Regulatory Genomics III: Pathway Analysis

Shamith Samarajiwa

Group Leader (Computational Biology & Data Science)
MRC Cancer Unit,
University of Cambridge.

Computational Biology MPhil: Genomics II Lectures
February 26th 2021

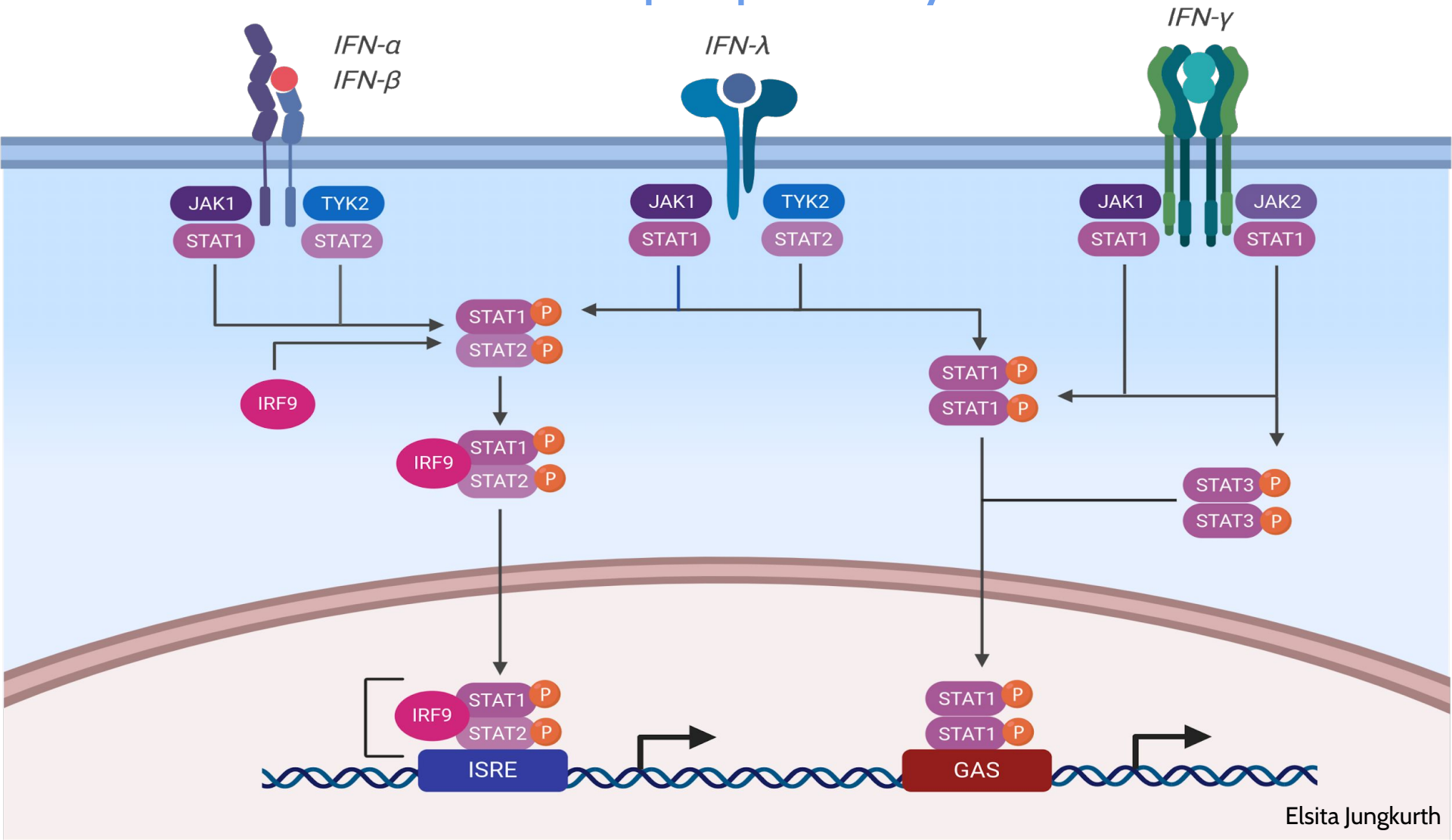
Overview

- What is a biological pathway?
- Statistical methods for pathway enrichment analysis
- Pitfalls of enrichment analysis
- Gene name disambiguation
- Gene Ontology Enrichment
- Pathway Enrichment
- Gene Set Enrichment Analysis (GSEA)
- Interactome and Network Analysis

What are biological pathways?

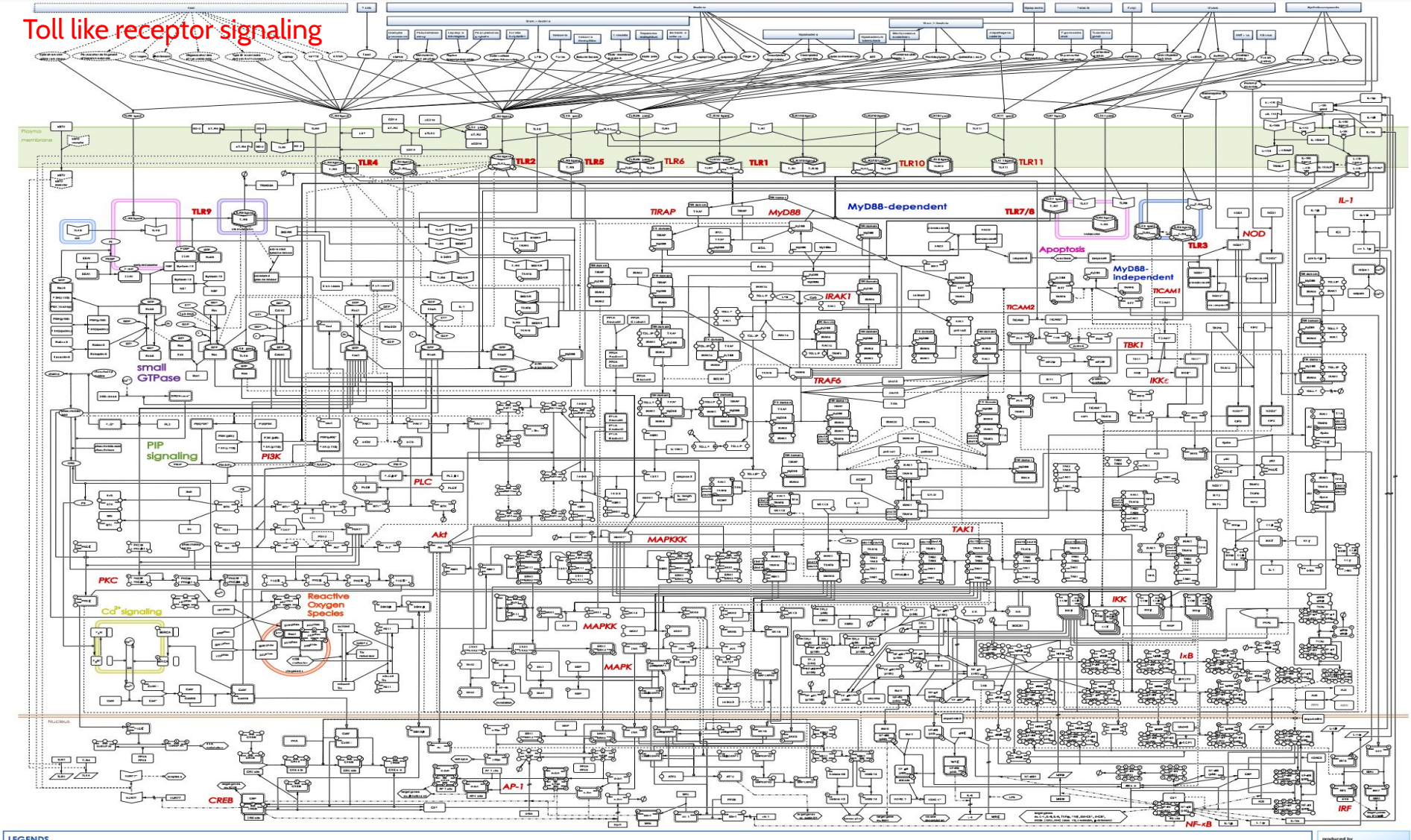
- Biological Pathways are a series of consecutive interactions between biochemical molecules acting in concert, to maintain and control of cellular information flow, energy and biochemical compounds in the cell leading to a change in molecular products or a cellular states.
- What's known as "pathways" are usually fragments or subnetworks of more complex networks.
- Information flow in pathways demonstrate directionality, feedforward and feedback loops, negative and positive regulation.
- There are different types of pathways:
    - Signalling
    - Genetic, Transcriptional, Regulatory
    - Metabolic (anabolic, catabolic, transport and energy)

A simple pathway

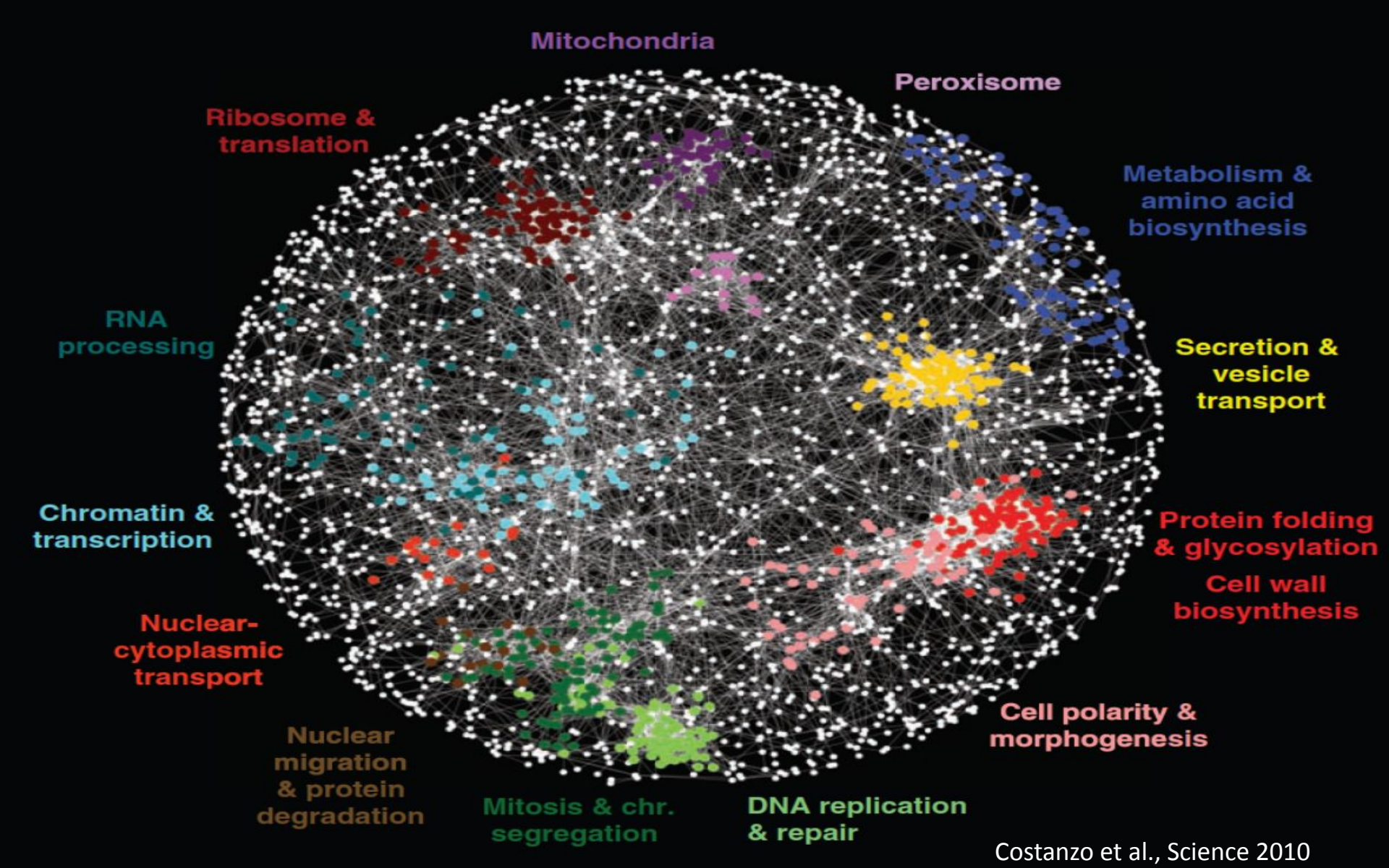


Elsita Jungkurth

Toll like receptor signaling



Oda and Kitano, Mol. Sys. Biol., 2006



Costanzo et al., Science 2010

Types of enrichment analysis

- Pathway Analysis - measure perturbation, disruption or change in a pathway activity by quantification of the associated enrichment probability.
- Other types of enrichment analysis:
  - Ontologies
  - Gene Sets
  - Networks
- Over Representation Analysis
- Functional Class Scoring

Limitations:

- Each functional category is analysed independently, without considering the system
- Fail to consider connectivity and topology of pathways

Where do gene lists come from?

- **Omics data**
- Comparison of Gene Lists
  - Are any genes/gene products significantly enriched or depleted in my list?
  - Differential enrichment between two lists

  ex: Fisher's exact test
- Ranked list of genes
  - Are any gene sets ranked significantly high or low in my ranked list of genes?

  ex: minHG test
- How do we assess significance?
  - Compare these lists against known sets of genes that act in concert to drive biological processes.
  - Compare to a background set/universe or difference between conditions.
  - Correct for multiple testing (FDR or FWER)

Statistical tests for enrichment

Enrichment of gene lists:

- Fisher's exact test & Hypergeometric test
- Chi-squared test
- Equality of two probabilities test
- Binomial test

Enrichment or depletion of a GO category within a class of genes: which test?

Isabelle Rivals ✉, Léon Personnaz, Lieng Taing, Marie-Claude Potier

Bioinformatics, Volume 23, Issue 4, 15 February 2007, Pages 401–407,

Enrichment of ranked lists:

- Kolmogorov-Smirnov (implemented in GSEA)
- Minimum hypergeometric test
- Wilcoxon Rank Sum test & Mann-Whitney U test

Fisher's Exact test

The formula below gives the exact hypergeometric probability of observing this particular arrangement of the data, assuming the given marginal totals, on the null hypothesis that men and women are equally likely to be studiers.

Fisher's exact test is a statistical significance test used in the analysis of contingency tables (significance of the deviation from a null hypothesis (e.g., P-value) can be calculated exactly).

	Men	Women	*Row Total*
Studying	***a***	***b***	*a + b*
Non-studying	***c***	***d***	*c + d*
Column Total	*a + c*	*b + d*	*a + b + c + d (=n)*

	Men	Women	*Row total*
Studying	**1**	**9**	*10*
Not-studying	**11**	**3**	*14*
Column total	*12*	*12*	*24*

Fisher showed that the probability of obtaining any such set of values was given by the hypergeometric distribution:

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b}\binom{c+d}{d}}{\binom{n}{b+d}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\;b!\;c!\;d!\;n!}$$

where $\binom{n}{k}$ is the binomial coefficient and the symbol ! indicates the factorial operator. With the data above (using the first of the equivalent forms), this gives:

$$p = \binom{10}{1}\binom{14}{11}/\binom{24}{12} = \frac{10!\;14!\;12!\;12!}{1!\;9!\;11!\;3!\;24!} \approx 0.001346076$$

Enrichment of ranked lists

- Arbitrary thresholds
- Possible problems with gene lists
  - No “natural” value to filter thresholds (Fold Change ≥ 2 fold adj.p-value ≤ 0.01 etc.)
  - Results change with different threshold settings
  - Results change with background set
  - Gene name mapping ambiguity
  - Loss of statistical power due to thresholding (no resolution of significant signal with different strengths, weak signals neglected)

Comparison to a background set (universe)

- All possible genes that could appear in your gene list
  - After zero or low count genes removed in RNA-seq
  - If using microarrays - the “universe of genes” limited by the legitimate probes on platform (i.e minus any blacklisted probes).

- Annotation resources
  - GFF3 or GTF files from UCSC, RefSeq, Gencode, Ensembl
  - Some annotation terms may be subsets of other terms (GO ontology)
  - Many Bioconductor packages are geared towards providing annotation.

Sampling Bias

Multiple sources of bias confound functional enrichment analysis of global -omics data

James A. Timmons, Krzysztof J. Szkop and Iain J. Gallagher

Genome Biology 2015 **16**:186 | **DOI:** 10.1186/s13059-015-0761-7 | © Timmons et al. 2015

Published: 7 September 2015

- Technology bias
- Detection bias: In a transcriptomic experiment not all genes can be detected with equal reliability, to the extent that some genes are never detected as being 'regulated'.
- Biological bias: The transcriptome of a given cell type or tissue is highly specialized, to the point that it can be used to determine the identity of an unknown RNA profile efficiently.

Some useful advice

Nature Reviews Genetics **9**, 509-515 (July 2008) | doi:10.1038/nrg2363

Use and misuse of the gene ontology annotations

Seung Yon Rhee[1], Valerie Wood[2], Kara Dolinski[3] & Sorin Draghici[4]

An introduction to effective use of enrichment analysis software

Hannah Tipney and Lawrence Hunter*

Center for Computational Pharmacology, University of Colorado Denver, Aurora, CO 80045, USA
**Correspondence to*: Tel: +1 303 724 3369; E-mail: hannah.tipney@ucdenver.edu

JOURNAL OF THE ROYAL SOCIETY Interface

Separate enrichment analysis of pathways for up- and downregulated genes

Guini Hong[1], Wenjing Zhang[1], Hongdong Li[1], Xiaopei Shen[1] and Zheng Guo[1,2]

Gene pairs with various types of functional links defined in pathways tend to have positively correlated expression levels.

Gene name disambiguation

Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics

Barry R Zeeberg[†], Joseph Riss[†], David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett and John N Weinstein

[†] Contributed equally

Gene name errors are widespread in the scientific literature

Mark Ziemann, Yotam Eren and Assam El-Osta

Genome Biology 2016 17:177 | DOI: 10.1186/s13059-016-1044-7 | © The Author(s). 2016

Published: 23 August 2016

Getting the latest gene nomenclature

HGNC
HUGO Gene Nomenclature Committee

Search everything | Search symbols, keywords or IDs

Use * to search with a root symbol (eg ZNF*)

Home | Downloads | Gene Families | Tools | Useful links | About | Newsletters | Contact Us | Help | VGNC | Request Symbol

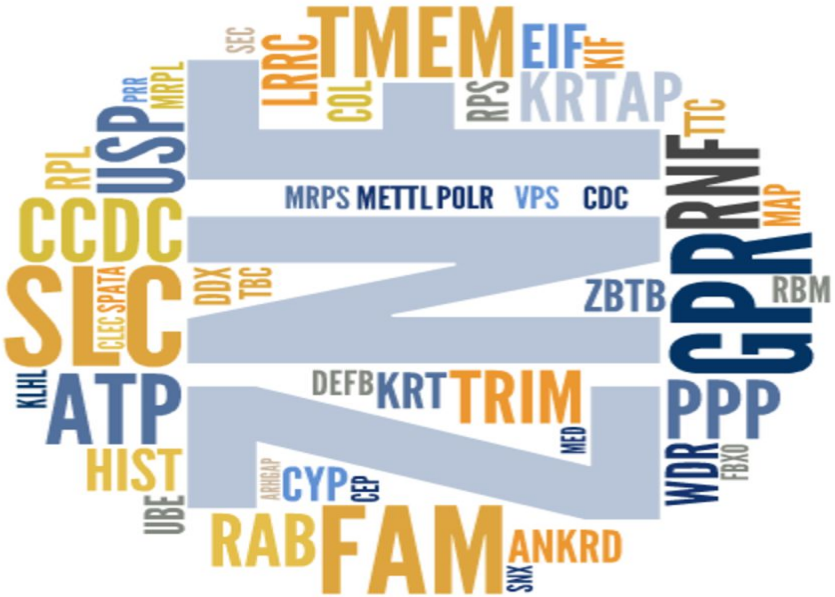
HGNC is responsible for approving unique symbols and names for human loci, including protein coding genes, ncRNA genes and pseudogenes, to allow unambiguous scientific communication.

genenames.org is a curated online repository of HGNC-approved gene nomenclature, gene families and associated resources including links to genomic, proteomic and phenotypic information.

Search our catalogue of more than 40,000 symbol reports using our improved search engine (see Search help), search lists of symbols using our Multi-symbol checker and identify possible orthologs using our HCOP tool.

Download our ready-made data files from our Statistics and Downloads page, create your own datasets using either our Custom Downloads tool or BioMart service, or write a script/program utilising our REST service.

Submit your gene symbol and name proposals to us to be accredited with HGNC approved nomenclature for use in publications, databases and presentations.



Extracting meaning from biomedical data

- Over Representation Analysis
- Functional Class Scoring
- Topology or Connectivity driven analysis

Common approaches:

  - Gene Ontology enrichment
  - Pathway enrichment
  - Gene Set Enrichment Analysis (GSEA)
  - Interactome or Phospho-proteome analysis
  - Network analysis

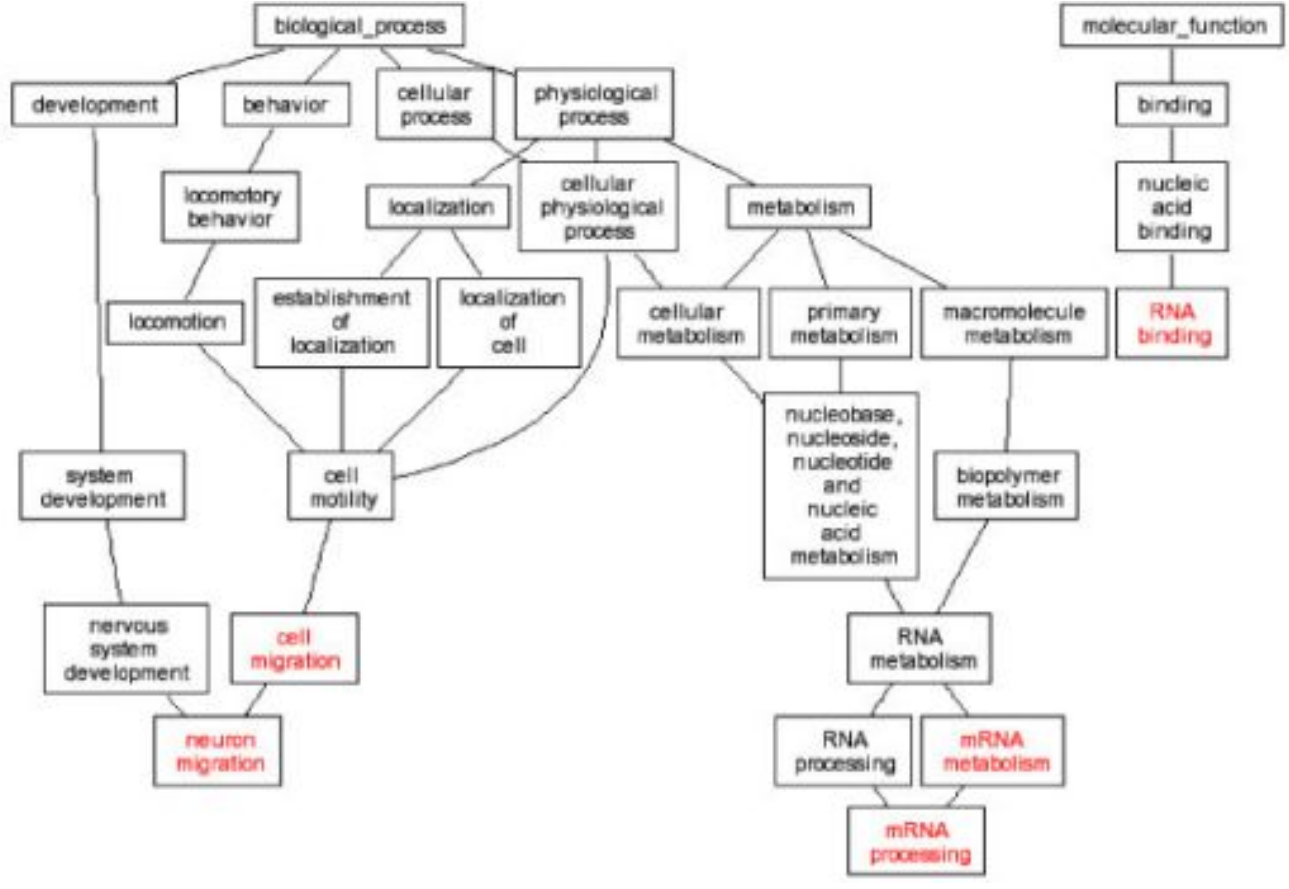
Gene Ontology

The Gene Ontology project provides an ontology (a machine readable controlled vocabulary) of defined terms representing properties of gene products and covers three domains:

- CC: **Cellular Component**, the parts of a cell or its extracellular environment;
- MF: **Molecular Function**, the elemental activities of a gene product at the molecular level, such as binding or catalysis;
- BP: **Biological Process**, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.

Each domain has 9 levels (generic to specific) organised as a directed di-acyclic graph.

Ontology DAG



Gene Ontology evidence codes

Quality and reliability (multiple lines of evidence) filter for GSEA.

- Inferred from Electronic Annotation (IEA)

- Inferred from Experiment (EXP)
- Inferred from Direct Assay (IDA)
- Inferred from Physical Interaction (IPI)
- Inferred from Mutant Phenotype (IMP)
- Inferred from Genetic Interaction (IGI)
- Inferred from Expression Pattern (IEP)

- Inferred from Sequence or structural Similarity (ISS)
- Inferred from Sequence Orthology (ISO)
- Inferred from Sequence Alignment (ISA)
- Inferred from Sequence Model (ISM)
- Inferred from Genomic Context (IGC)
- Inferred from Biological aspect of Ancestor (IBA)
- Inferred from Biological aspect of Descendant (IBD)
- Inferred from Key Residues (IKR)
- Inferred from Rapid Divergence(IRD)
- Inferred from Reviewed Computational Analysis (RCA)

A collection of GO resources

- NIH DAVID
- AMIGO (GO consortium) & Panther
- WebGestalt
- BINGO (Cytoscape)
- GOrilla, g:Profiler and Revigo
- FatiGO (Babelomics-5 toolkit)
- L2L
- topGO (Bioconductor)
- STEM (short time series expression miner)

*** Welcome to DAVID 6.8 with updated Knowledgebase (more info). ***
*** If you are looking for DAVID 6.7, please visit our development site. ***

Recommending: A paper published in ***Nature Protocols*** describes step-by-step procedure to use DAVID!

Shortcut to DAVID Tools

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

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

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

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

Welcome to DAVID 6.8

2003 - 2016

The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) v6.8 comprises a full Knowledgebase update to the sixth version of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

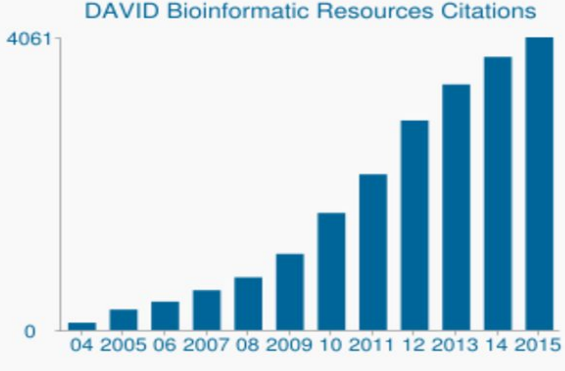
- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view.
- Search for other functionally related genes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

Search

What's Important in DAVID?

- New requirement to cite DAVID
- IDs of Affy Exon and Gene arrays supported
- Novel Classification Algorithms
- Pre-built Affymetrix and Illumina backgrounds
- User's customized gene background
- Enhanced calculating speed

Statistics of DAVID



- > 21,000 Citations
- Average Daily Usage: ~2,600 gene lists/sublists from ~800 unique researchers.

AmiGO 2

AmiGO 2 | Home | Search | Browse | Tools & Resources | Help | Feedback | About | AmiGO 1.8

AmiGO 2

More information on quick search

Quick search

Search

Search Templates

Use predefined **templates** to explore Gene Ontology data.

Go »

Advanced Search

Interactively **search** the Gene Ontology data for annotations, gene products, and terms using a powerful search syntax and filters.

Search

Browse the Ontology

Use the drill-down **browser** to view the ontology structure with annotation counts.

Go »

GOOSE

Use GOOSE to query the legacy GO database with **SQL**.

Go »

Term Enrichment Service

Your genes here...

biological process

Homo sapiens

Statistics

View the most recent **statistics** about the Gene Ontology data in AmiGO.

Go »

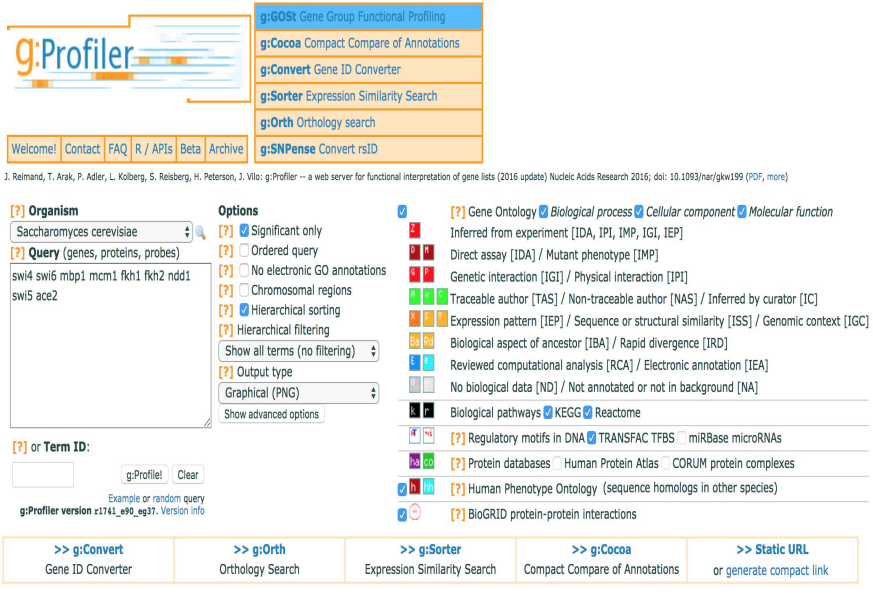
And Much More...

Many **more tools** are available from the software list, such as alternate searching modes, Visualize, non-JavaScript pages.

Go »

Build your own annotated gene sets

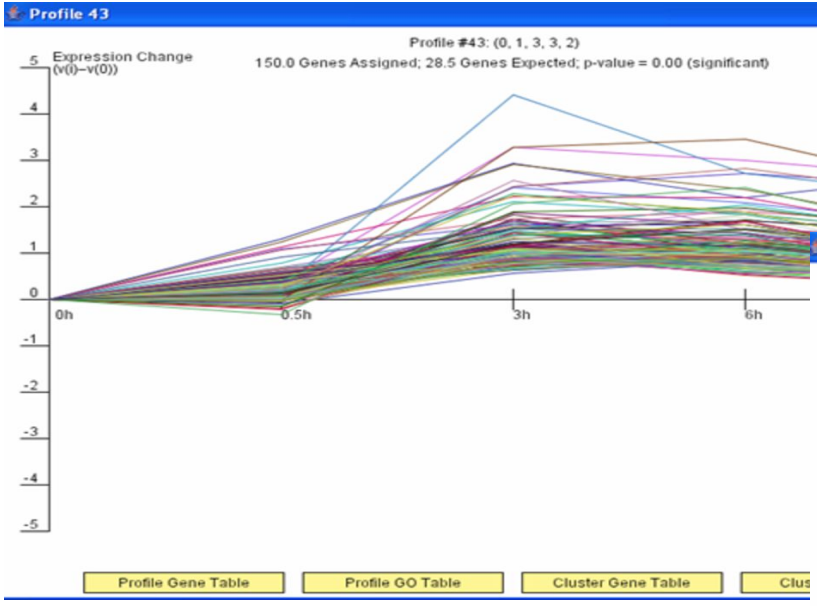
g:Profiler and Revigo



Reimand et al., Nucleic Acids Res. 2016

Supek et al., PLoS One. 2011

STEM: Short Time Series Expression Miner



Integrates ontologies and clustering methods to identify biologically related genes with similar expression patterns across time.

Ernst & Bar-Joseph BMC Bioinformatics. 2006

Pathway Analysis Resources

Pathguide » the pathway resource list

Navigation

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

Search

Organisms: All

Availability: All

Standards: All

Reset | Search

Analysis

- Statistics
- Database Interactions

Complete Listing of All Pathguide Resources

Pathguide contains information about **702** biological pathway related resources and molecular interaction related resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

News

Major new update of Pathguide
September 2017
We now have information for over 690 resources!

Major new update of Pathguide
August 2013
We now have information about ~550 resources!

Protein-Protein Interactions

Database Name (Order: alphabetically | by web popularity)	Full Record	Availability	Standards
2P2Idb - The Protein-Protein Interaction Inhibition Database	Details	Free	
3D-Interologs - 3D-Interologs	Details	Free	
3DID - 3D interacting domains	Details	Free	
ACSN - Atlas of Cancer Signalling Network	Details	Free	BioPAX
ADAN - Prediction of protein-protein interaction of modular domains	Details	X	
AHD2.0 - Arabidopsis Hormone Database 2.0	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X	
aMAZE - Protein Function and Biochemical Pathways Project	Details	X	
ANAP - Arabidopsis Network Analysis Pipeline	Details	Free	
ANIA - ANnotation and Integrated Analysis of the 14-3-3 interactome	Details	Free	
AnimalTFDB - Animal Transcription Factor Database	Details	Free	

OPEN ACCESS Freely available online

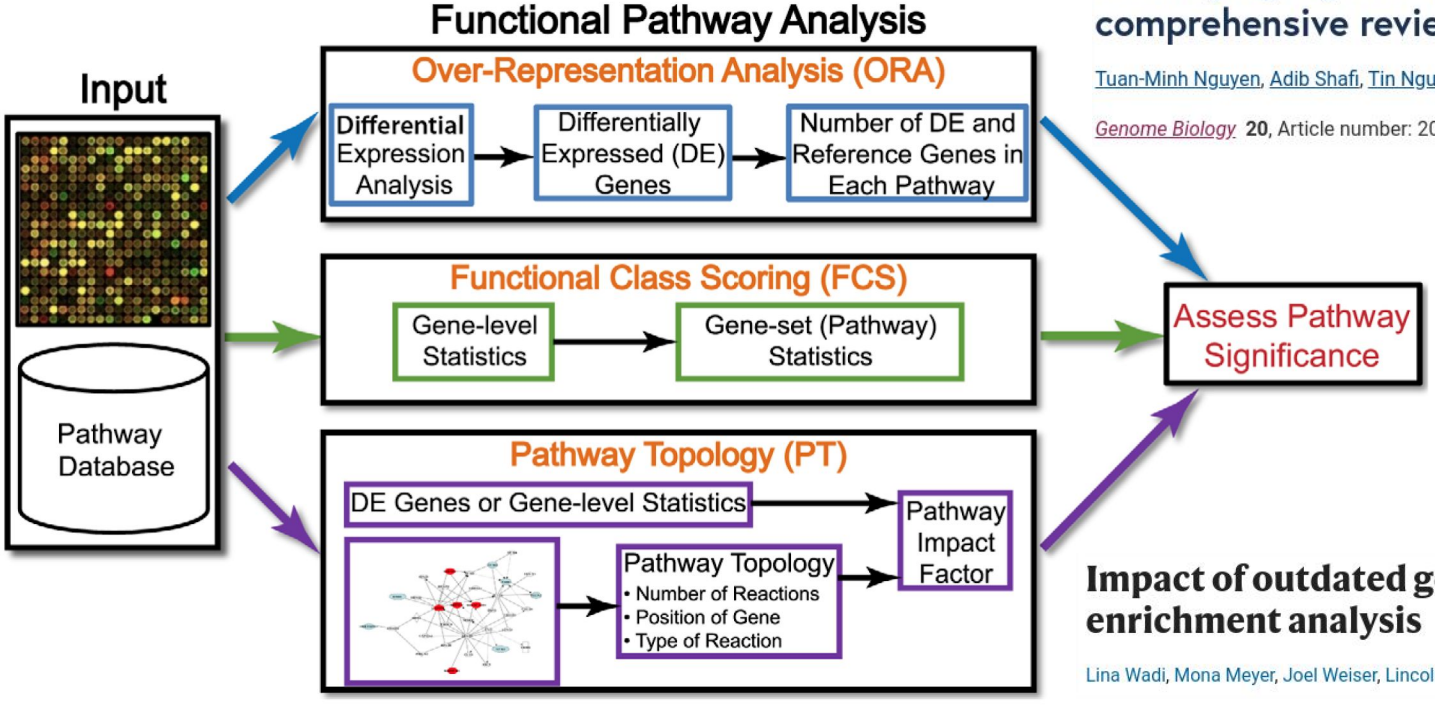
PLOS COMPUTATIONAL BIOLOGY

Review

Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges

Purvesh Khatri[1,2]*, Marina Sirota[1,2], Atul J. Butte[1,2]*

1 Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, United States of America, **2** Lucile Packard Children's Hospital, Palo Alto, California, United States of America



Identifying significantly impacted pathways: a comprehensive review and assessment

Tuan-Minh Nguyen, Adib Shafi, Tin Nguyen & Sorin Draghici

Genome Biology **20**, Article number: 203 (2019) | Cite this article

Impact of outdated gene annotations on pathway enrichment analysis

Lina Wadi, Mona Meyer, Joel Weiser, Lincoln D Stein & Jüri Reimand

A collection of resources and tools for over representation analysis of pathways

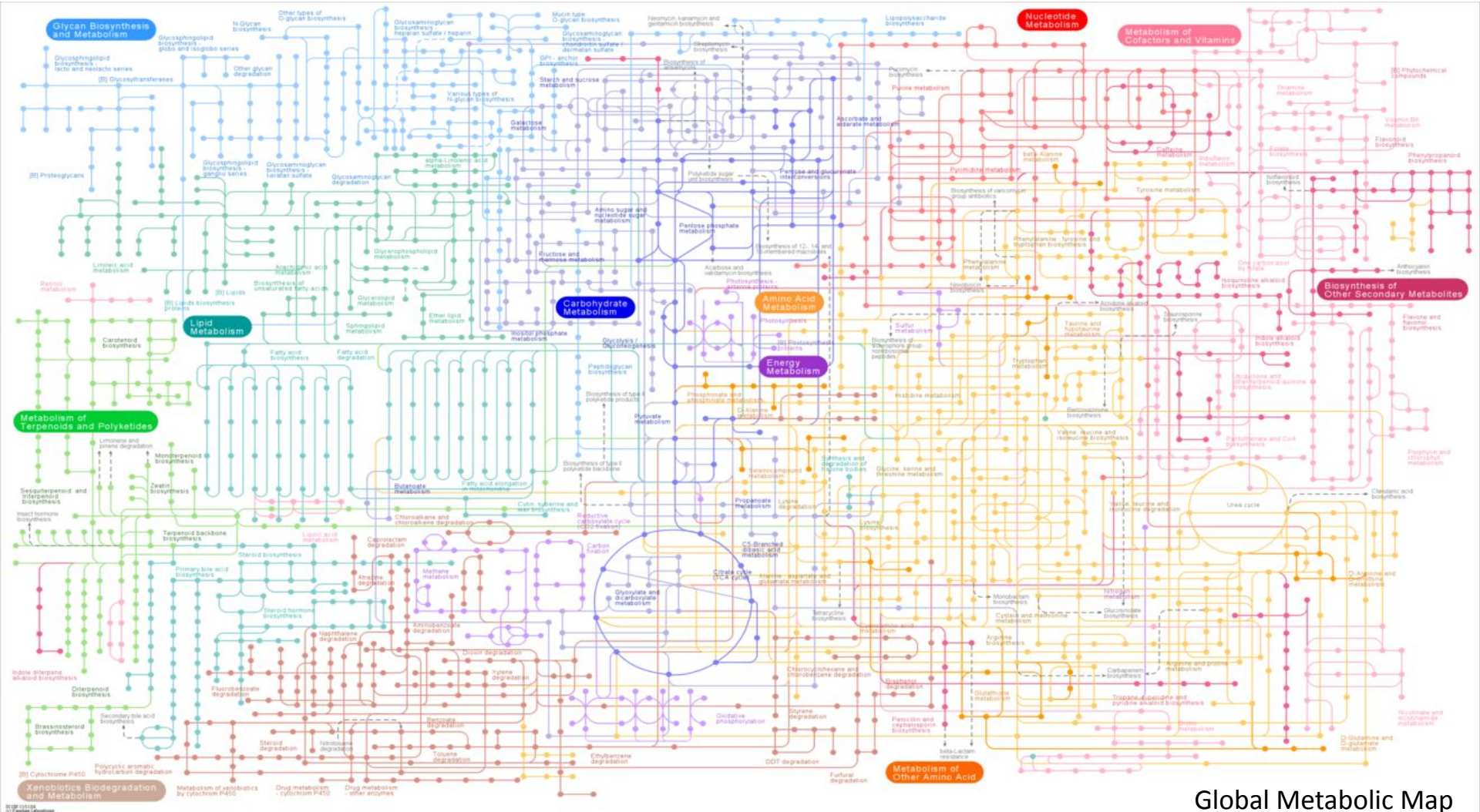
- KEGG (KEGGRest -Bioconductor)
- Reactome (Reactome PA & FIviz apps)
- Panther
- BioCyc (HumanCyc)
- WikiPathways
- Pathway Commons
- SPIA
- OmniPath
- Pathview (Bioconductor)
- ClusterProfiler (Bioconductor)
- Paradigm
- Commercial: Ingenuity, Metacore, Pathway studio

BEWARE!

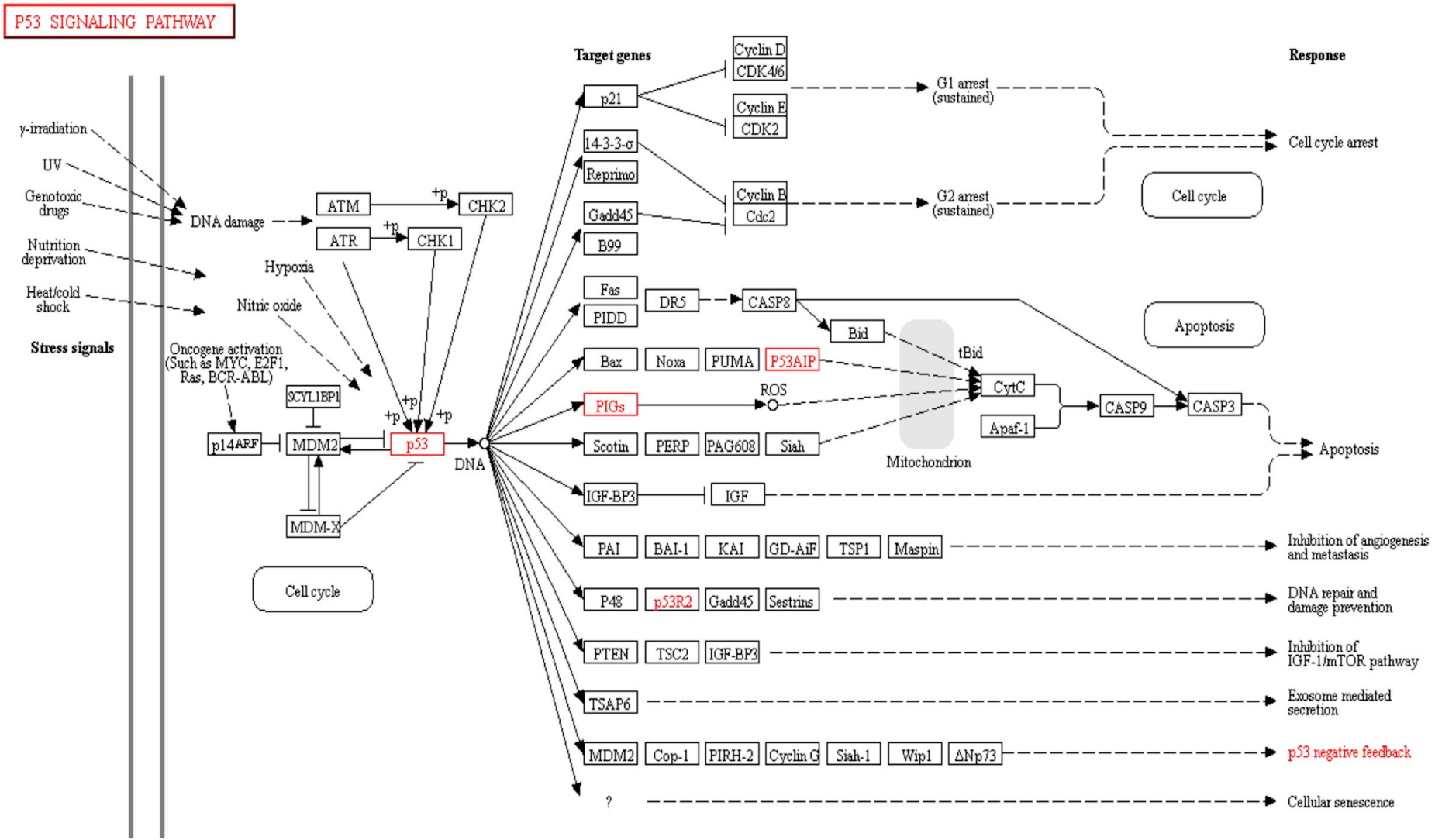
Pathways and Gene Sets from most databases are subjective and biased.

Most public pathway resources don't provide evidence for node connectivity.

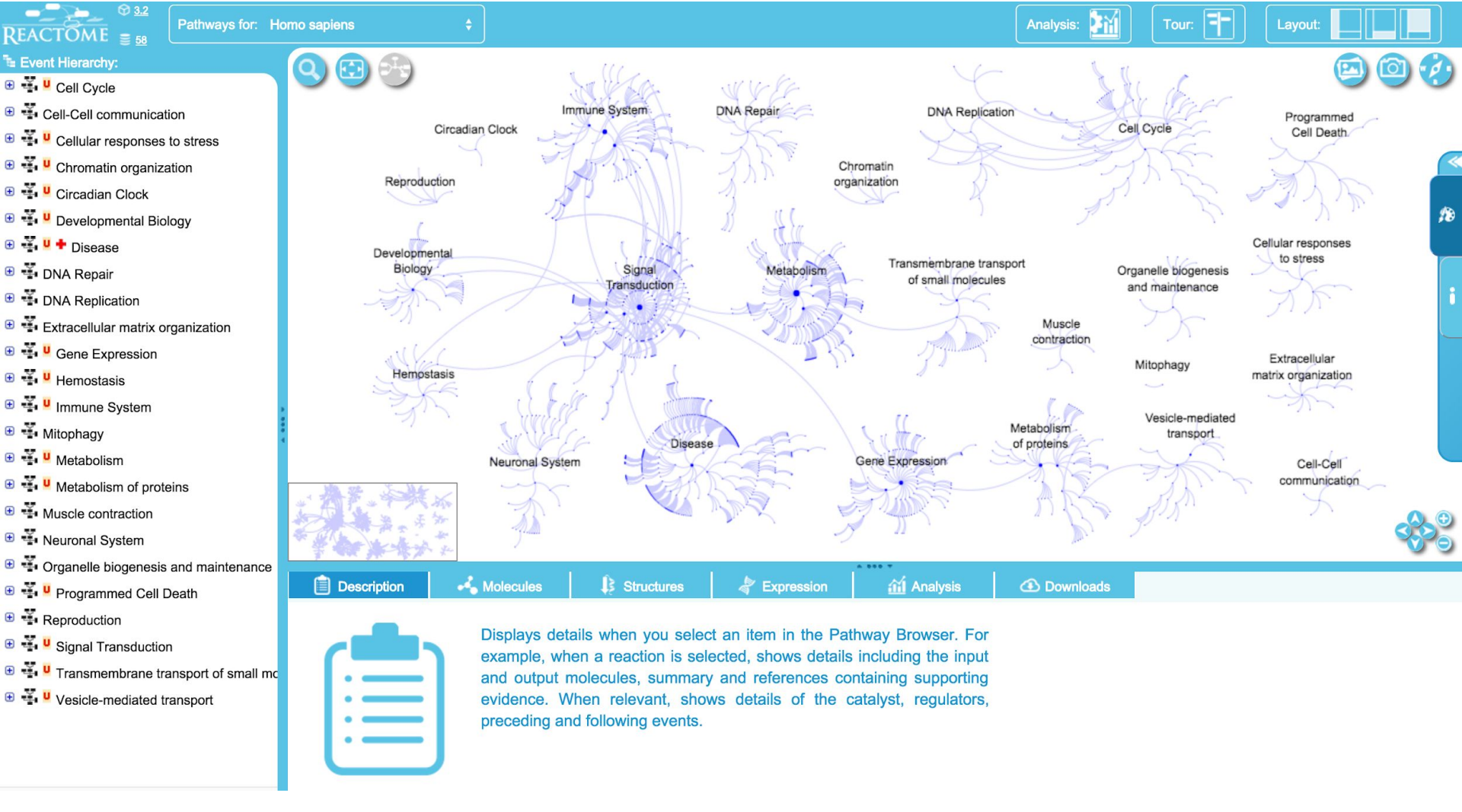
KEGG: Kyoto Encyclopedia of Genes and genomes



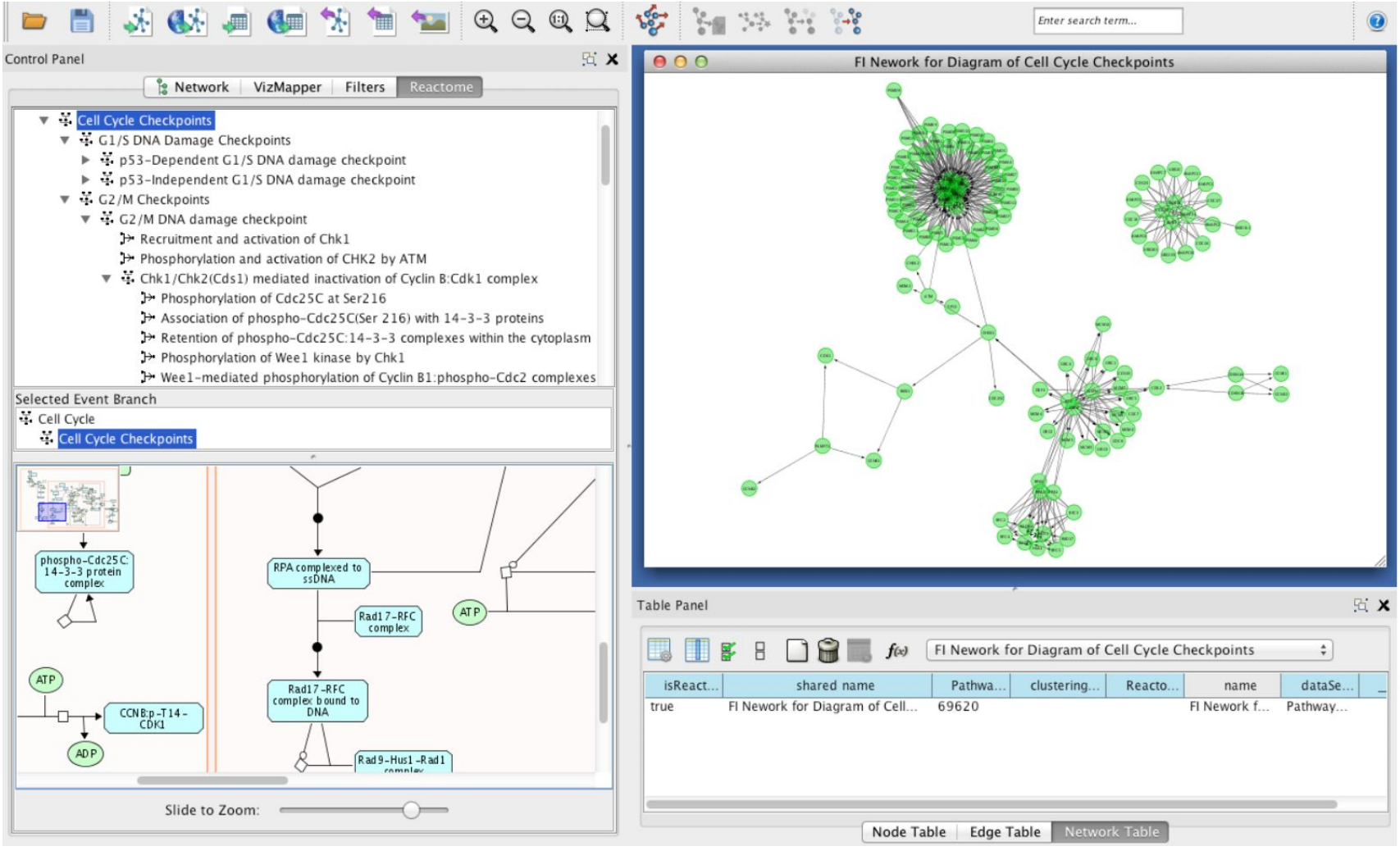
Global Metabolic Map



Reactome

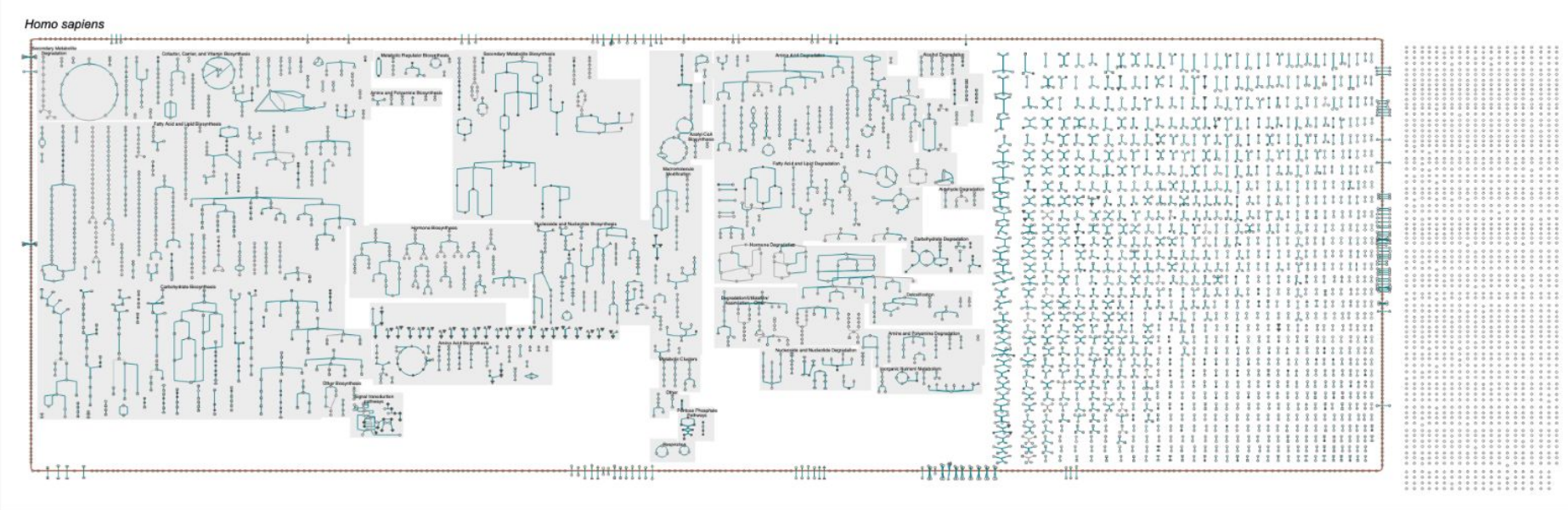

REACTOME
3.2
58
Pathways for: Homo sapiens
Analysis:
Tour:
Layout:
Event Hierarchy:
Cell Cycle
Cell-Cell communication
Cellular responses to stress
Chromatin organization
Circadian Clock
Developmental Biology
Disease
DNA Repair
DNA Replication
Extracellular matrix organization
Gene Expression
Hemostasis
Immune System
Mitophagy
Metabolism
Metabolism of proteins
Muscle contraction
Neuronal System
Organelle biogenesis and maintenance
Programmed Cell Death
Reproduction
Signal Transduction
Transmembrane transport of small mo
Vesicle-mediated transport
Circadian Clock
Immune System
DNA Repair
DNA Replication
Cell Cycle
Programmed Cell Death
Reproduction
Chromatin organization
Developmental Biology
Signal Transduction
Metabolism
Transmembrane transport of small molecules
Organelle biogenesis and maintenance
Cellular responses to stress
Muscle contraction
Hemostasis
Mitophagy
Extracellular matrix organization
Neuronal System
Disease
Gene Expression
Metabolism of proteins
Vesicle-mediated transport
Cell-Cell communication
Description
Molecules
Structures
Expression
Analysis
Downloads
Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events.

Reactome FIviz



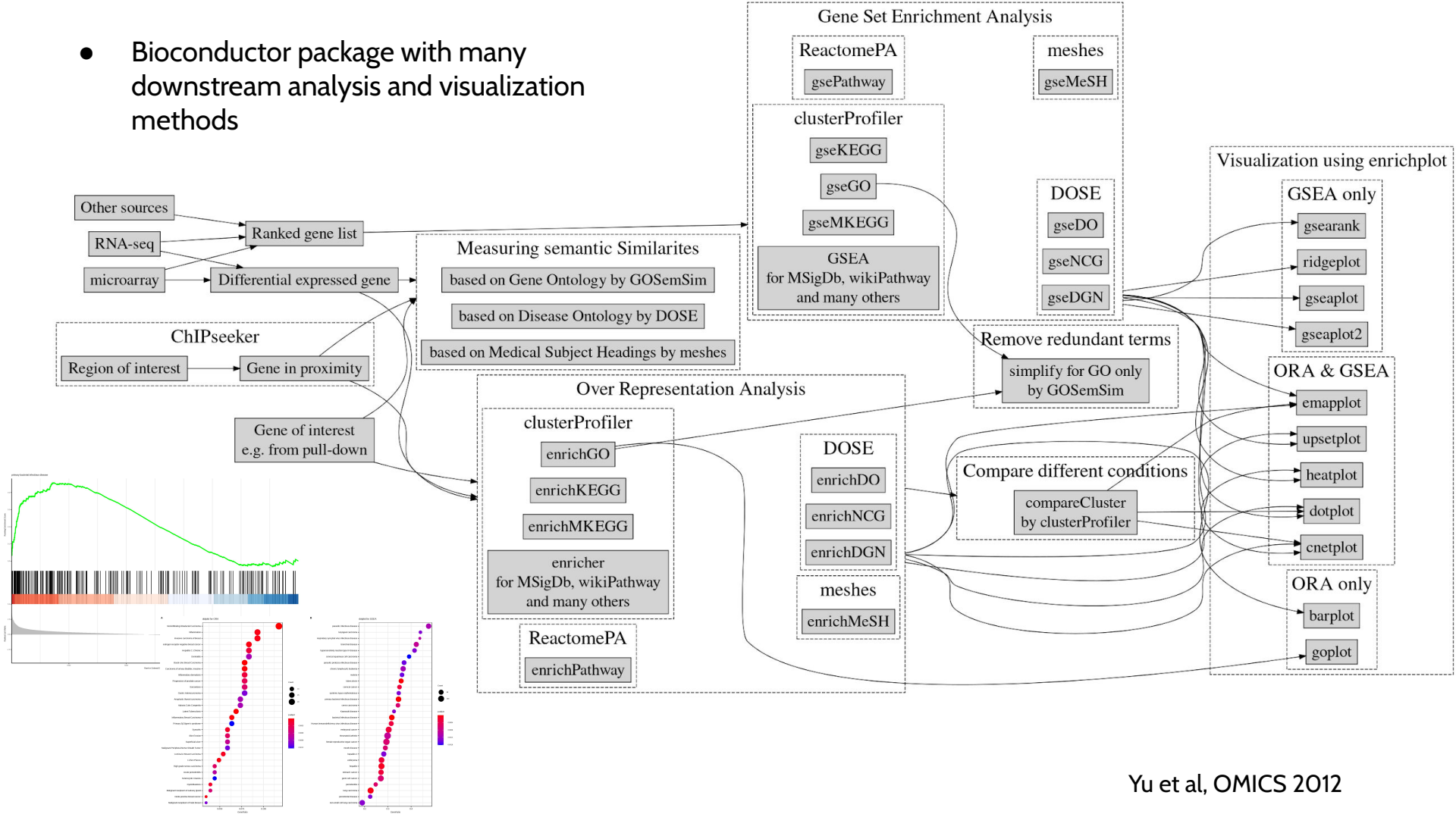
BioCyc (HumanCyc)

- HumanCyc provides an encyclopedic reference on human ~350 metabolic pathways.
- It provides a zoomable human metabolic map diagrams, and it has been used to generate a steady-state quantitative model of human metabolism.
- Associated PathwayTools provide a variety of functionality.



ClusterProfiler

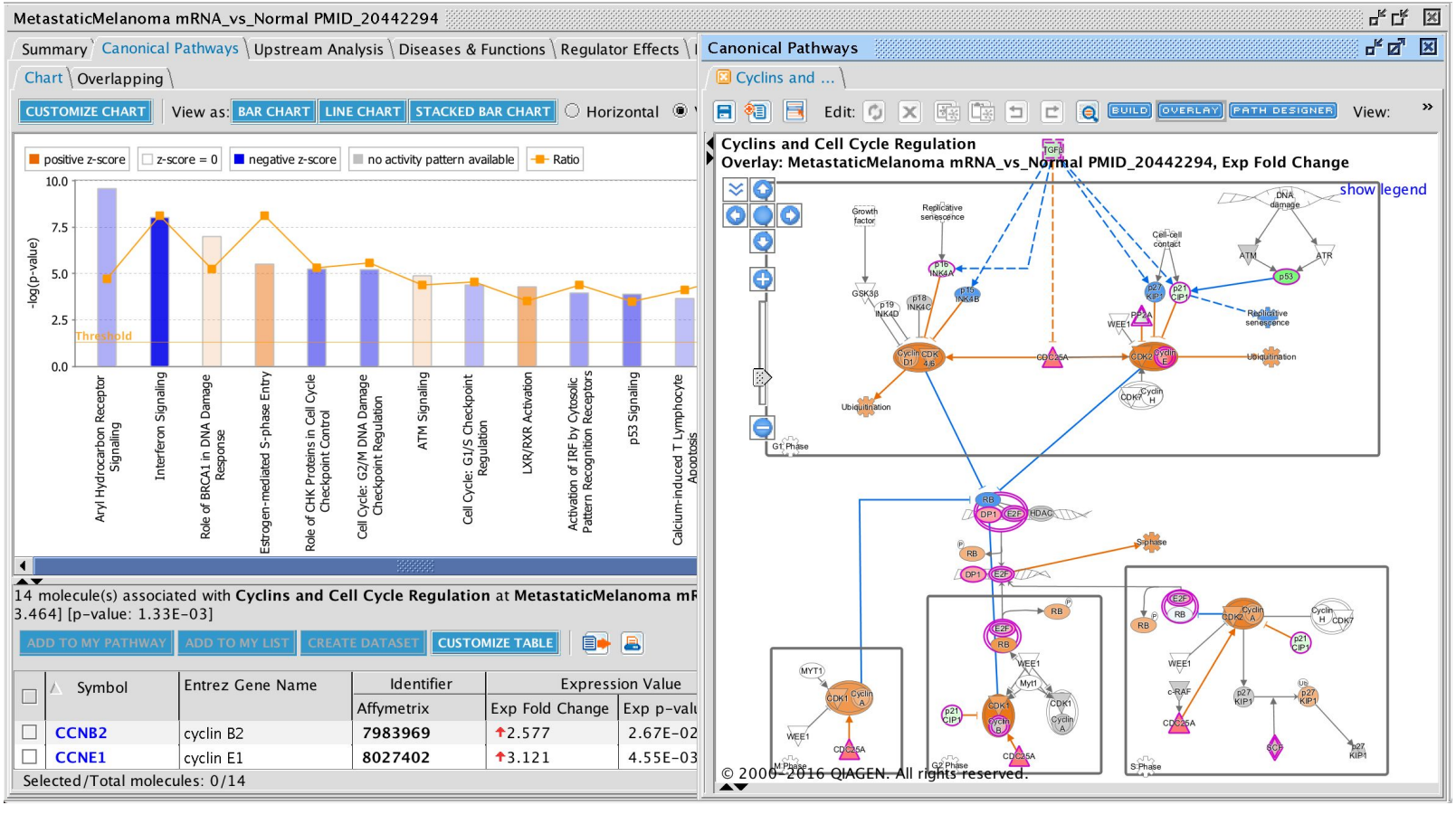
- Bioconductor package with many downstream analysis and visualization methods



Yu et al, OMICS 2012

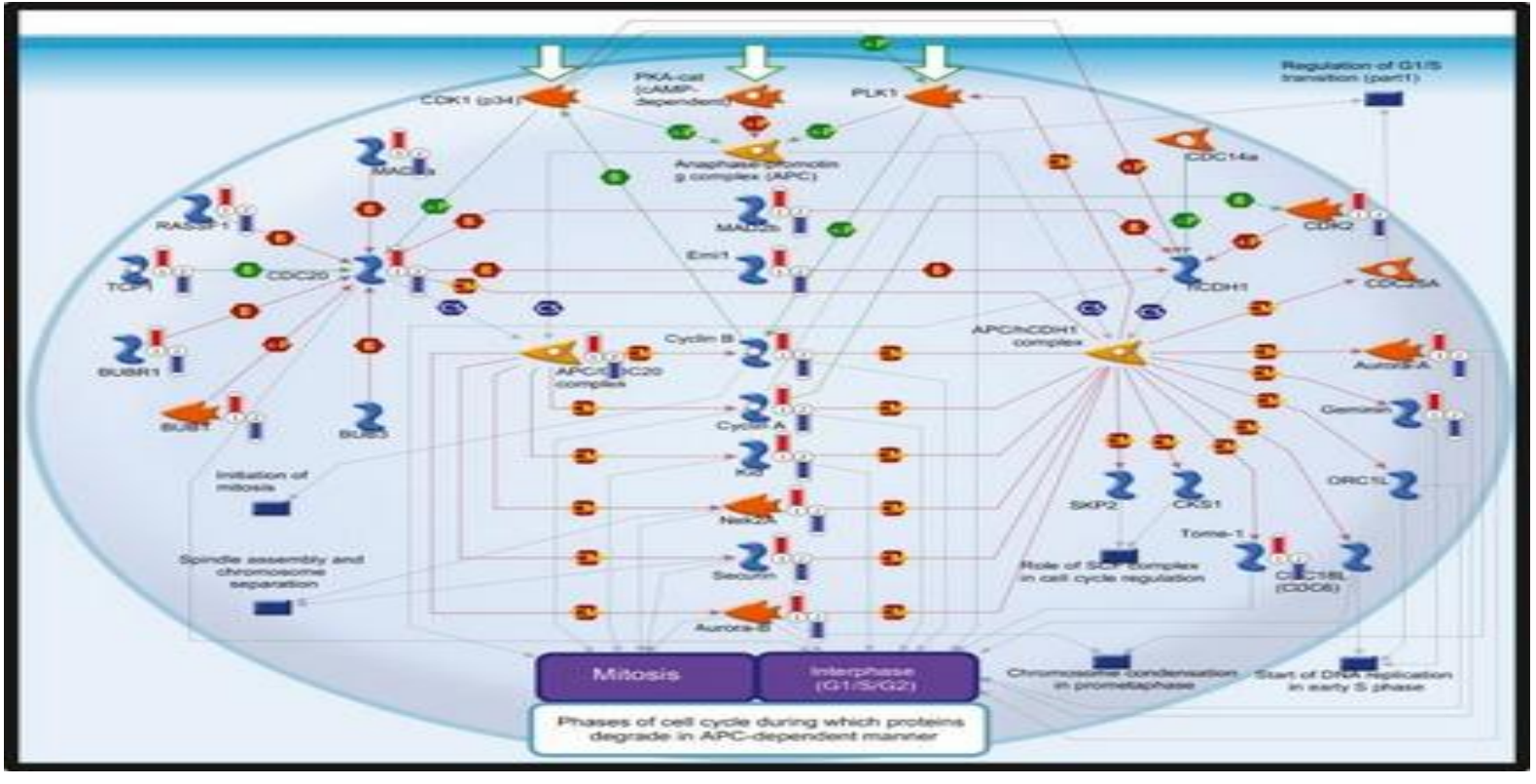
Commercial Pathway Resources

IPA: Ingenuity Pathway Analysis



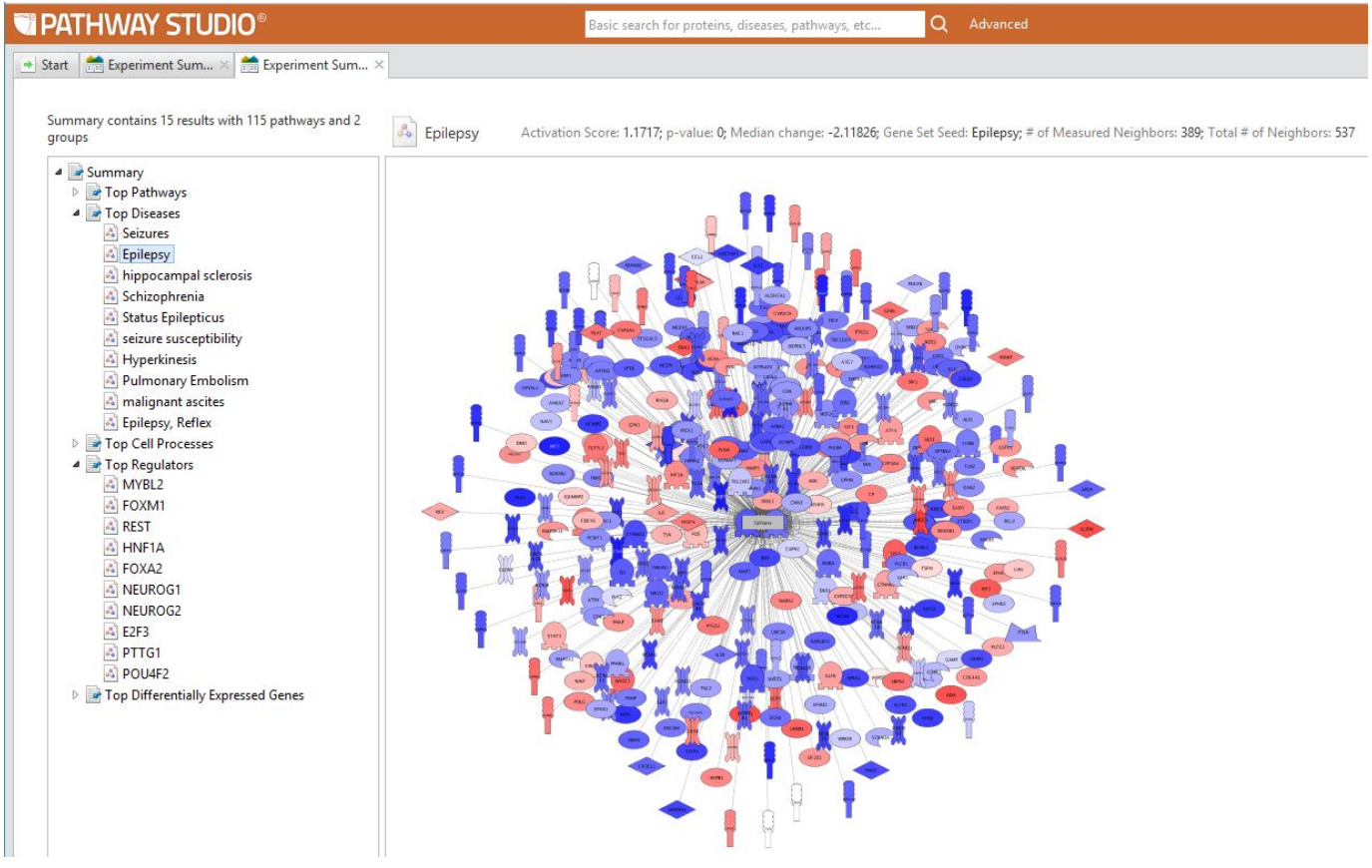
Commercial Pathway Resources

MetaCore



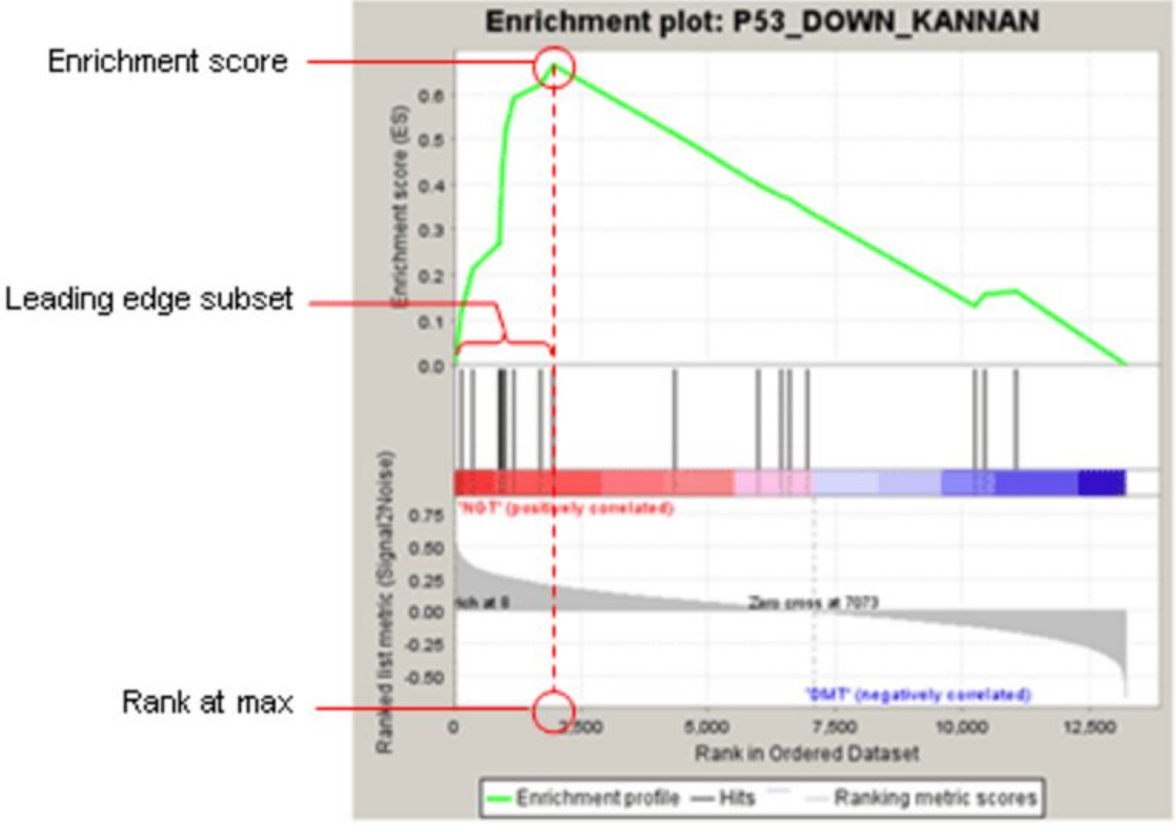
Commercial Pathway Resources

PathwayStudio



Gene Set Enrichment Analysis (GSEA)

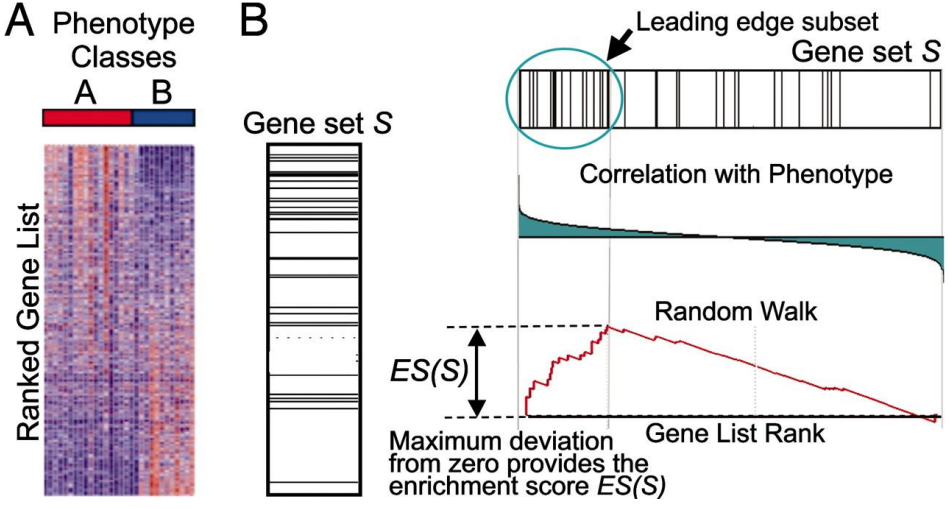
- Enrichment Score (ES), which reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes.



GSEA calculates the Enrichment Score (ES) by walking down the ranked list of genes, increasing a running-sum statistic when a gene is in the gene set and decreasing it when it is not.

Interpreting GSEA

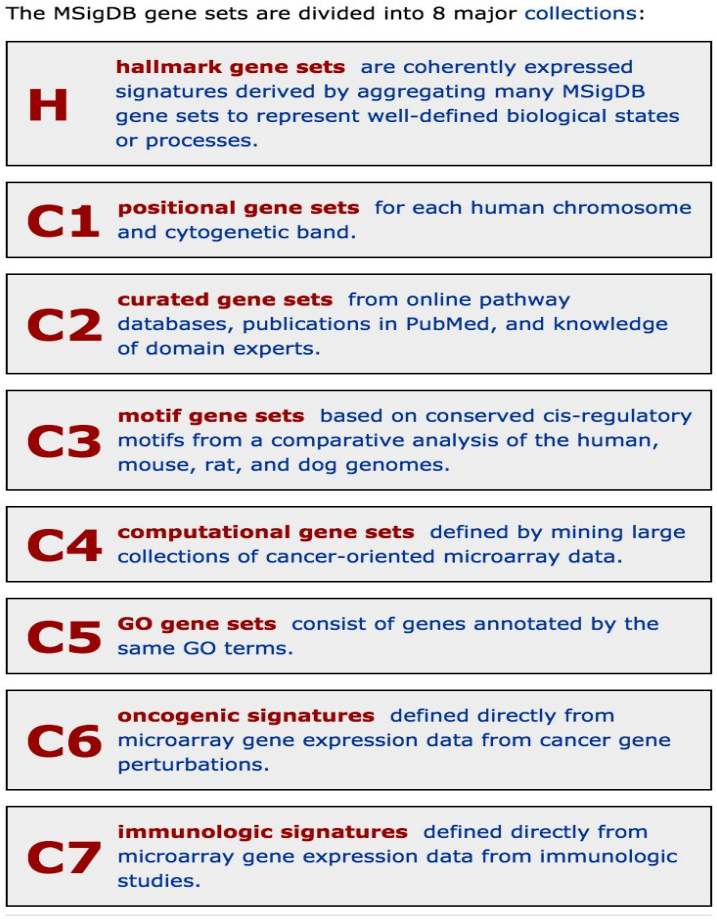
- The magnitude of the increment depends on the correlation of the gene with the phenotype. The ES is the maximum deviation from zero encountered in walking the list. A positive ES indicates gene set enrichment at the top of the ranked list; a negative ES indicates gene set enrichment at the bottom of the ranked list.
- Gene sets with a distinct peak at the beginning or end of the ranked list are generally the most interesting.
- The **leading edge subset** of a gene set is the subset of members that contribute most to the ES.
- Use FDR of 25%



MSigDB

- The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software.
- GMT file format
- Used together with BROAD institute GSEA java package or web app.

- Can also be used with GSEABase and EGSEA Bioconductor analysis tools.



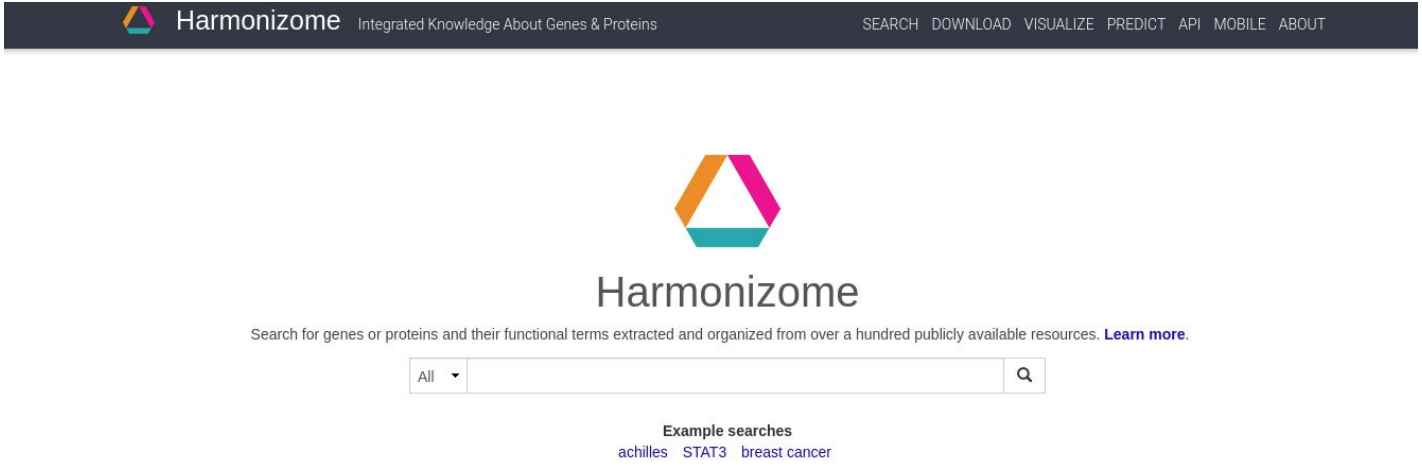
EnrichR & Harmanizome

Nucleic Acids Res. 2016 Jul 8; 44(Web Server issue): W90–W97.
Published online 2016 May 3. doi: 10.1093/nar/gkw377
PMCID: PMC4987924
PMID: 27141961

Enrichr: a comprehensive gene set enrichment analysis web server 2016 update

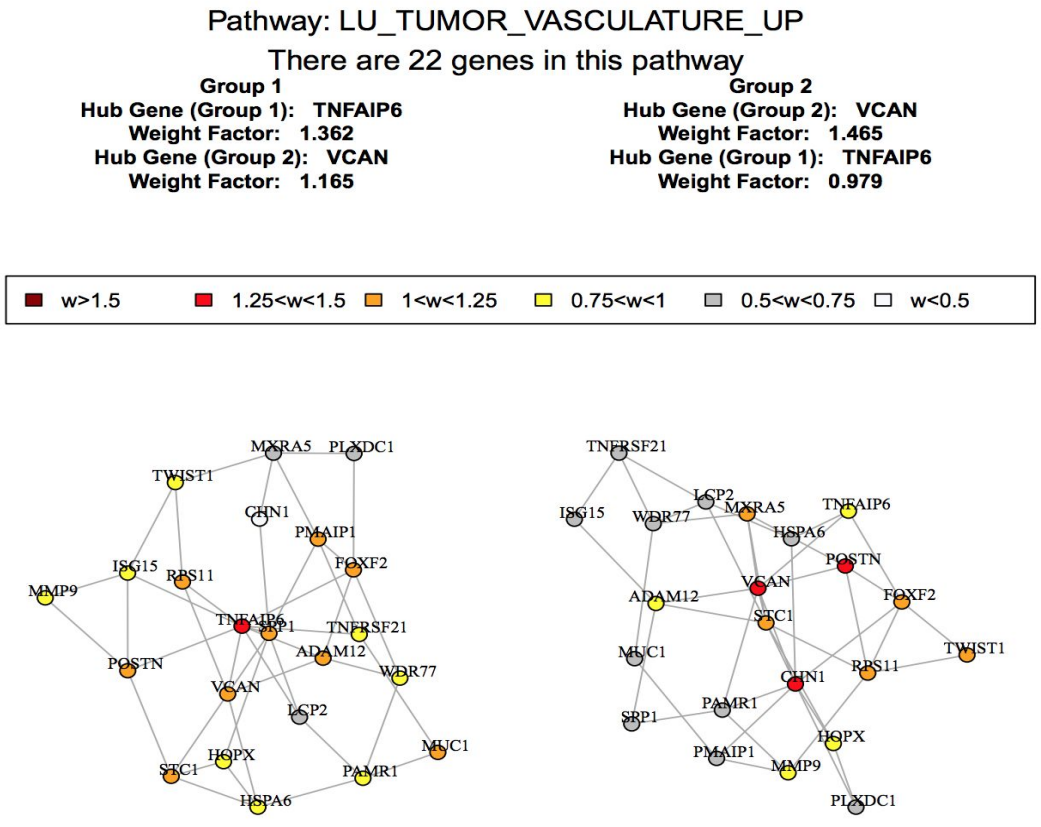
Maxim V. Kuleshov,[1] Matthew R. Jones,[1] Andrew D. Rouillard,[1] Nicolas F. Fernandez,[1] Qiaonan Duan,[1] Zichen Wang,[1] Simon Koplev,[1] Sherry L. Jenkins,[1] Kathleen M. Jagodnik,[2] Alexander Lachmann,[1] Michael G. McDermott,[1] Caroline D. Monteiro,[1] Gregory W. Gundersen,[1] and Avi Ma'ayan[1,*]

Python API and Web interface



Other Bioconductor/R Enrichment Analysis Tools

- Camera
- limma:
  - roast
  - mroast
  - romer
- GSAR
- GSA



Be aware of the statistical tests used and their assumptions!

Exploring Topology and Connectivity

- Macromolecules (DNA, metabolites, other entities) and Gene products (Proteins, Non-coding RNA) interact in specific ways to regulate phenotypes.
- Evolutionarily defines and functional properties such as Protein-Protein interactions, subcellular localization, Co-Expression, Protein domains, Post-translational modification etc can be used for determining connectivity.
- Tools to explore topology and connectivity can help us understand complex biological processes.
    - SPIA
    - STRING
    - GeneMania
    - GeneWalk
    - Cytoscape & Bioconductor packages

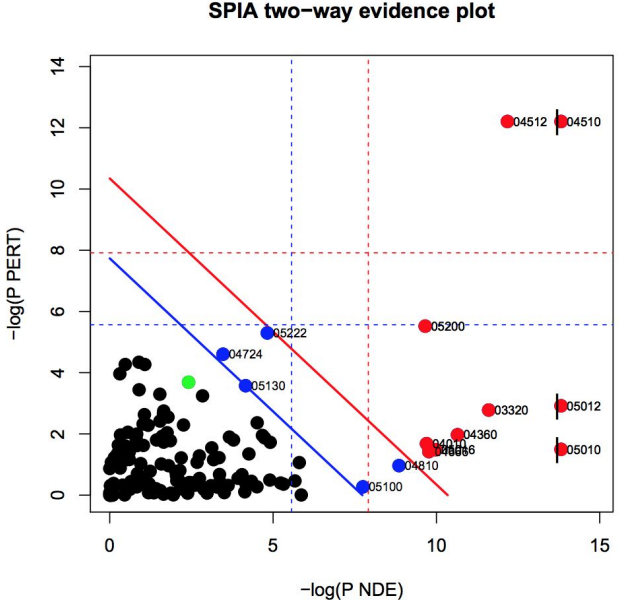
SPIA: Signaling Pathway Impact Analysis

Bioinformatics. 2009 Jan 1; 25(1): 75–82.
Published online 2008 Nov 5. doi: 10.1093/bioinformatics/btn577
PMCID: PMC2732297
PMID: 18990722

A novel signaling pathway impact analysis

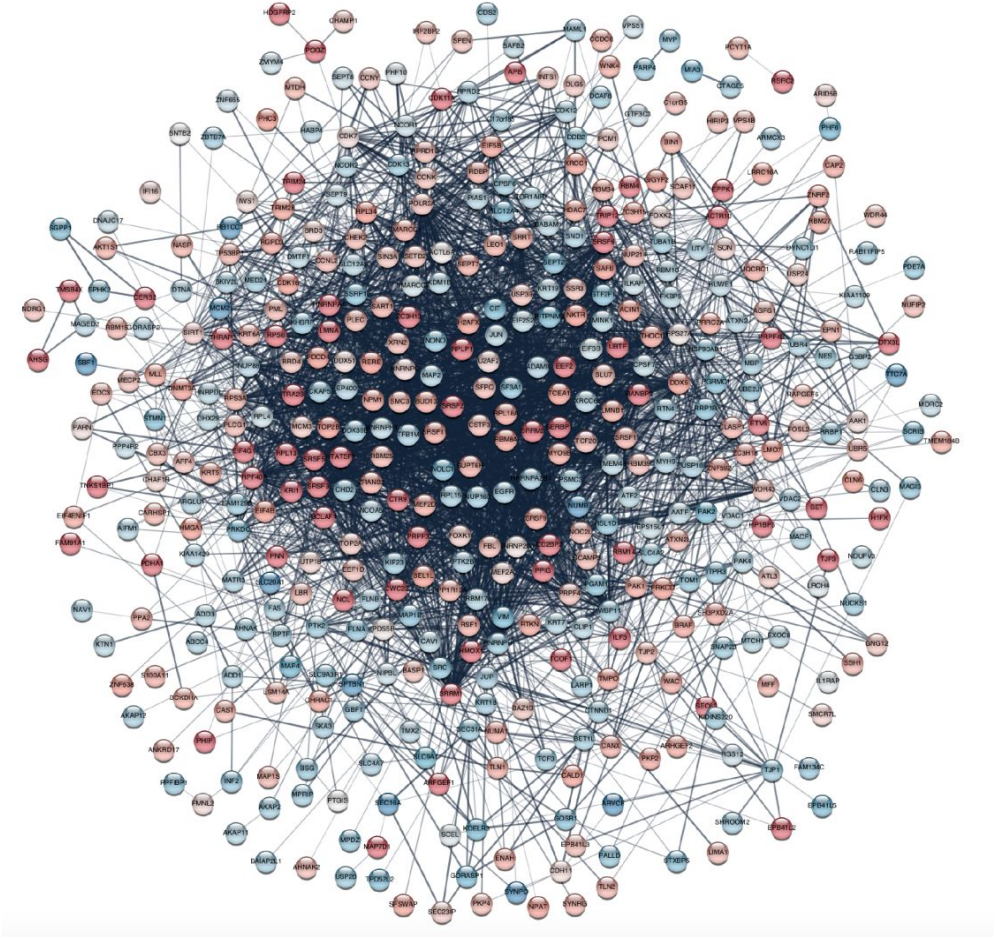
Adi Laurentiu Tarca,[1,2] Sorin Draghici,[1,*] Purvesh Khatri,[1] Sonia S. Hassan,[2] Pooja Mittal,[2] Jung-sun Kim,[2] Chong Jai Kim,[2] Juan Pedro Kusanovic,[2] and Roberto Romero[2]

SPIA uses the information from a set of differentially expressed genes and their fold changes, as well as pathways topology in order to assess the significance of the pathways in the condition under the study.



STRING

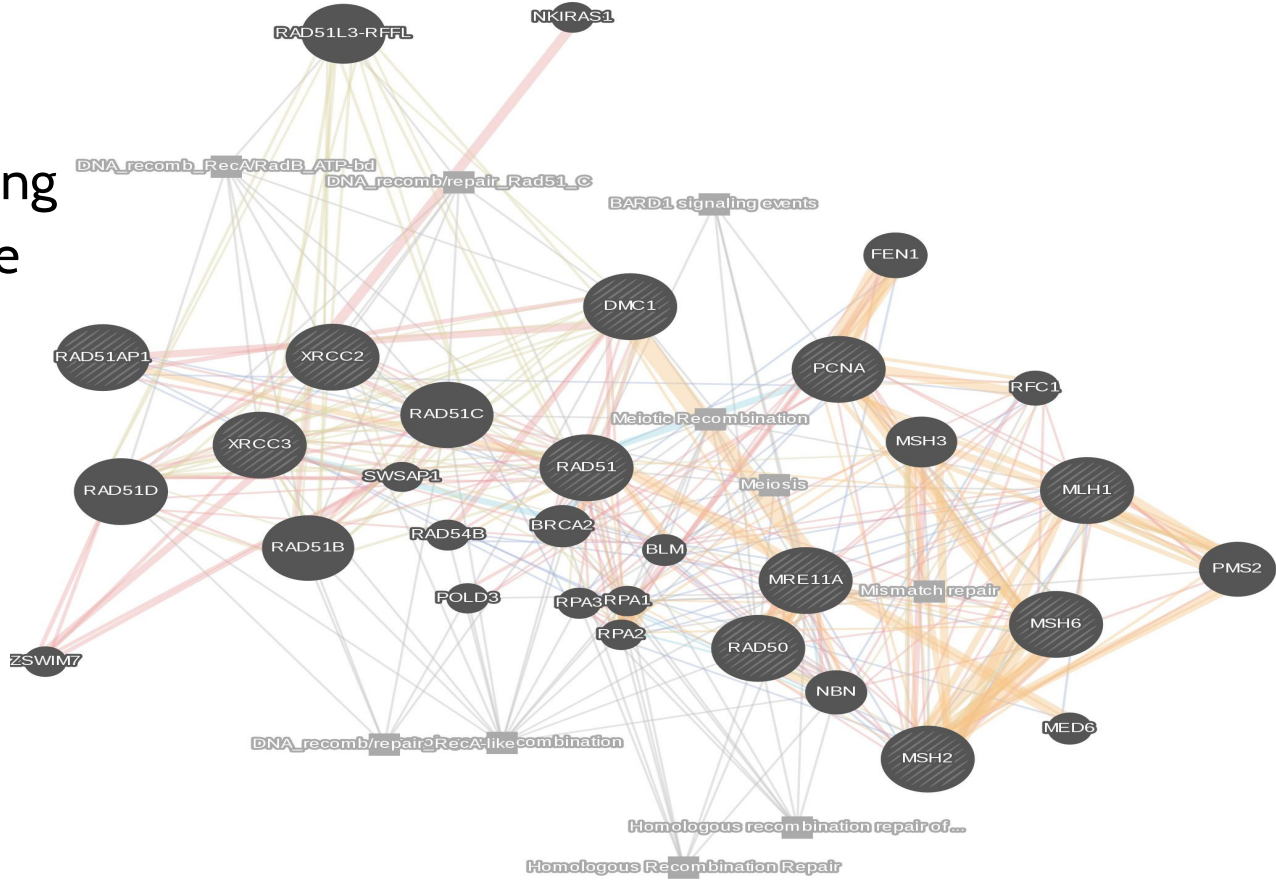
- Web app (https://string-db.org/) and STRINGdb bioconductor package
- Determines how a list of genes or proteins interacts
- Integration of protein–protein interactions, including direct (physical) as well as indirect (functional) associations, litterature evidence and co-expression.



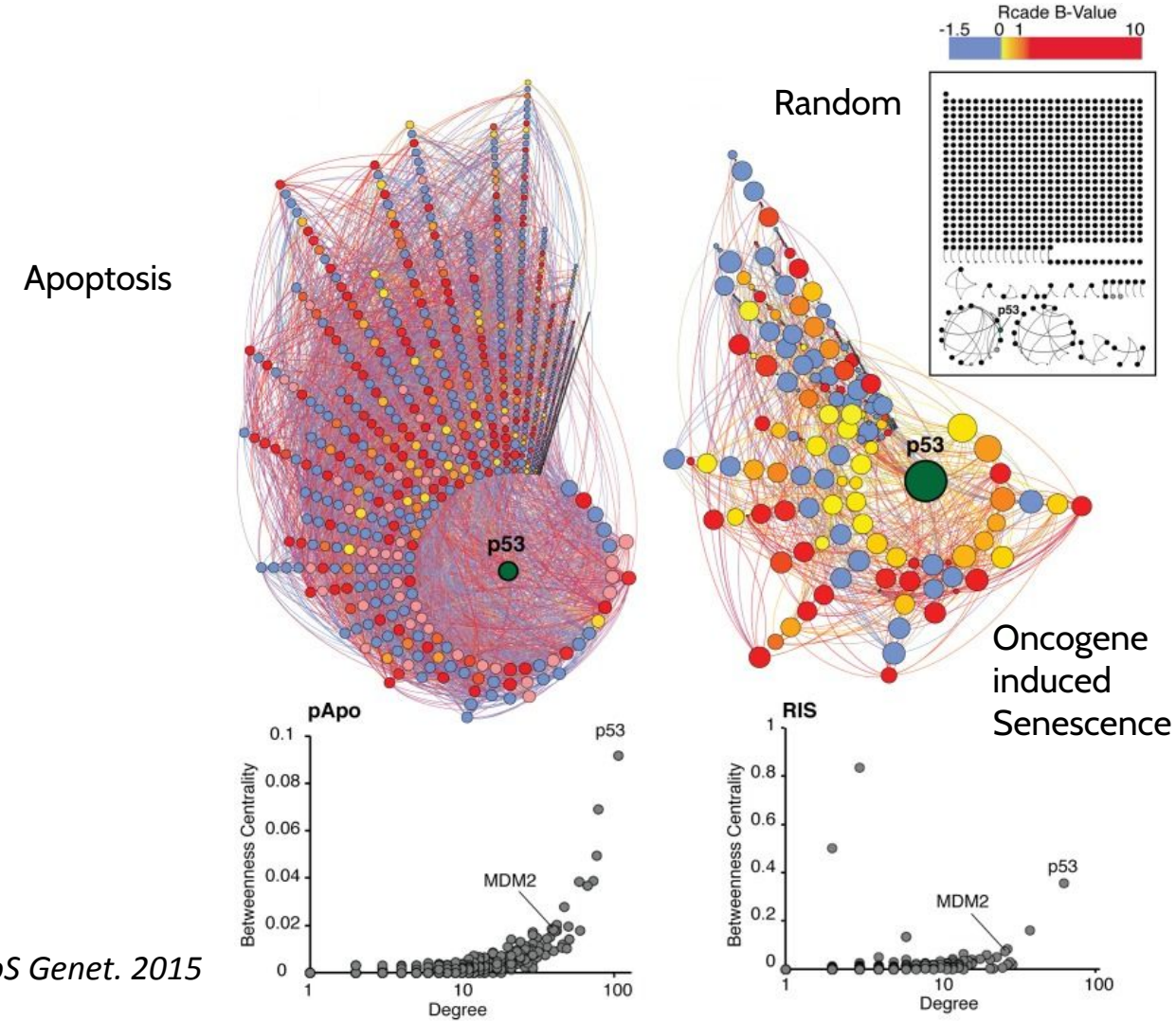
Doncheva et al, J. Proteome Res. 2019

GeneMania

- Web and Cytoscape apps.
- Utilizes connectivity of many external genomic and proteomic datasets.
- Semi-supervised learning algorithm with message passing.



Network topology of p53 regulated genes



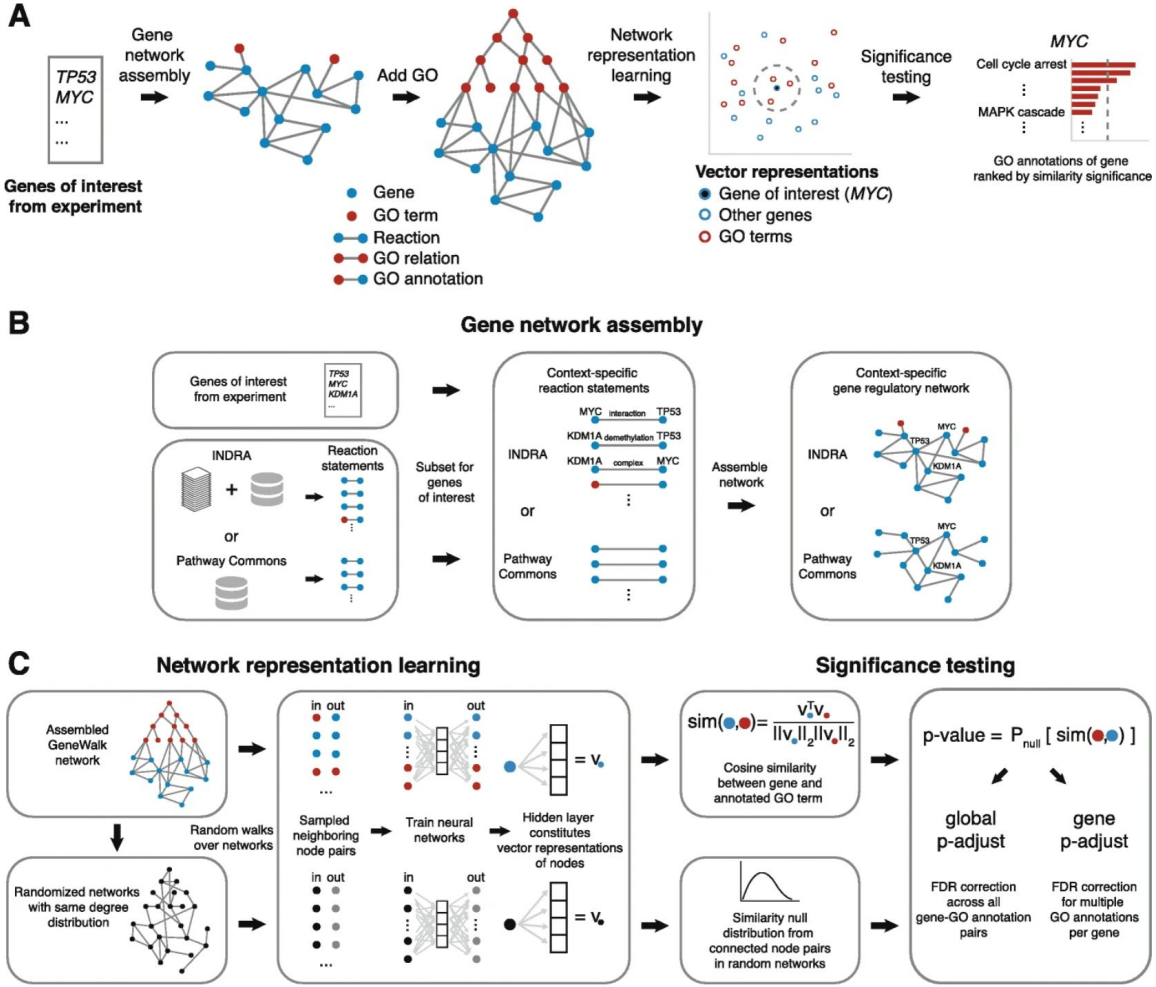
Kirschner & Samarajiwa et al., PLoS Genet. 2015

GeneWalk

GeneWalk uses representation learning to quantify the similarity between vector representations of each gene and its GO annotations, yielding annotation significance scores that reflect the experimental context.

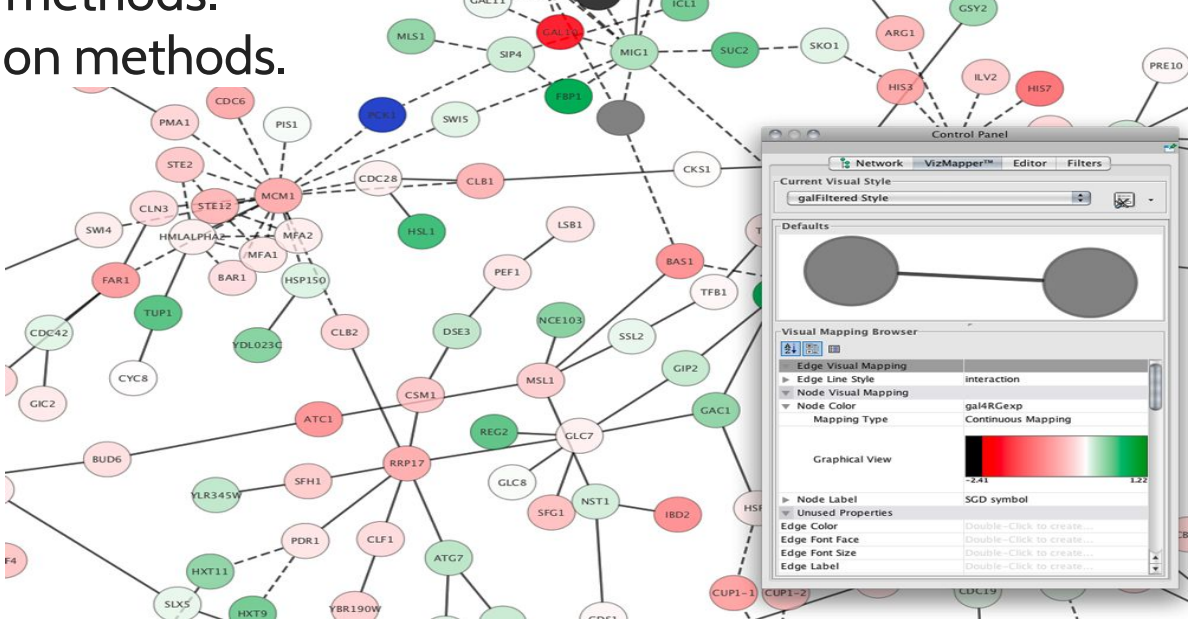
GeneWalk identifies relevant gene functions for a biological context using network representation learning

Robert Ietswaart, Benjamin M. Gyori, John A. Bachman, Peter K. Sorger & L. Stirling Churchman



Cytoscape

- *Cytoscape* is an open source software platform for *analysing* and *visualizing* molecular interaction networks and biological pathways and *integrating* these networks with annotations, gene expression profiles and other state data.
- Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization.
- Large number of network analysis apps and algorithms.
- Network topology analysis methods.
- Many layout and visualization methods.



References

1. Szklarczyk et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan;43(Database issue):D447-52. 1
2. Shannon P et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks Gen. Res. 2003 Nov; 13(11):2498-504
3. Montojo et al., GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics. 2010 Nov 15;26(22):2927-8.