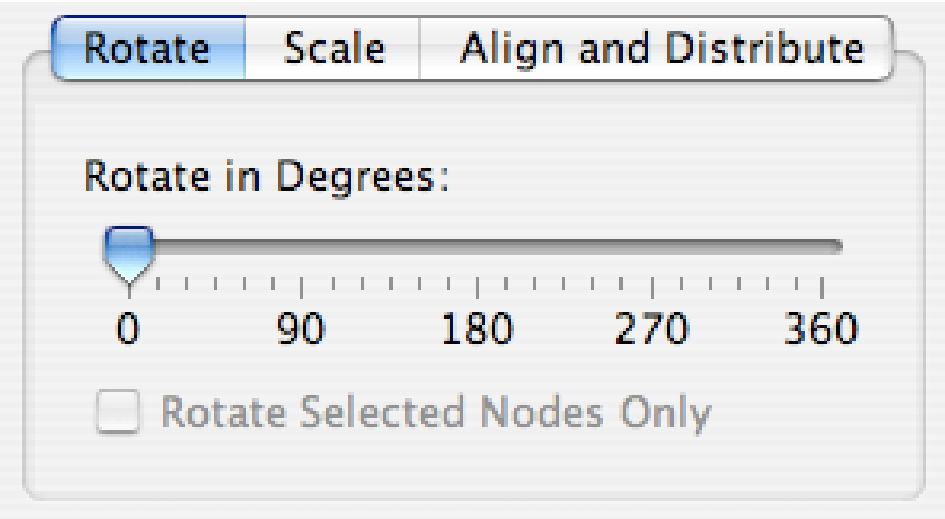
to Gary Bader

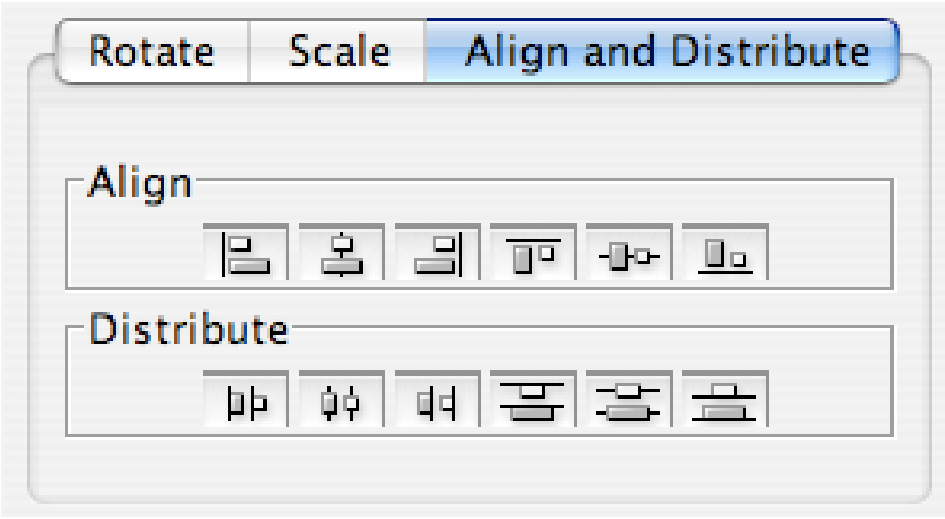


to Gary Bader

Network Layout

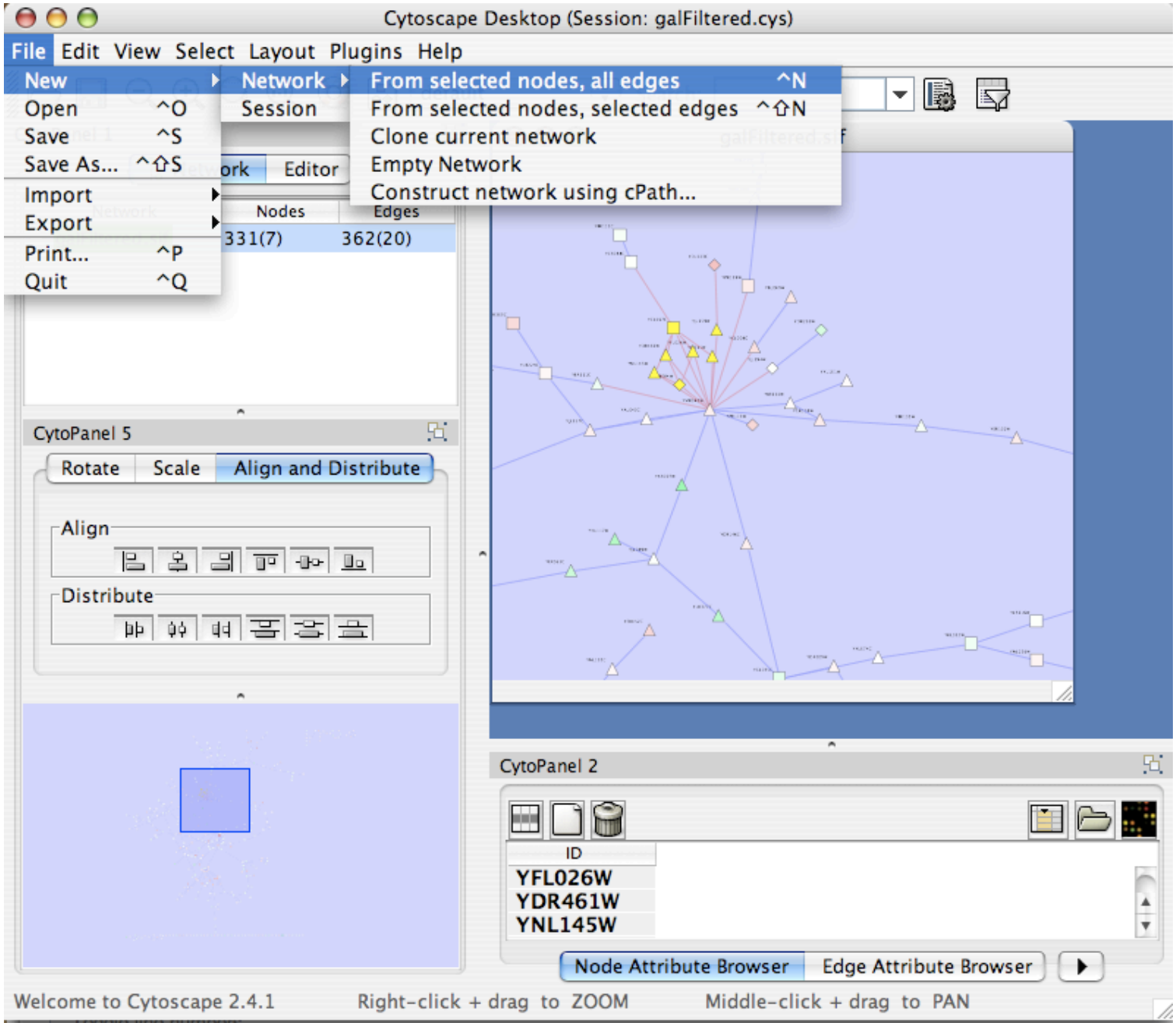
- Many algorithms available through apps
- Demo: Move, zoom/pan, rotate, scale, align





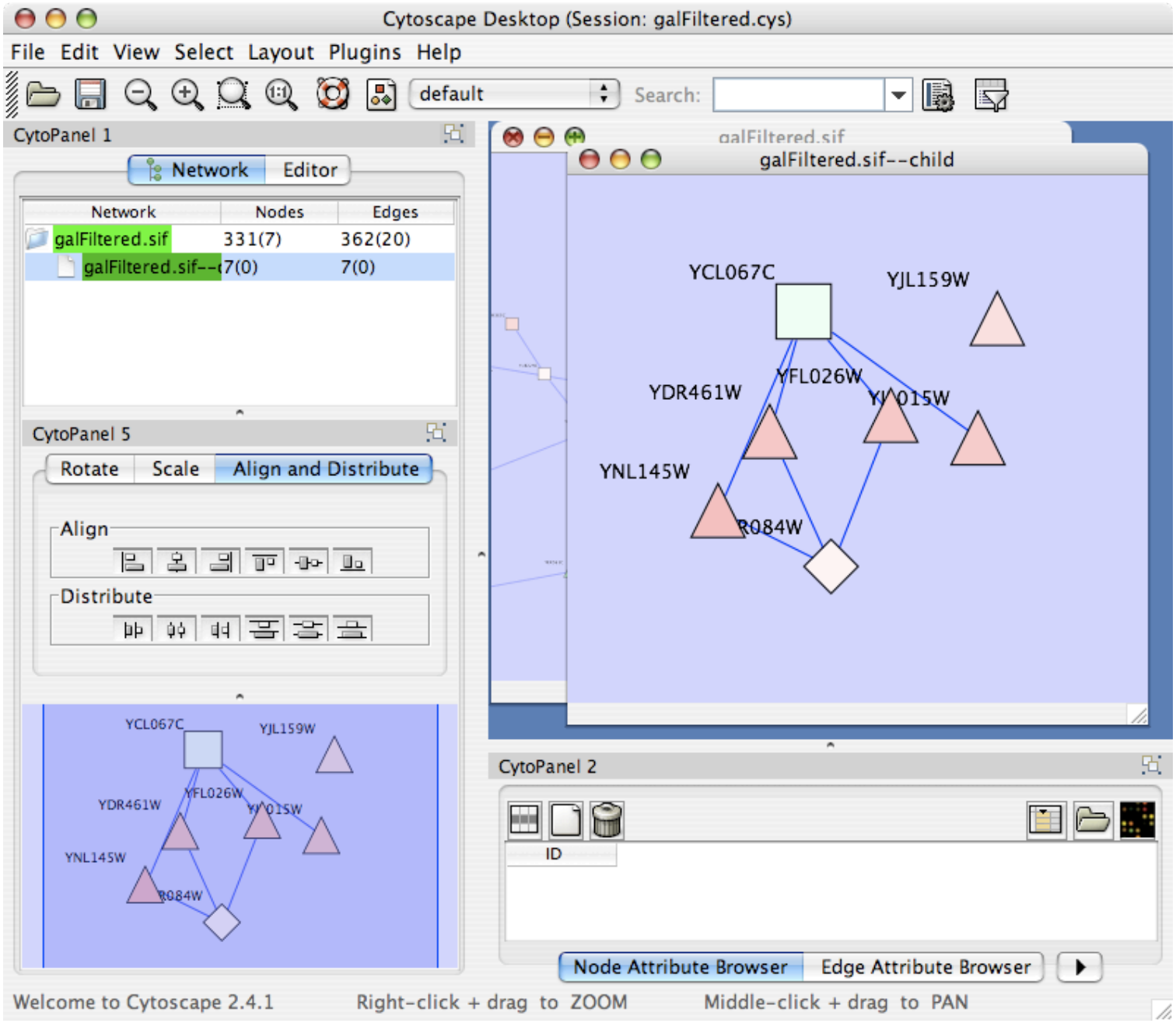
to Gary Bader

Create Subnetwork



to Gary Bader

Create Subnetwork



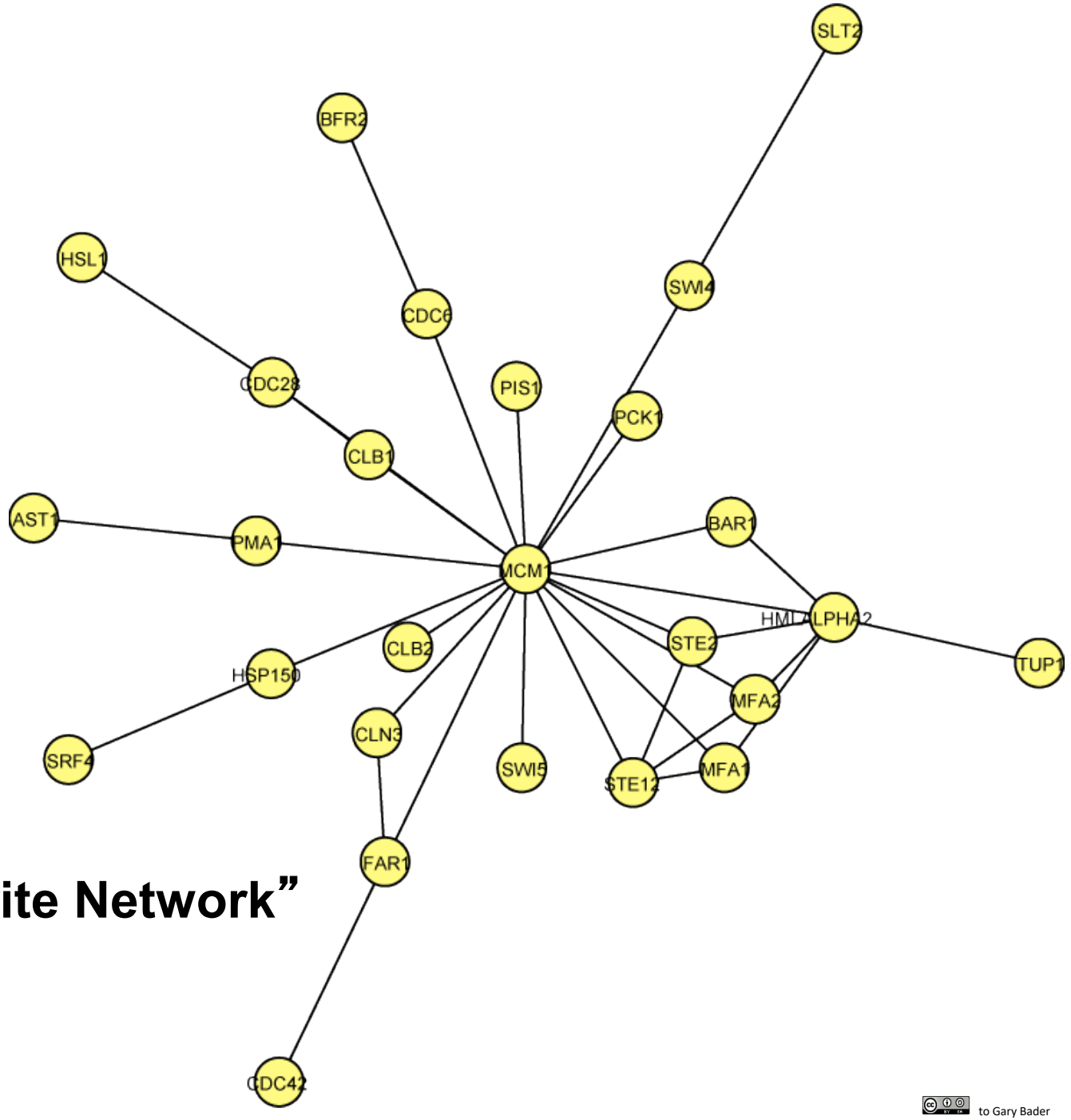
to Gary Bader

Visual Style

- Customized views of experimental data in a network context
- Network has **node** and **edge** attributes
    - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
    - E.g. node/edge size, shape, colour…
- E.g. Visualize gene expression data as node colour gradient on the network

to Gary Bader

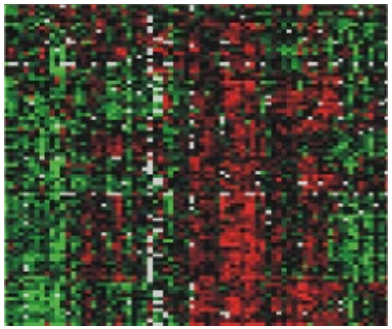
Visual Style

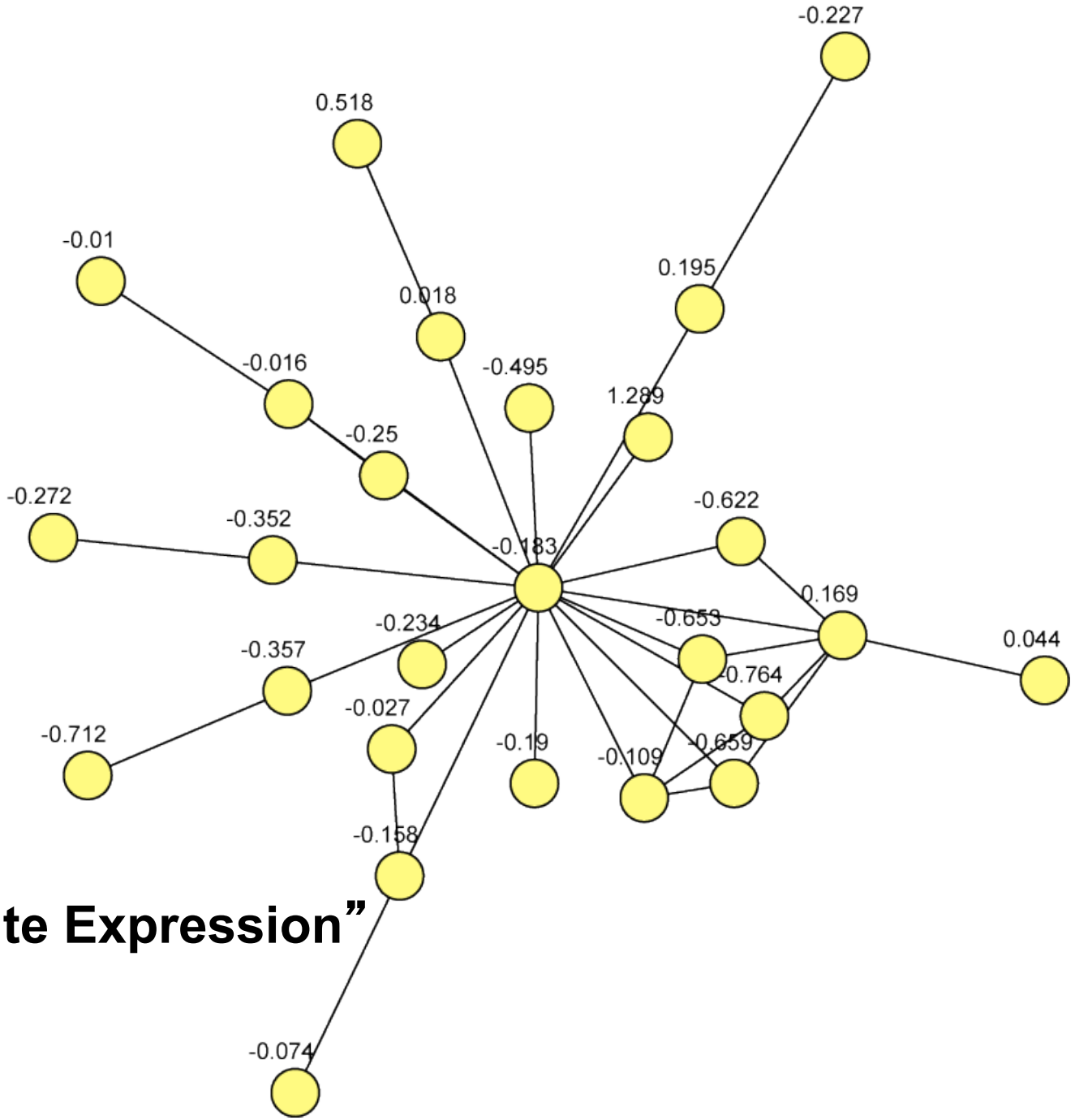


Load “Your Favorite Network”

to Gary Bader

Visual Style

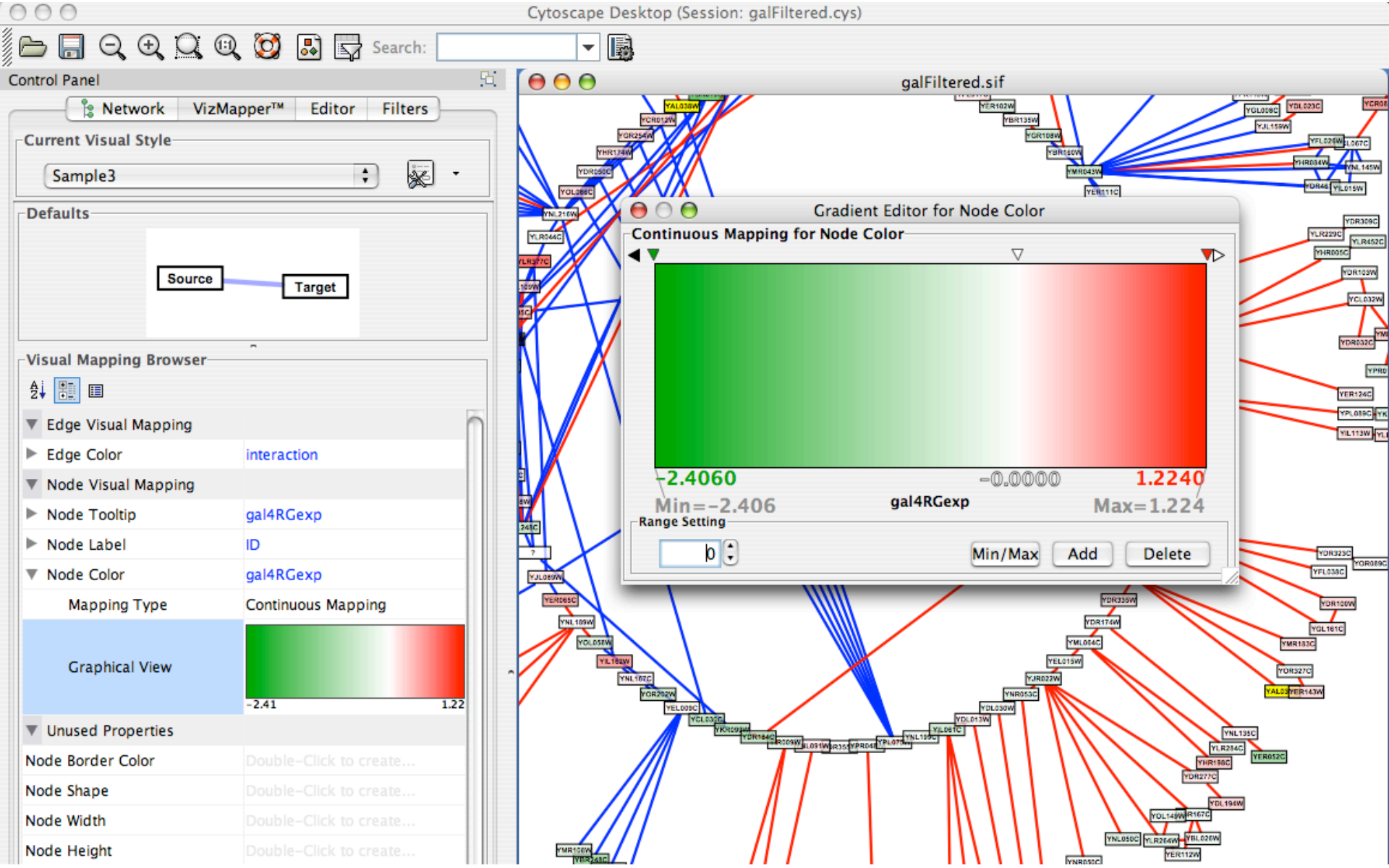




Load “Your Favorite Expression” Dataset

to Gary Bader

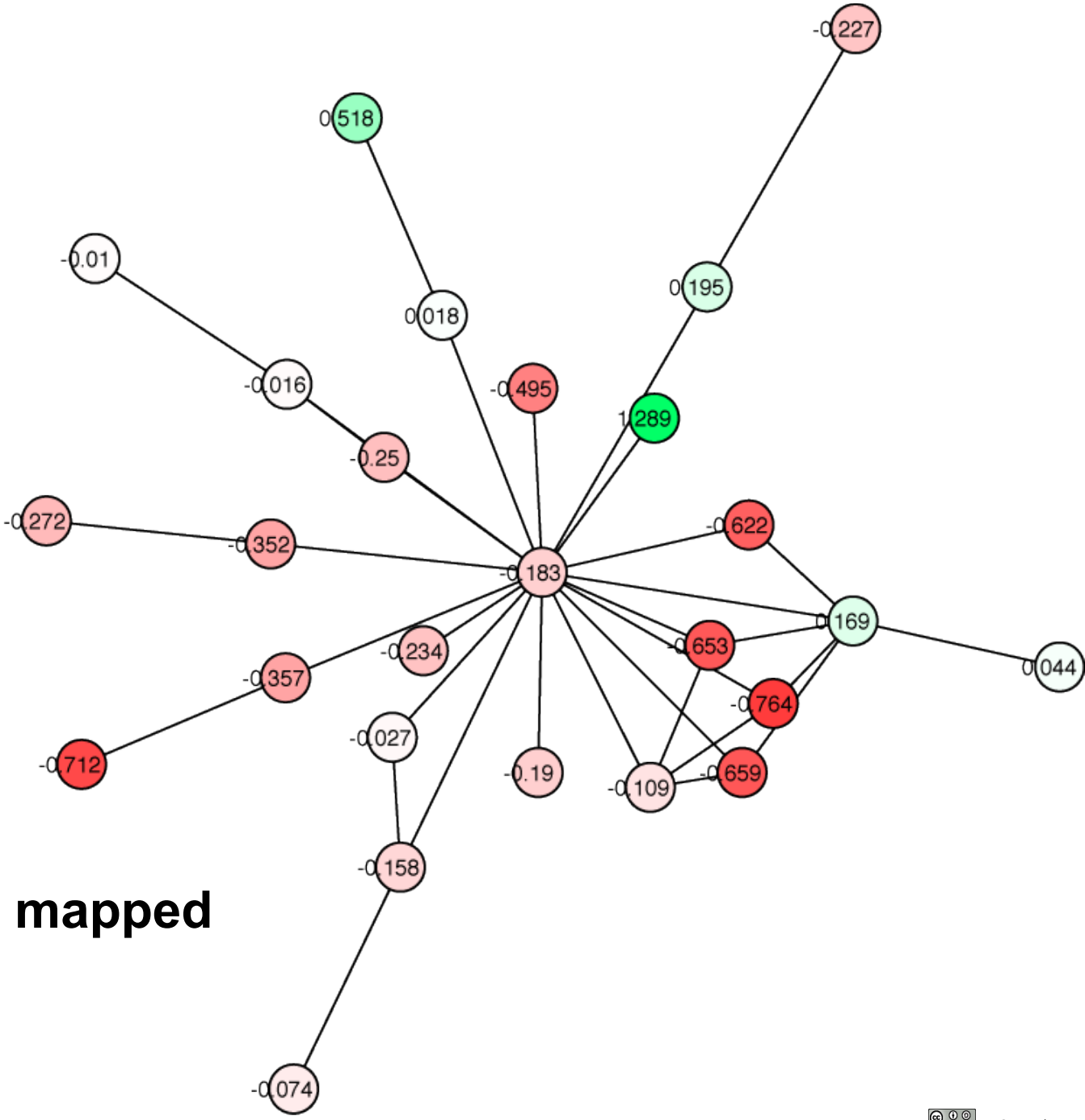
Visual Style



Map expression values to node colours using a continuous mapper

to Gary Bader

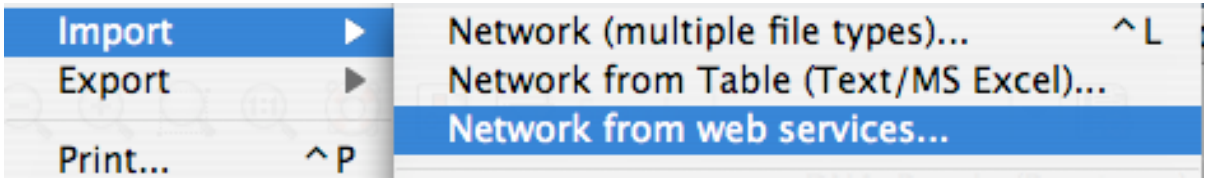
Visual Style

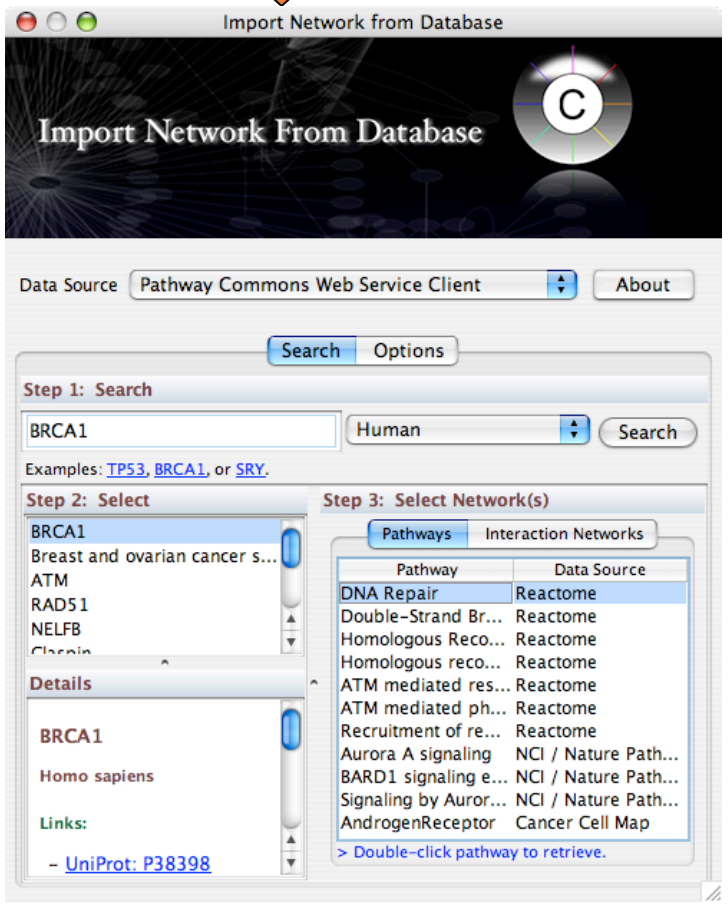


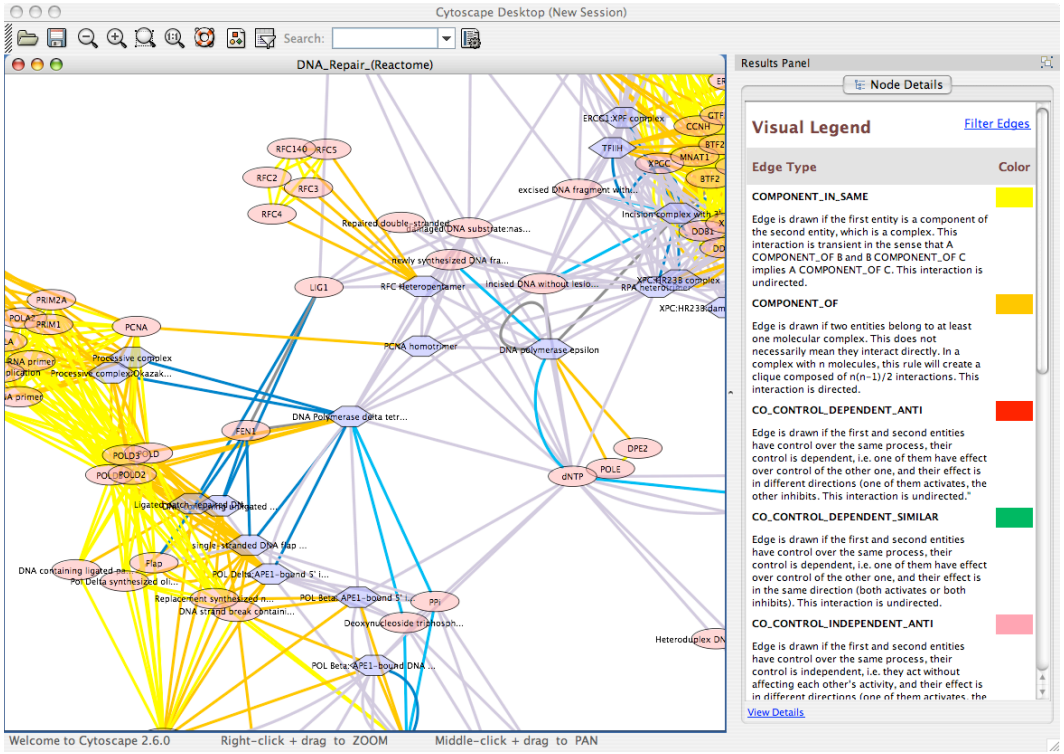
Expression data mapped to node colours

to Gary Bader

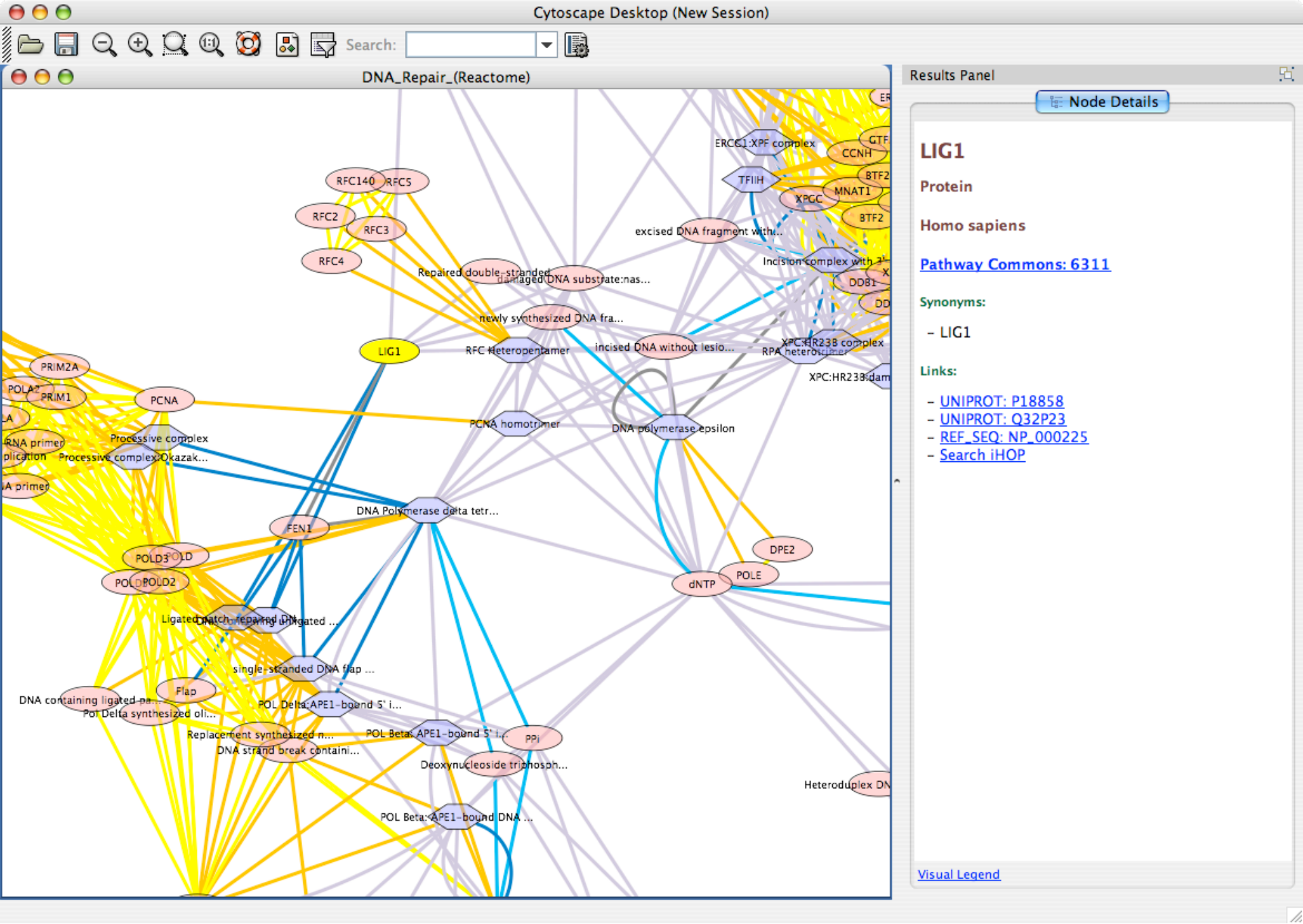
Interaction Database Search



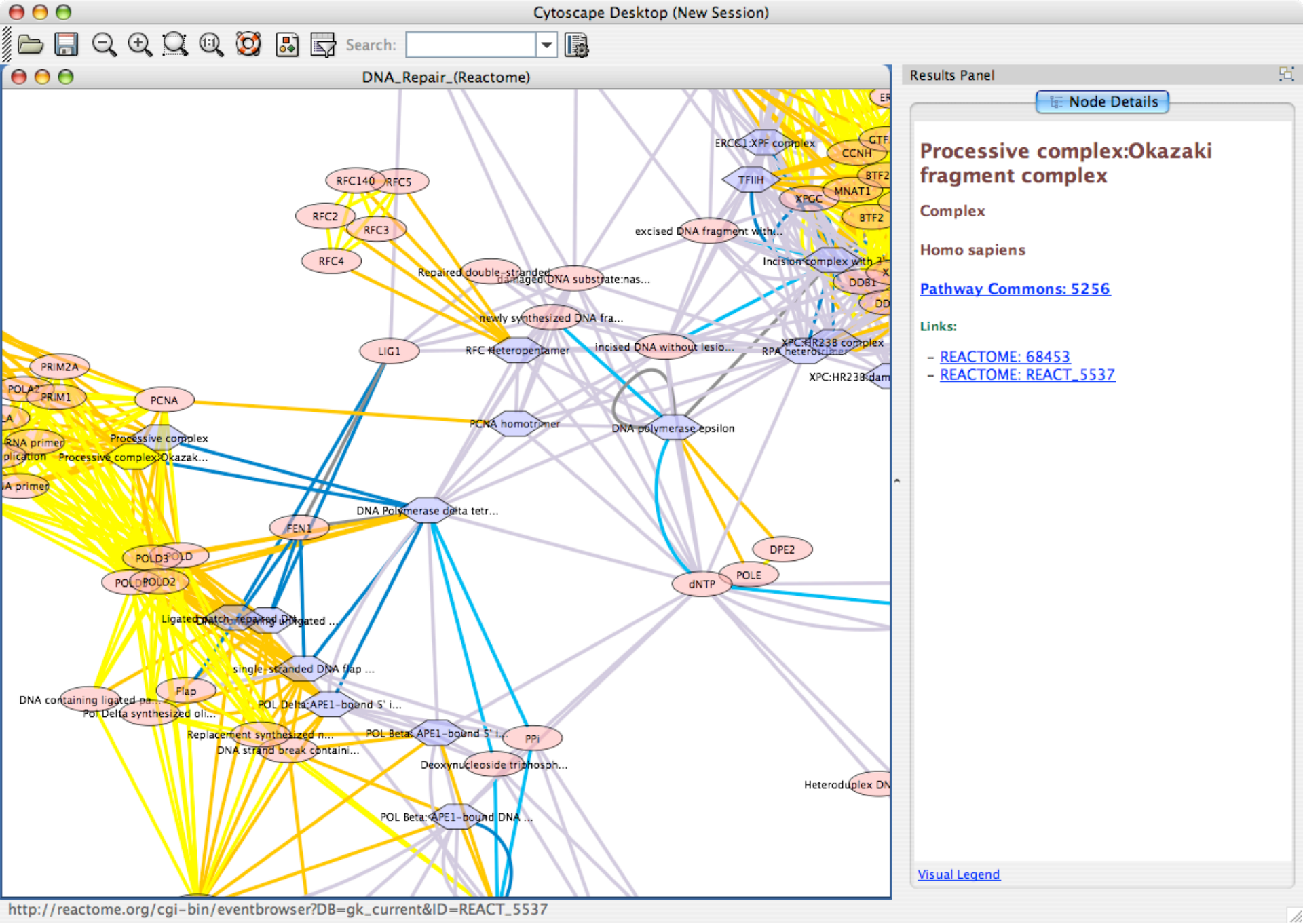




to Gary Bader



to Gary Bader



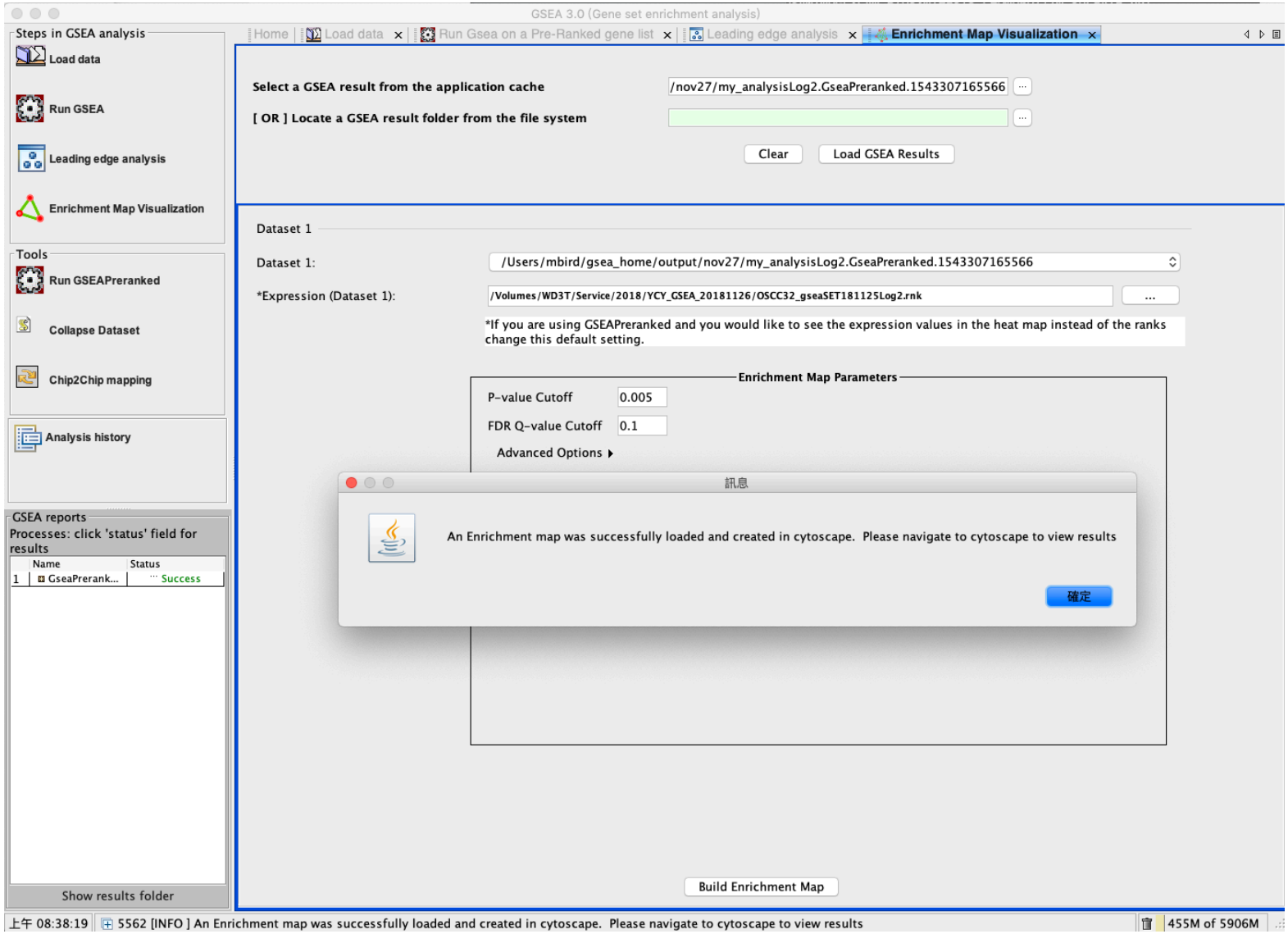
to Gary Bader

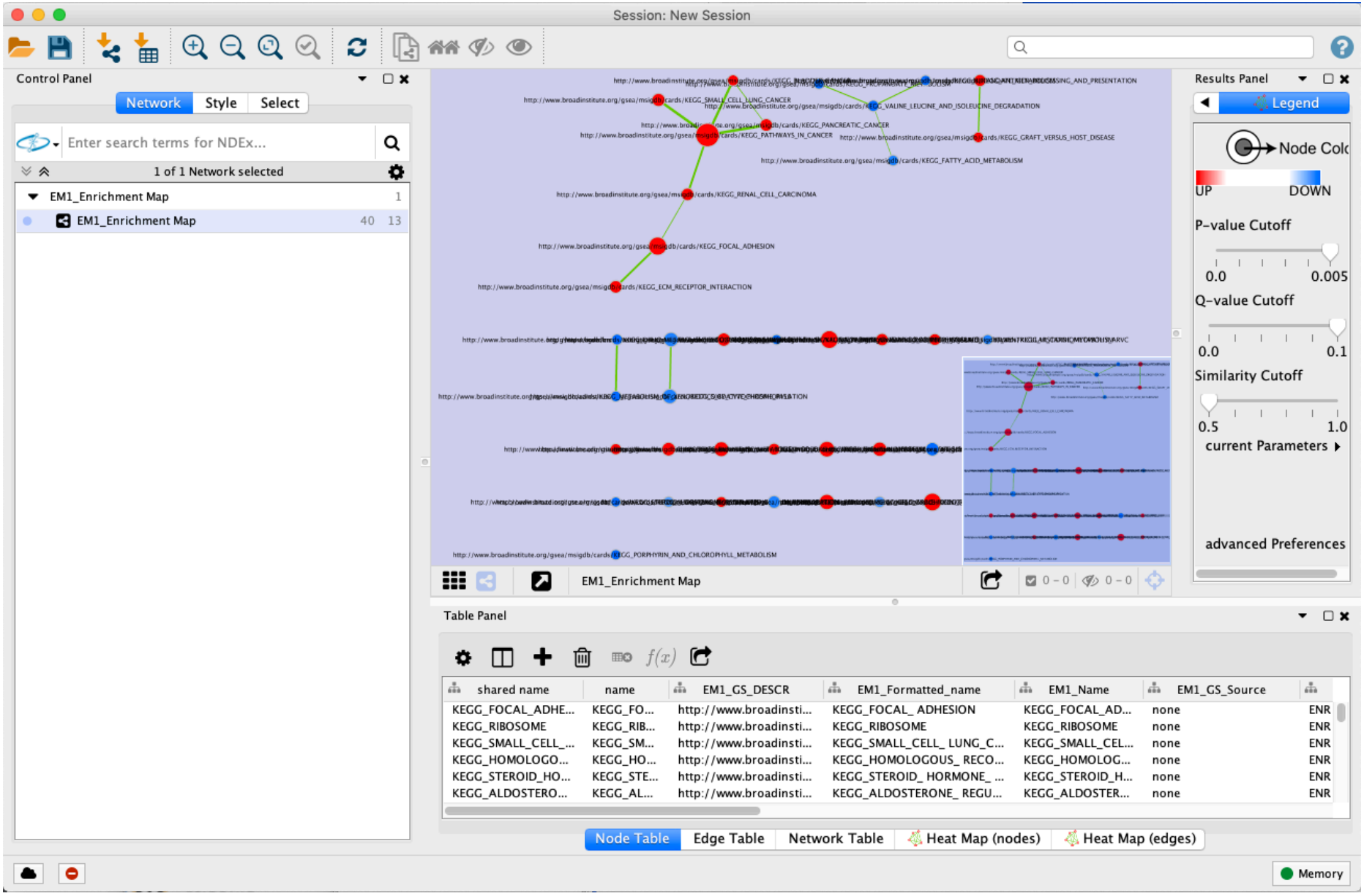

Exercises

Cytoscape Exercise:

- 安裝與執行cytoscape 3.x
- 匯入GSEA Build Enrichment Map檔案
- 練習修改Layout, Style
- 安裝App

GSEA <-> Cytoscape




Session: New Session
Control Panel
Network
Style
Select
Enter search terms for NDEx...
1 of 1 Network selected
EM1_Enrichment Map 1
EM1_Enrichment Map 40 13
Results Panel
Legend
Node Color
UP
DOWN
P-value Cutoff
0.0
0.005
Q-value Cutoff
0.0
0.1
Similarity Cutoff
0.5
1.0
current Parameters
advanced Preferences
EM1_Enrichment Map
Table Panel
shared name
name
EM1_GS_DESCR
EM1_Formatted_name
EM1_Name
EM1_GS_Source
KEGG_FOCAL_ADHE...
KEGG_FO...
http://www.broadinsti...
KEGG_FOCAL_ ADHESION
KEGG_FOCAL_AD...
none
ENR
KEGG_RIBOSOME
KEGG_RIB...
http://www.broadinsti...
KEGG_RIBOSOME
KEGG_RIBOSOME
none
ENR
KEGG_SMALL_CELL_...
KEGG_SM...
http://www.broadinsti...
KEGG_SMALL_CELL_ LUNG_C...
KEGG_SMALL_CEL...
none
ENR
KEGG_HOMOLOGO...
KEGG_HO...
http://www.broadinsti...
KEGG_HOMOLOGOUS_ RECO...
KEGG_HOMOLOG...
none
ENR
KEGG_STEROID_HO...
KEGG_STE...
http://www.broadinsti...
KEGG_STEROID_ HORMONE_ ...
KEGG_STEROID_H...
none
ENR
KEGG_ALDOSTERO...
KEGG_AL...
http://www.broadinsti...
KEGG_ALDOSTERONE_ REGU...
KEGG_ALDOSTER...
none
ENR
Node Table
Edge Table
Network Table
Heat Map (nodes)
Heat Map (edges)
Memory