Systems Biology and Networks

MMG 835, SPRING 2016
Eukaryotic Molecular Genetics

George I. Mias

Department of Biochemistry and Molecular Biology
gmias@msu.edu

What is Systems Biology

- **Wikipedia:** ***"Systems biology Systems biology is the computational and mathematical modeling of complex biological systems. An emerging engineering approach applied to biological scientific research, systems biology is a biology-based inter-disciplinary field of study that focuses on complex interactions within biological systems, using a holistic approach (holism instead of the more traditional reductionism) to biological research.***

What is Systems Biology

- **nature.com** : ***"Systems biology is the study of biological systems whose behaviour cannot be reduced to the linear sum of their parts' functions. Systems biology does not necessarily involve large numbers of components or vast datasets, as in genomics or connectomics, but often requires quantitative modelling methods borrowed from physics."***

What is Systems Biology

- **Encyclopedia of Systems Biology (Springer New York, 2013).**

- ***"Systems biology refers to the <u>quantitative</u> analysis of the <u>dynamic</u> interactions among several components of a biological system and aims to understand the behavior of <u>the system as a whole</u>. Systems biology involves the development and application of systems theory concepts for the study of complex biological systems through iteration over <u>mathematical modeling, computational simulation and biological experimentation.</u> Systems biology could be viewed as a tool to increase our understanding of biological systems, to develop more directed experiments, and to allow accurate predictions."***

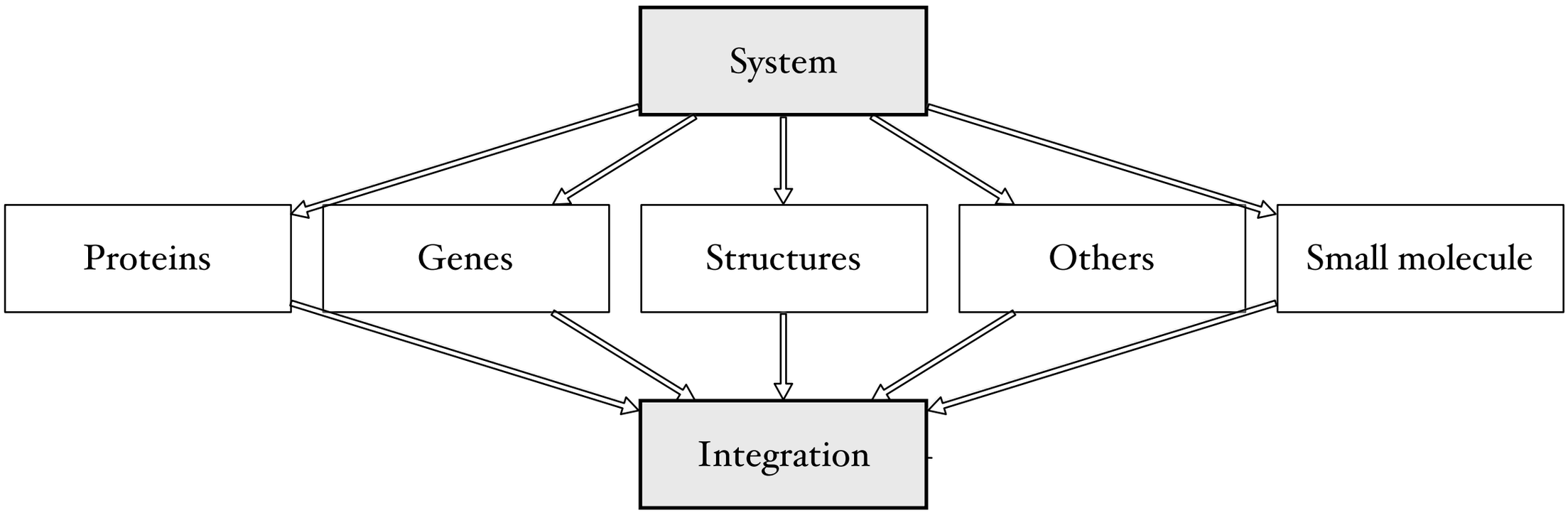
Systems Biology

- Systematic
- Novel approach (in biology at least)
- interdisciplinary
- Non-reductionist
    - Reductionist: Study the subcomponents for a system in detail, each one
    - Whole is greater the sum of its parts

Systems Biology

- Multiple inputs of information in a complex system
- More mathematical than traditional biology

Systems Biology

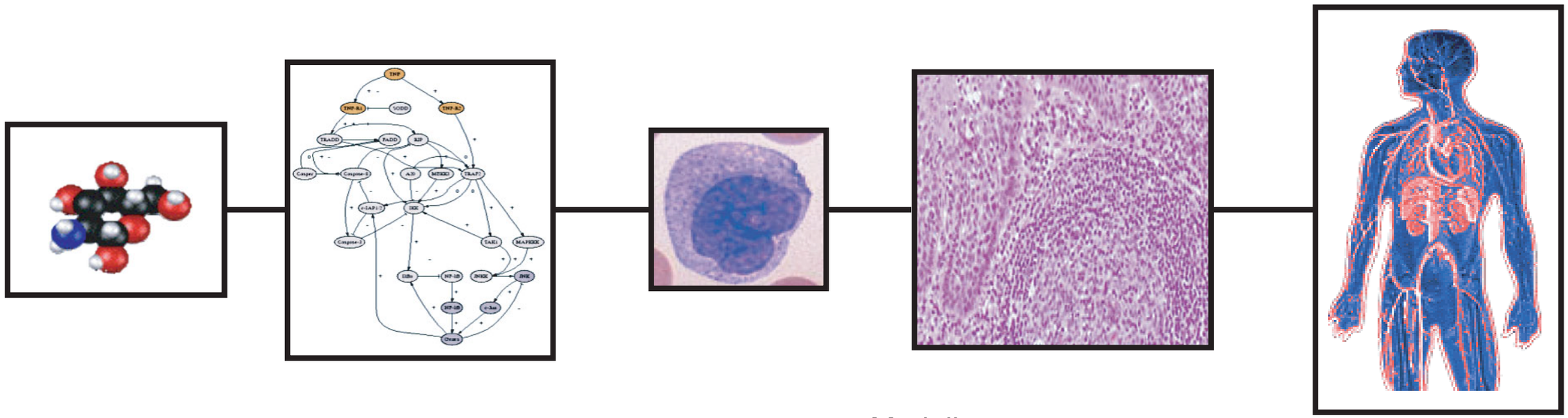


Systems Biology

- Molecular components
- Cell subsystems
- parts of an organism
- the organism
- the environment
- set of environments

- **Study of**
  - Function
  - Networks
  - Signals
  - Interactions of components

Systems Biology



Butcher et al., Nature Biotechnology 22(10), p1253 (2004)

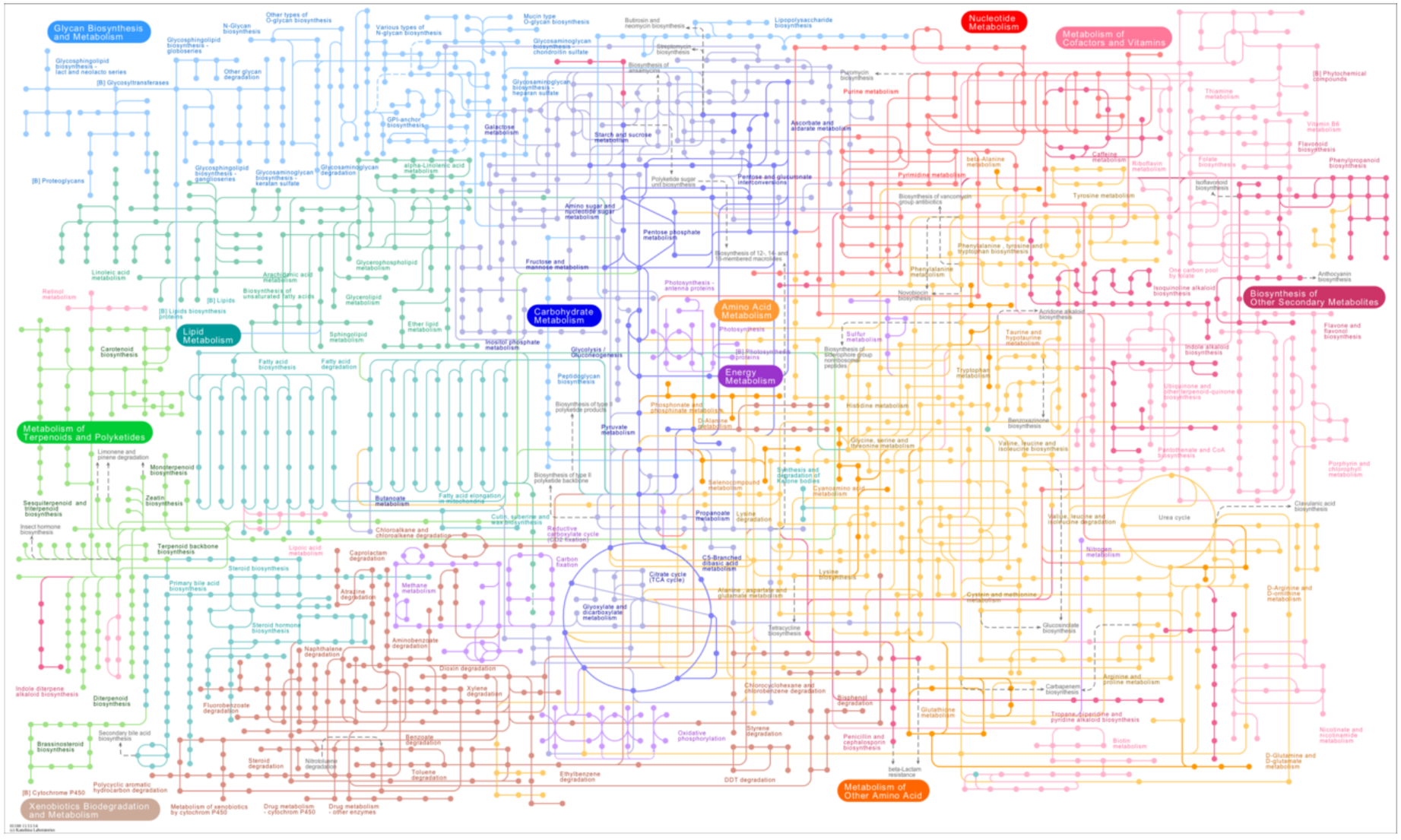
Systems Biology

- Examples

- Omics approaches
  - Human Genome Project
  - Mass spectrometry
    - Proteomics
    - Metabolomics
  - name-it-omics

Systems Biology

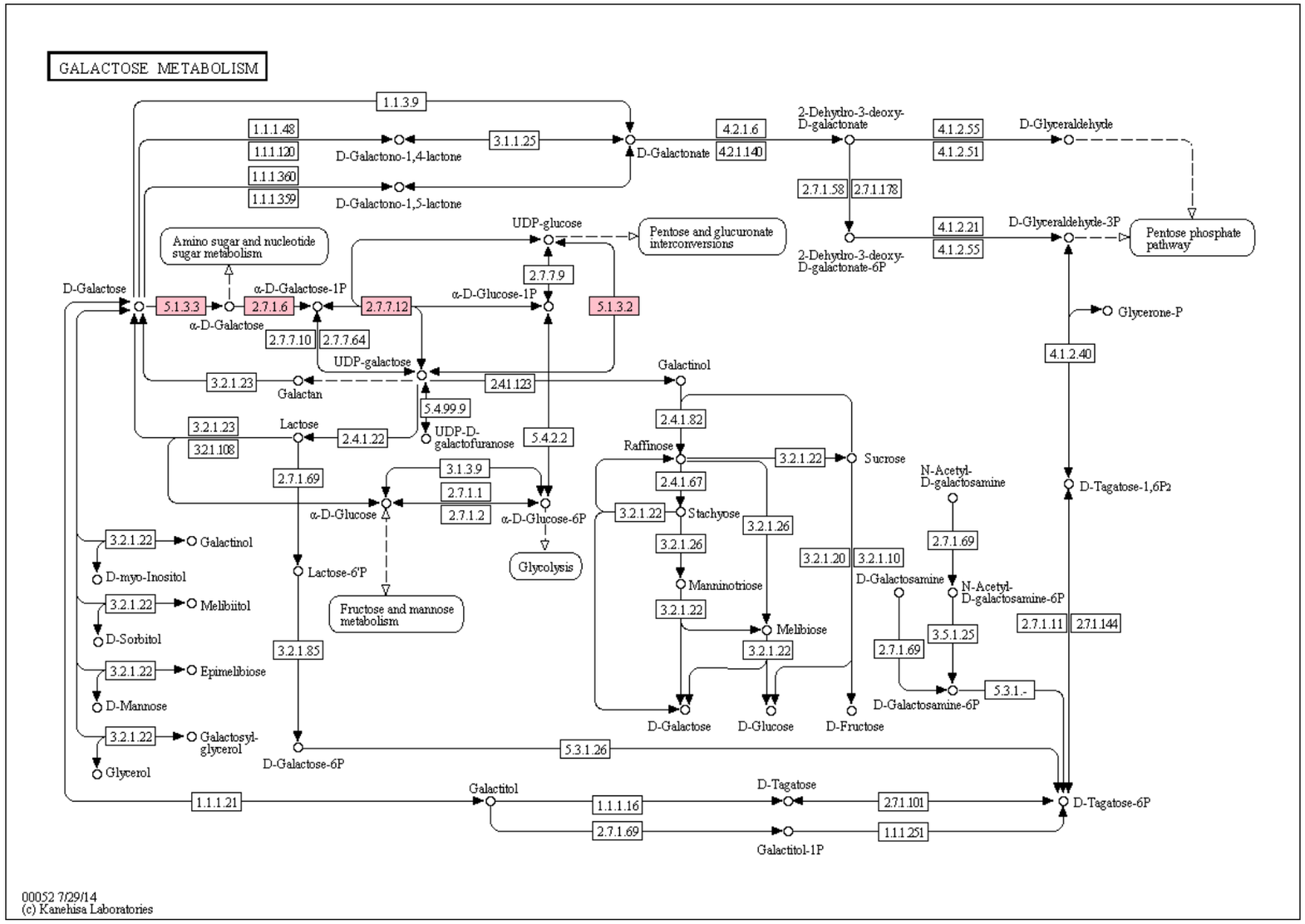
- Example: Metabolism



http://www.genome.jp/kegg/pathway/map/map01100.html (11/18/2014)
KEGG: Kyoto Encyclopedia of Genes and Genomes

Systems Biology

- Example: Metabolism: Galactose Metabolism (degradation)

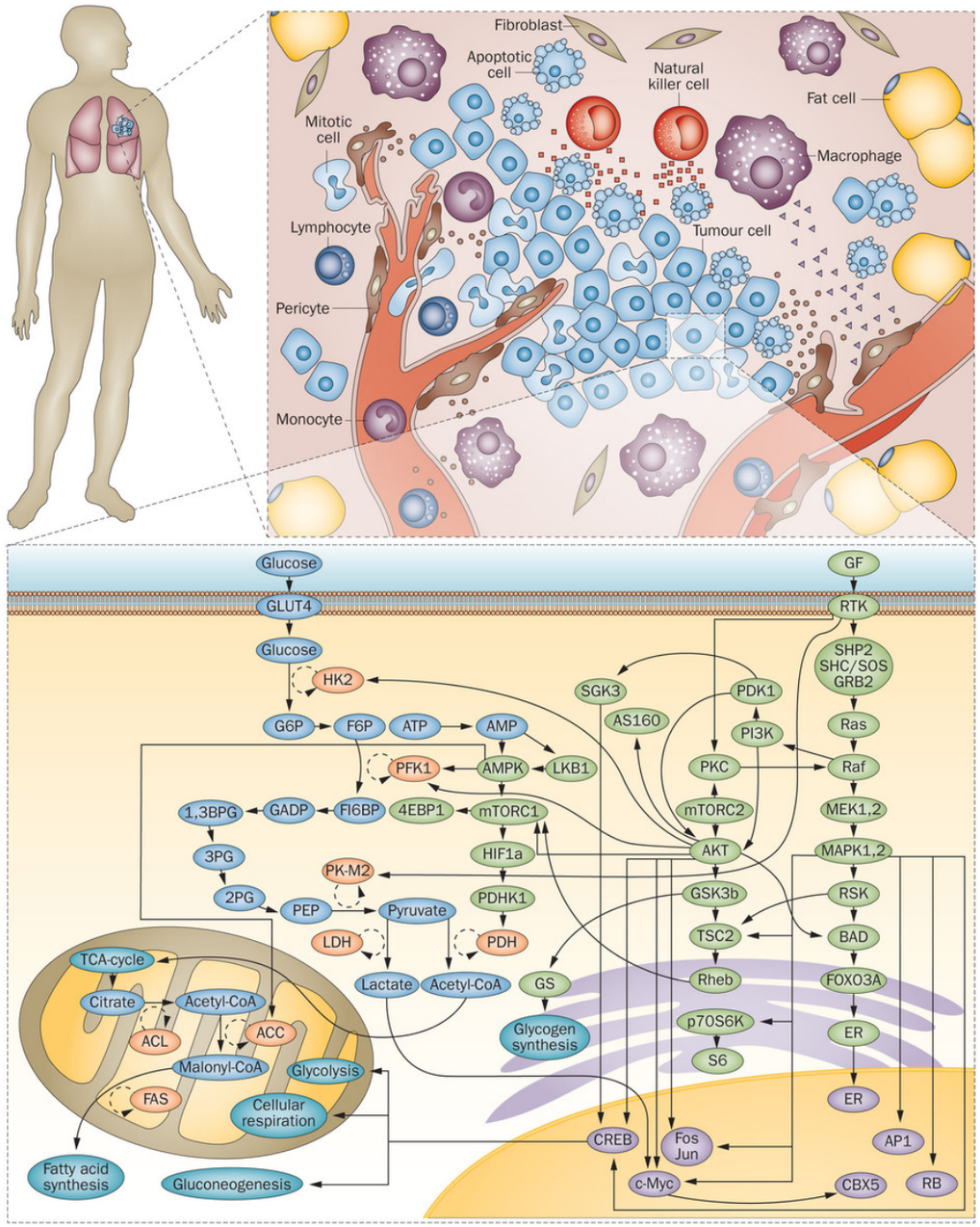


http://www.genome.jp/kegg/pathway/map/map01100.html (11/18/2014)

KEGG: Kyoto Encyclopedia of Genes and Genomes

Systems Biology

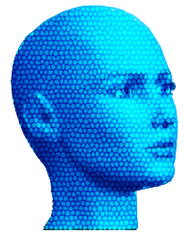
- Example: Cancer
  - Different Networks
  - Homeostasis
  - Molecular components



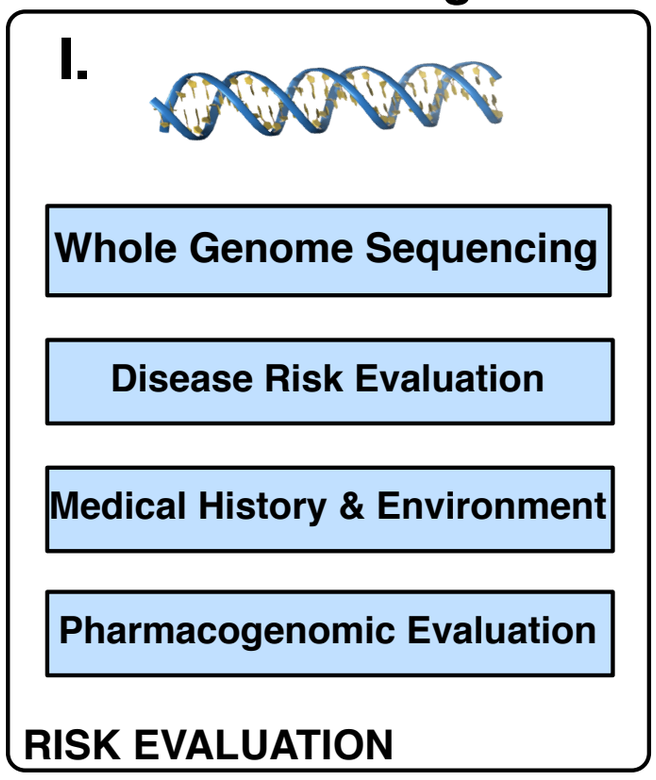
Werner, H. M. J. *et al.* (2014) *Nat. Rev. Clin. Oncol.* doi:10.1038/nrclinonc.2014.6

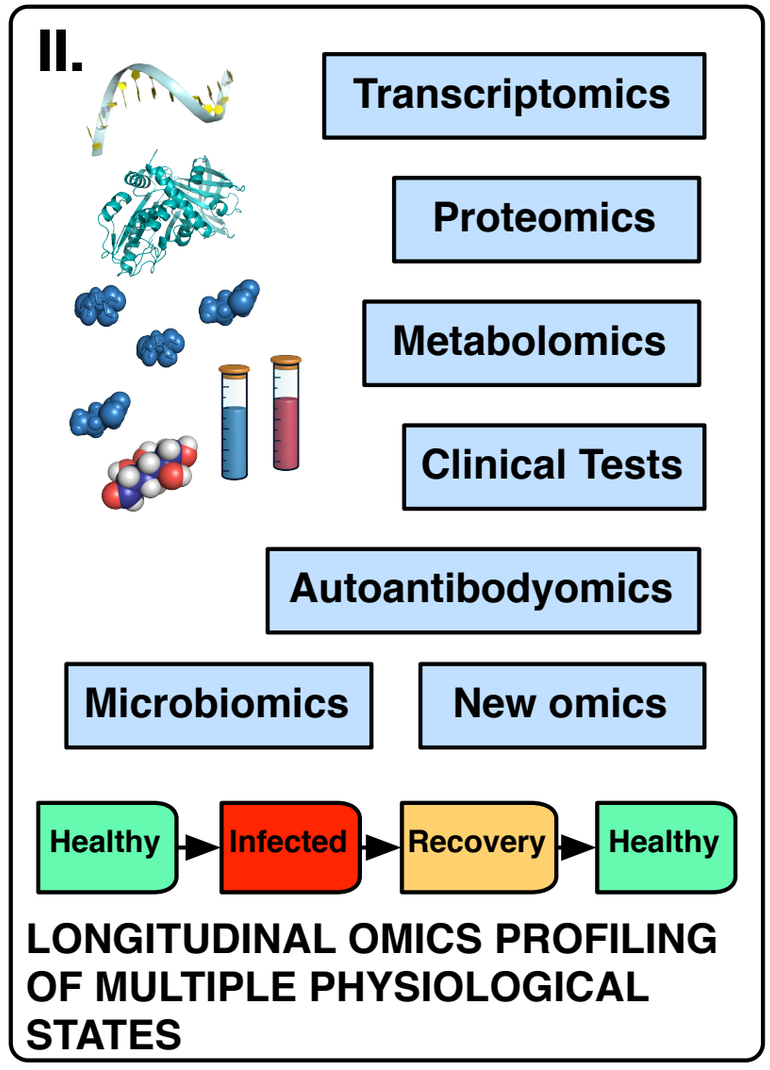
Systems Medicine

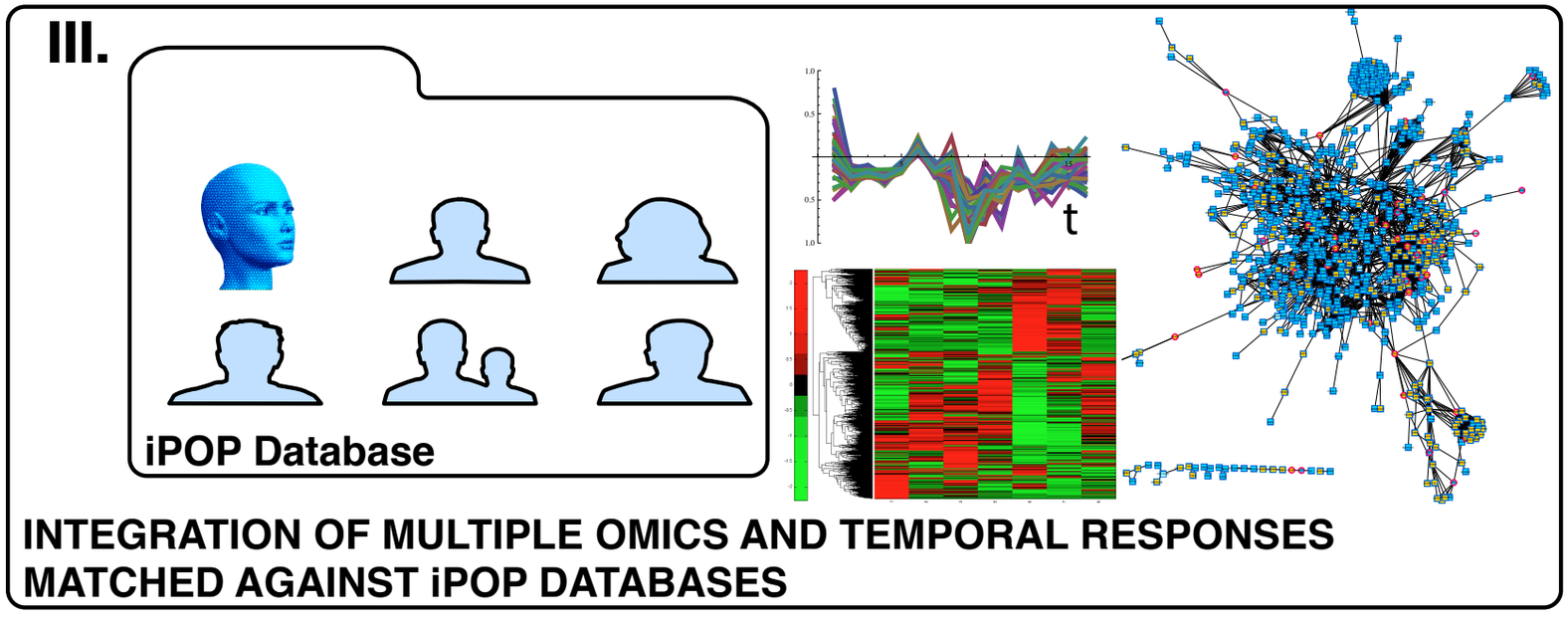
- Personalized
  - Determine risks
  - Monitor
  - Integrate



integrative
Personal
Omics
Profiling







Mias and Snyder, Quantitative Biology 1(1) p. 71 (2013)
Chen*, Mias*, Li-Pook-Than*, Jiang* et al Cell 148,1293 (2012)

http://goo.gl/iamZth

Systems Biology

- Experiments
- Models
- data
- theory
- computation

- **Nobert Wiener (1894-1964)**
  - **cybernetics and systems controls)**
- **20th century biochemistry**
- **21st century Leroy Hood and others**

Systems Biology

- Experiments

- **BIG**

DATA!

Reformulate biological problems in terms of mathematical models.

Models need computational approaches

Big data handling in storing, retrieving useful information and relaying/displaying this information

Data - Omics

Molecular Components

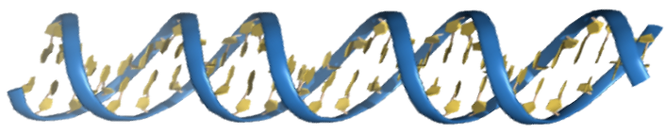
Nucleic acids
Proteins
Lipids
Carbohydrates

Genomics

Illumina

DNA VARIANTS

A

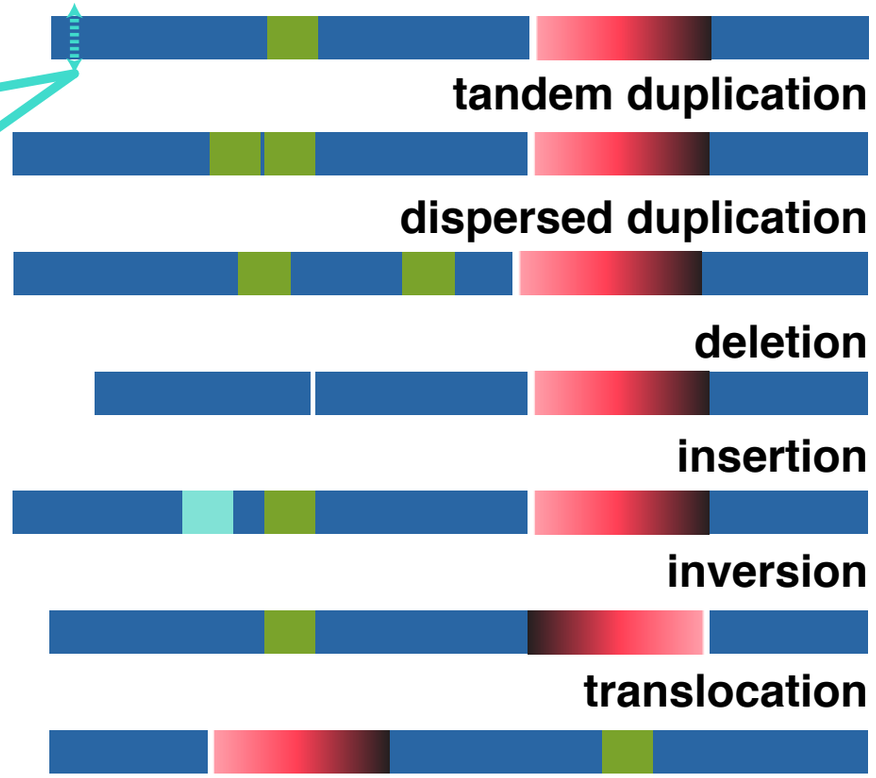


bp Level Variation

reference	*ggcttccaggaactc*
point mutation	ggcttccag**a**aactc *ggcttccaggaactc*
insertion	ggcttccagg**g**aactc *ggcttccagg aactc*
deletion	ggcttccaggactc *ggcttccagg aactc*

MinION

Structural Variation [>1000 bp]



Ion Torrent

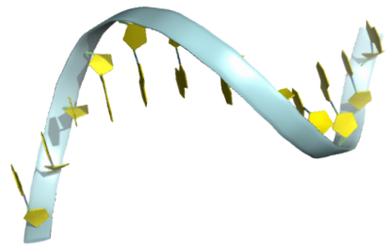
SOLiD

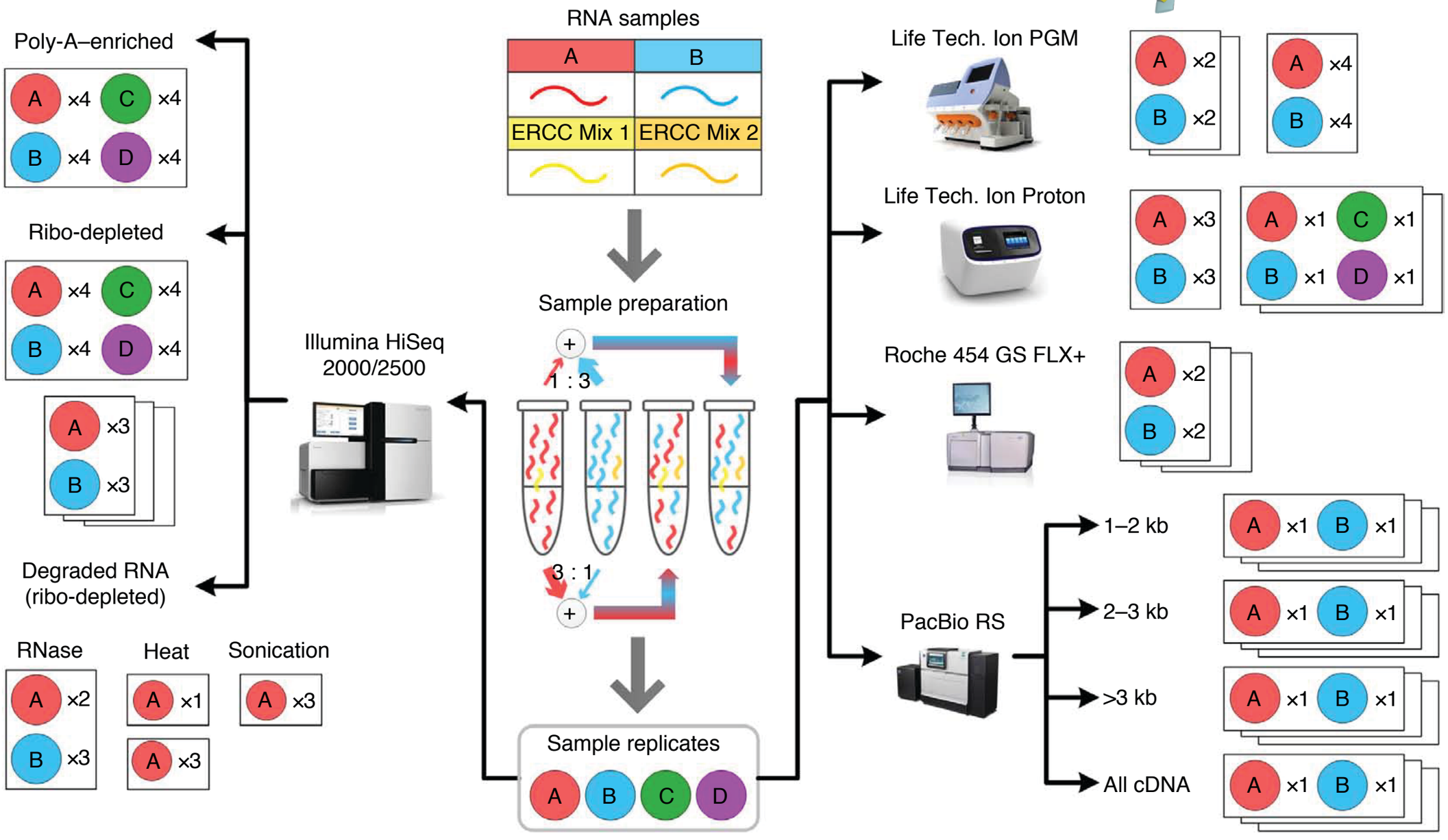
PacBio

454

Mias and Snyder, Quantitative Biology 1(1) p. 71 (2013)

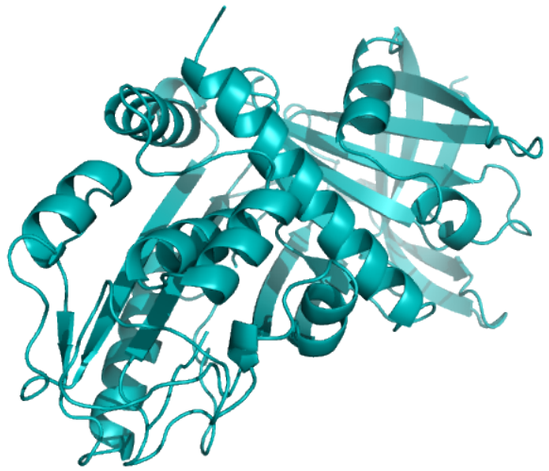
Transcriptomics

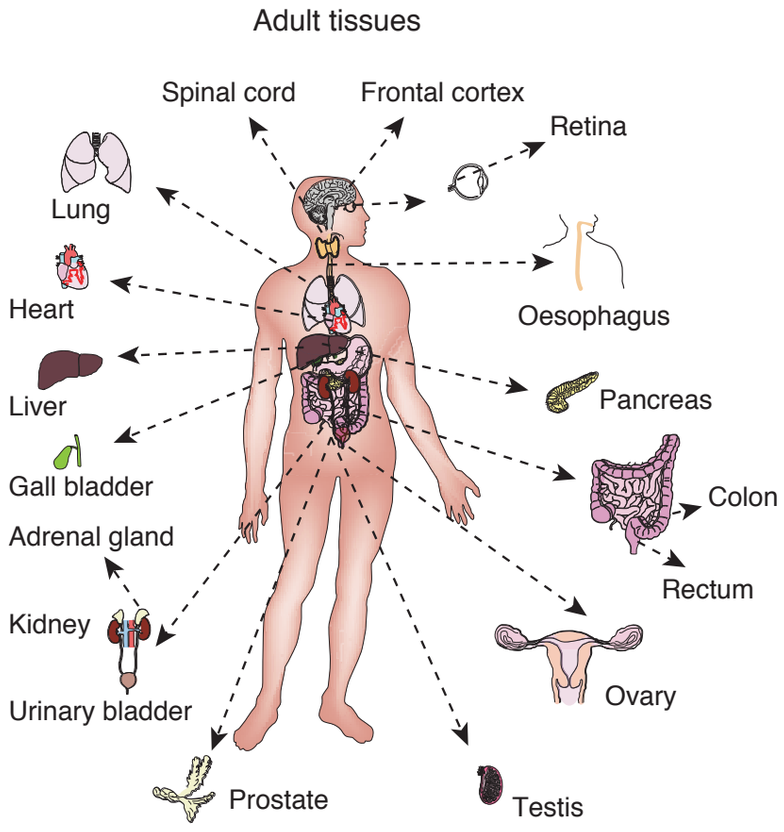


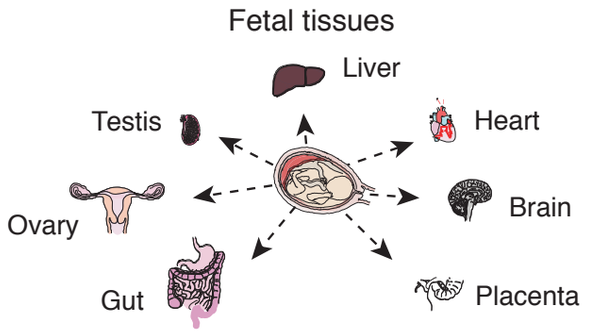


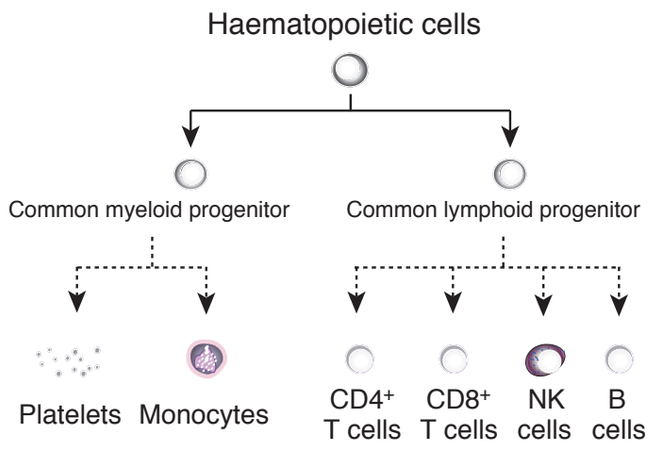
Li, Tighe et al., Nature Biotechnology 12(9), p. 915 (2014)

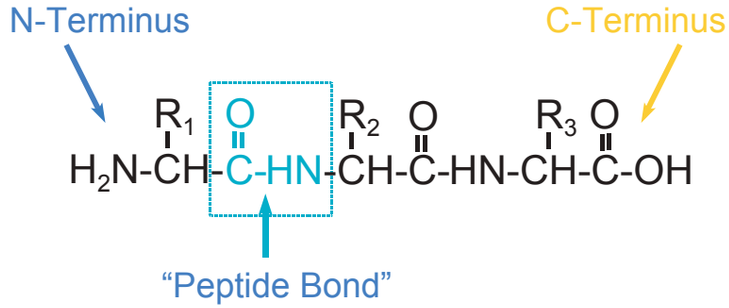
Proteomics

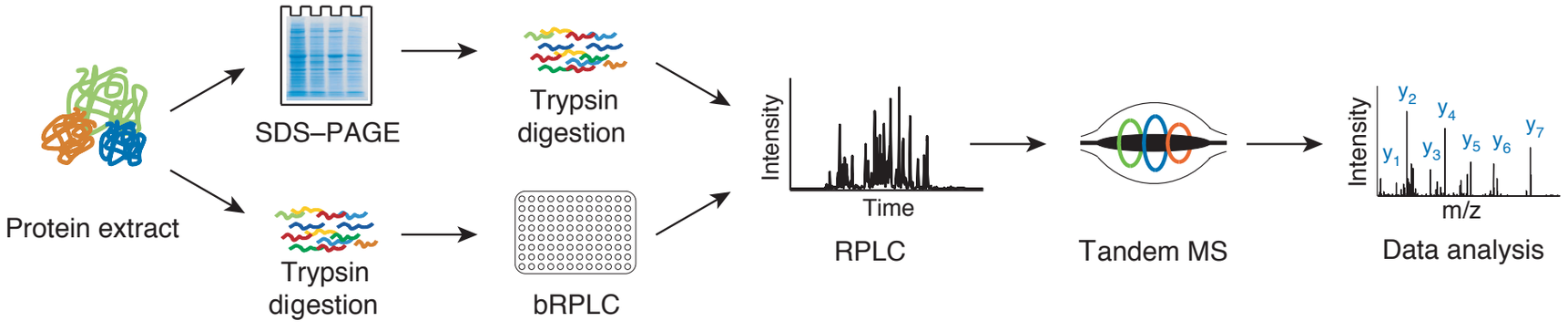












A draft map of the human proteome

Kim et al., Nature 509, p. 575 2014

Metabolomics

thermofisher.com

NMR

Other varieties

agilent.com

all metabolites in cells

- small molecules
- lipids
- peptides
- amino acids
- nucleic acids
- organic acids

Lipidomics

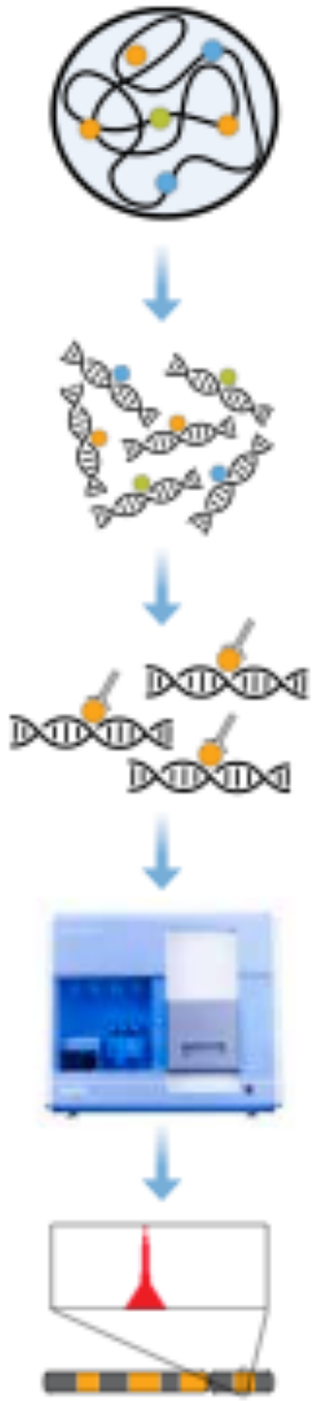
- lipids
- metabolic signaling
- energy storage
- cell proliferation
- cell migration
- apoptosis
- cellular membrane

Interactions

Whole-Genome Chromatin IP Sequencing

(ChIP-Seq)

identify binding sites of DNA-associated proteins



Nucleus

Crosslink and Fractionate Chromatin

ChIP: Enriched DNA Binding Sites

Sequence

Binding Site Mapping

http://res.illumina.com/documents/products/datasheets/datasheet_chip_sequence.pdf

Interactions

Protein Arrays
Peptide Array
Bead Methods
Two Hybrid Methods
Many more!

Practical Issues

Not quantitative enough
Expensive enough
Inconclusive
Not exhaustive
Not dynamic

Systems Biology

- **Applications**
  - Genotype to Phenotype
  - In-silico Cell
  - Physiological Models
  - Personalized Precise and Predictive Medicine

Modeling

What's new?

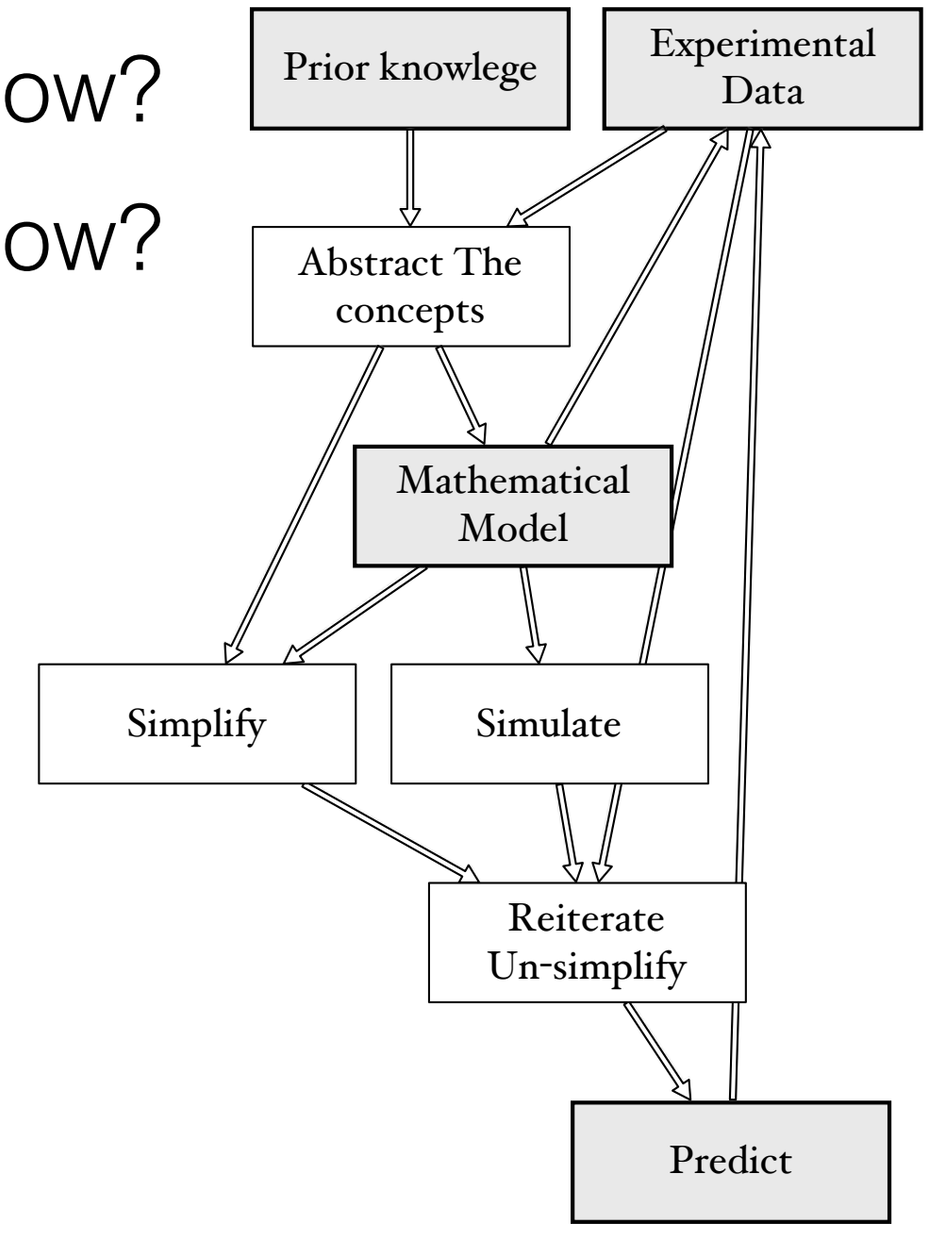
- Technological Advancements (e.g. mass spectrometry/sequencing/ imaging).
- High Performance Computing
- Information Storage abilities
- Information sharing abilities

How to approach it

What is the question?

What do we know?

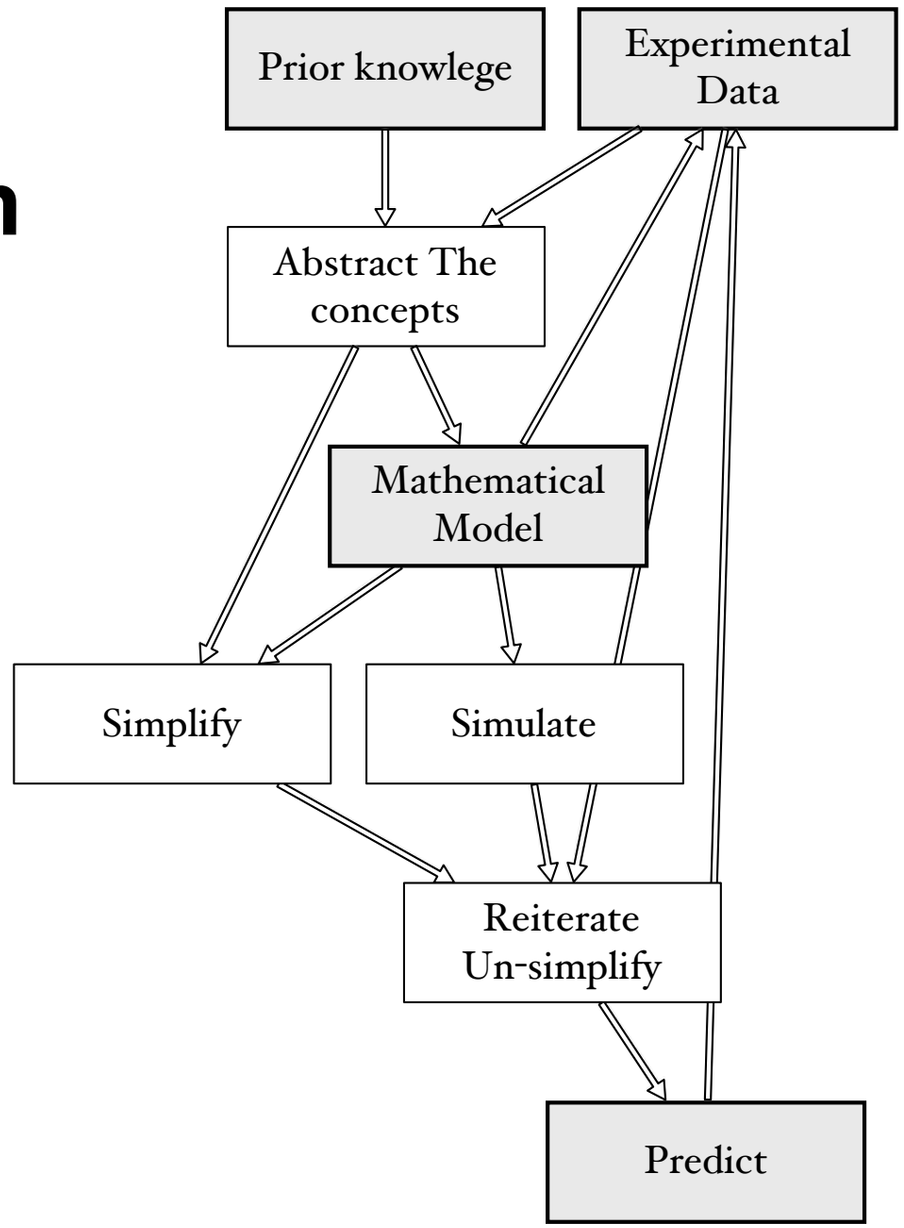
What can we know?



How to approach it

Underlying system

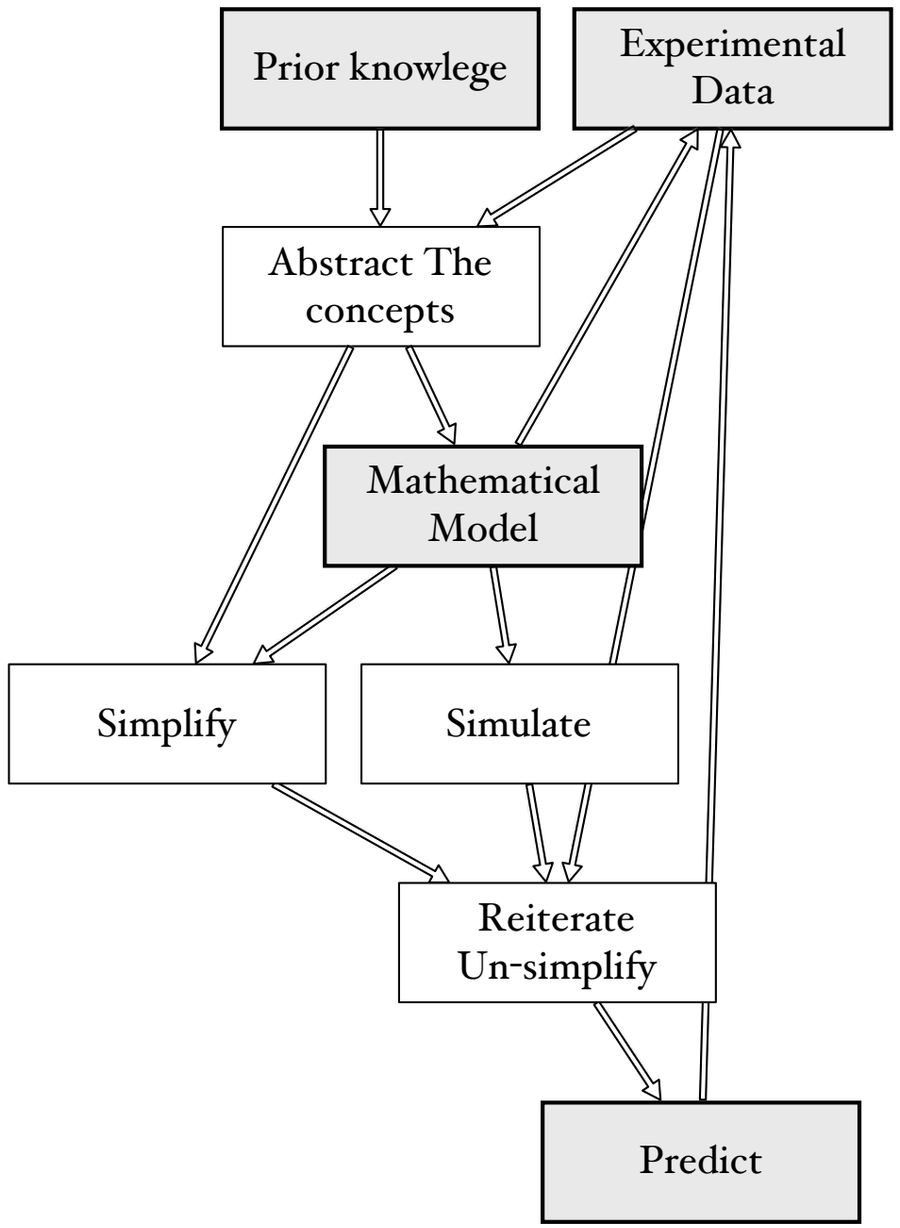
- characteristics
- responses
- interactions



How to approach it

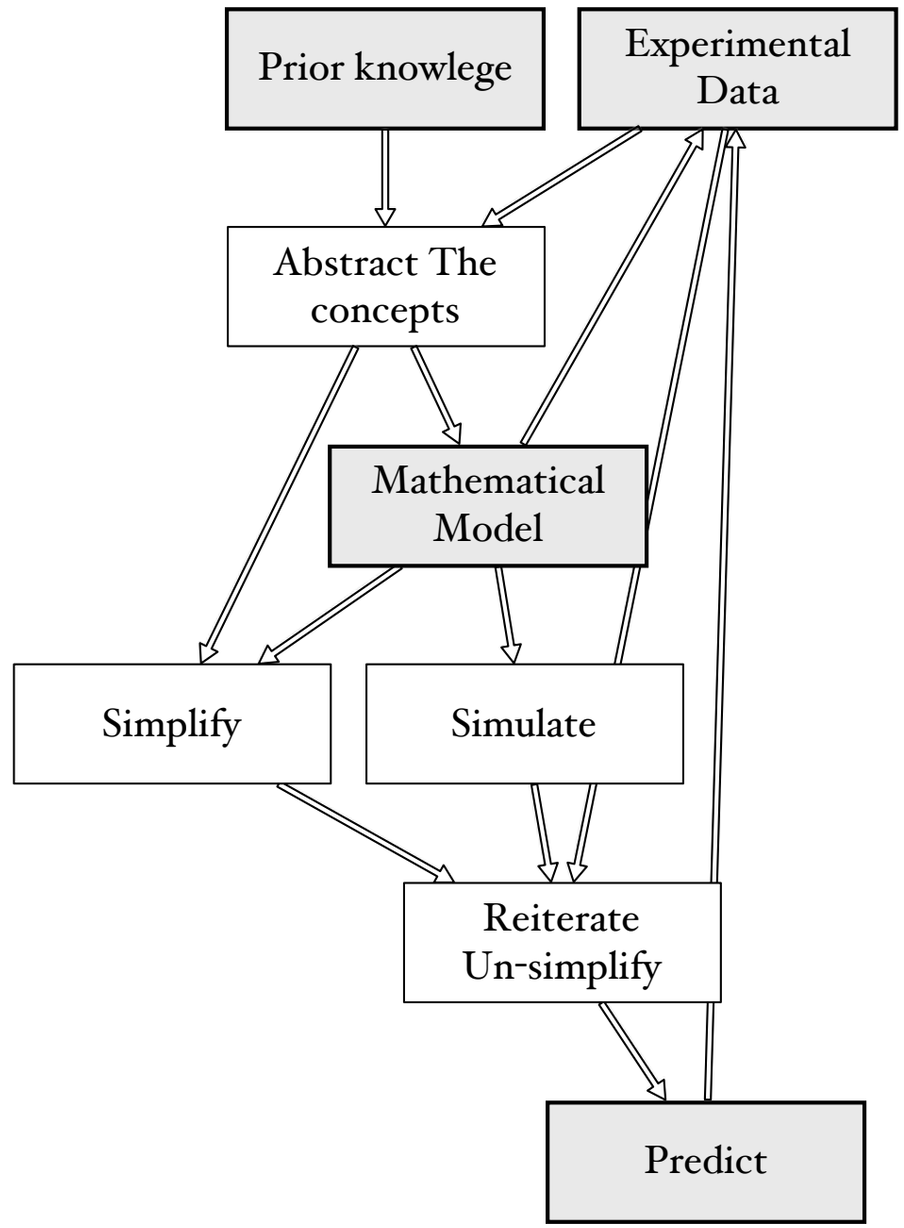
Pick your model (Based on system)

- atomic level
- cells
- molecular components
- dynamic Vs. Static
- How much to coarse grain?
- How many parameters?



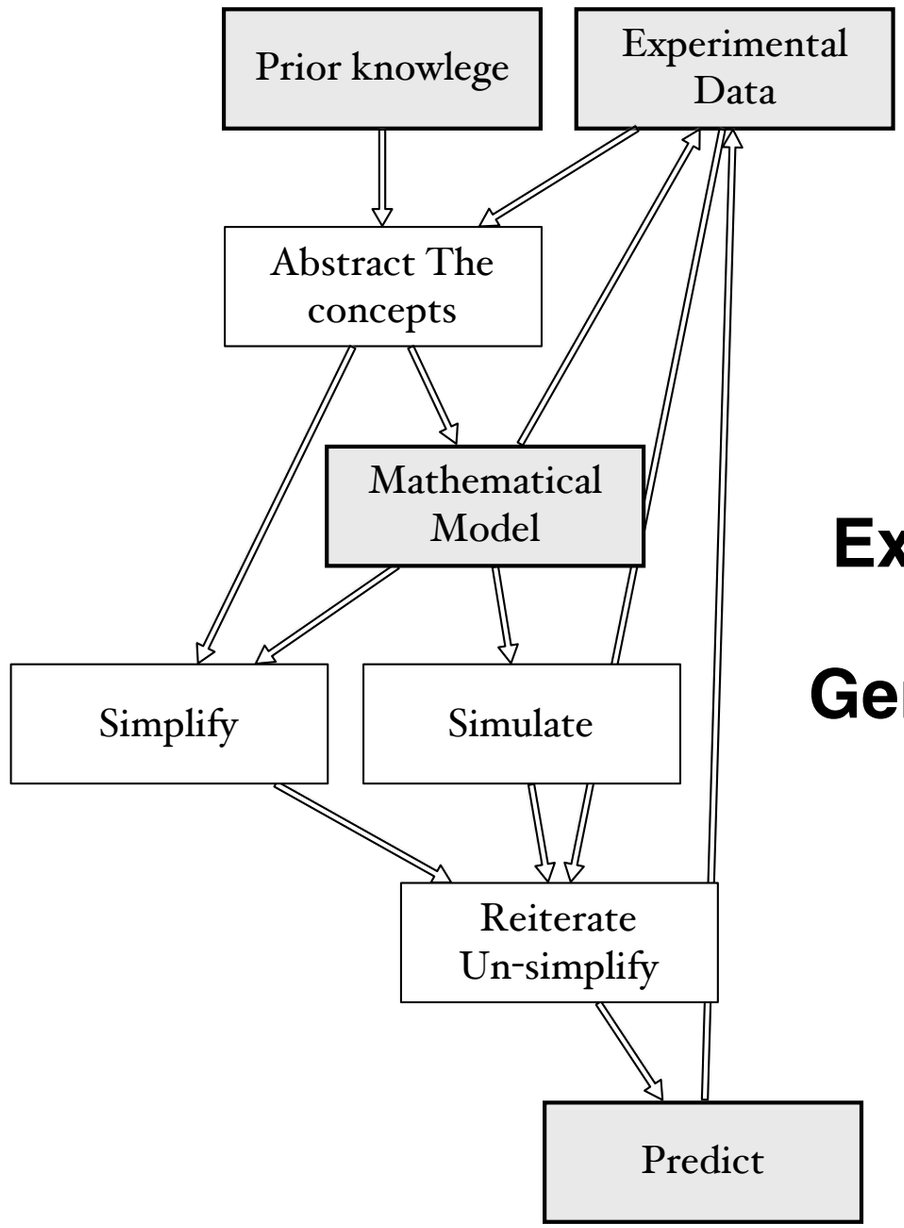
How to approach it

- Interactions and Networks
- Assumptions
- New knowledge required
- Storage of information



How to approach it

- Interactions and Networks
- Assumptions
- New knowledge required
- Storage of information

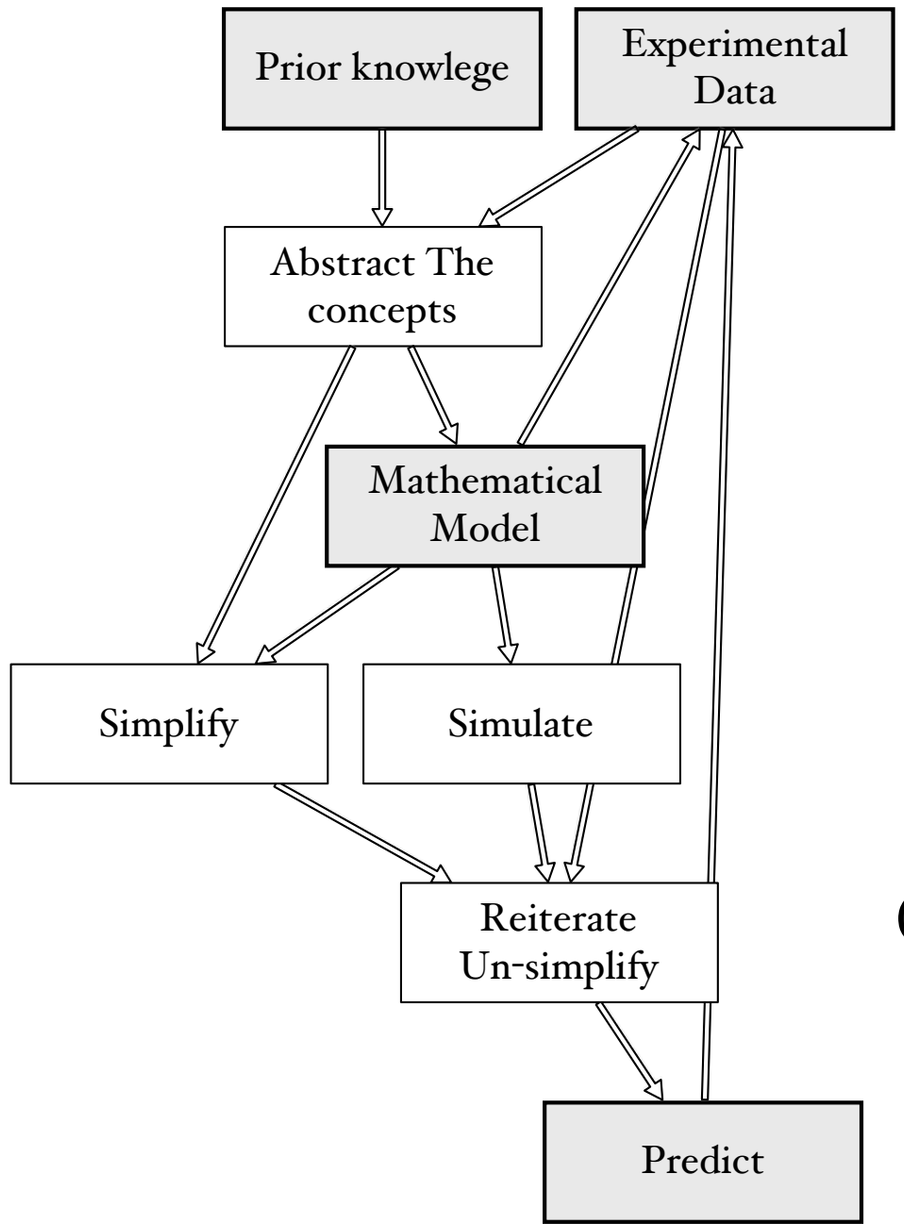


Experimental Tests

Generate Hypotheses

How to approach it

- Interactions and Networks
- Assumptions
- New knowledge required
- Storage of information



Validation

may invalidate

cannot confirm

Biochemical Models

Biochemical Models

Chemical Reactions

A transforms to B

$$A \rightarrow B$$

- conversion
- modification
- dimerization

A associates with B to form C

$$A + B \rightarrow C$$

- association
- synthesis

C dissociates to A and B

$$C \rightarrow A + B$$

- dissociation
- decomposition

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Chemical Reactions

null species (e.g. constant abundance)

$A \rightarrow \varnothing$

- degradation

$\varnothing \rightarrow B$

- production
- discarded reactants

Elementary Irreversible Reactions

- Single step
- one direction

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Chemical Reactions

Can put it all together: e.g.

$$A + B \rightleftharpoons AB \rightarrow C$$

- 2 elementary reactions
- C covalent modification of AB

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Chemical Reactions

A transforms to B

$$A \rightarrow B$$

$$\text{Reaction Rate} = k\,[\text{A}]$$

mass action

Rate of changes proportional to concentration

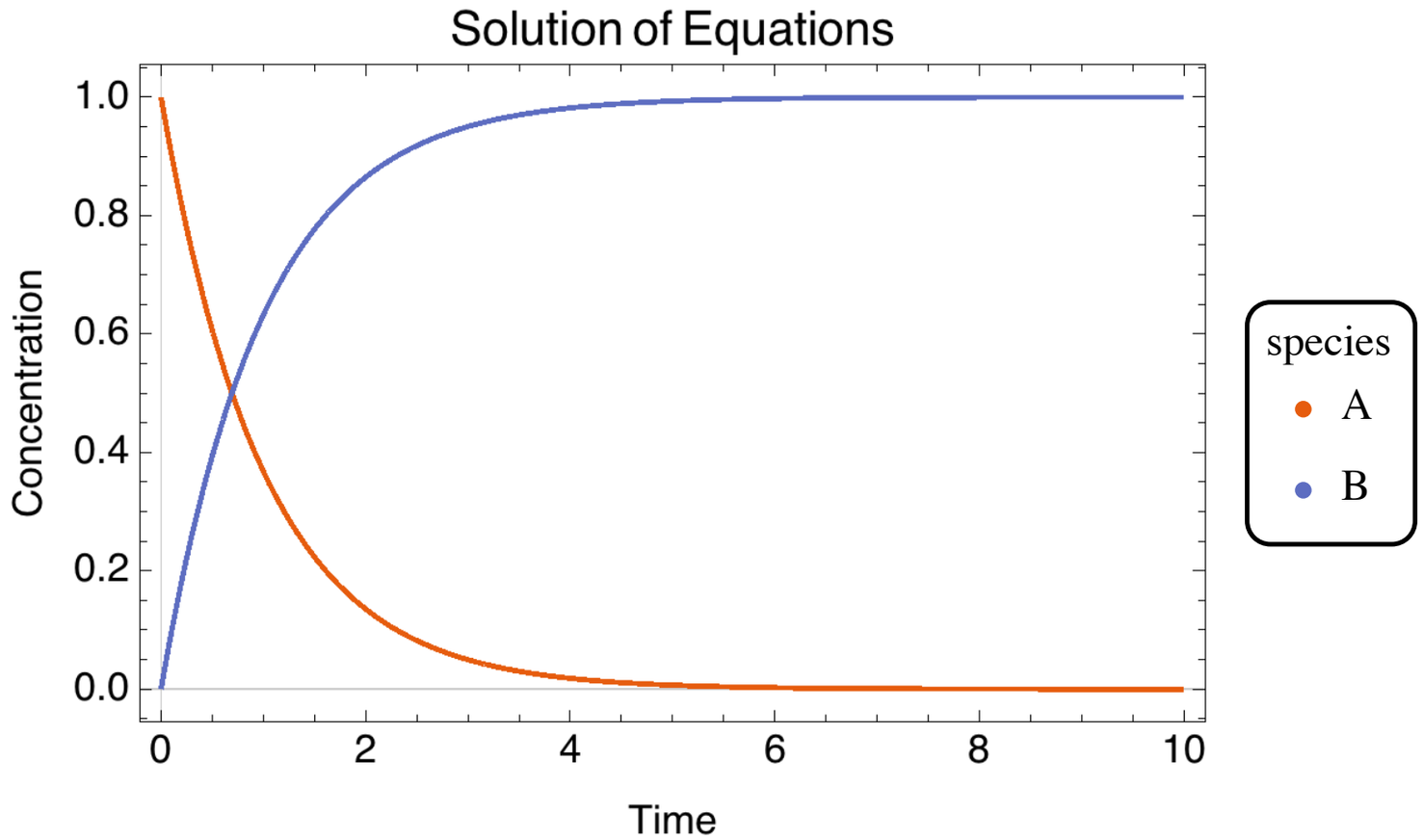
$$\frac{d[A]}{dt} = -k(A); \frac{d[B]}{dt} = k(A)$$

Biochemical Models

Chemical Reactions

A transforms to B

$$\frac{d[A]}{dt} = -k(A); \frac{d[B]}{dt} = k(A)$$



Rate of changes proportional to concentration

Biochemical Models

Chemical Reactions

Chemical Reaction	Rate Equation for One Species
$A + B \rightarrow C$	$\frac{dc}{dt} = k_f ab = k_f(a_0 - c)(b_0 - c)$
$2A \rightleftharpoons B$	$\frac{db}{dt} = k_f(a_0 - 2b)^2 - k_r b$
$A \rightleftharpoons 2B$	$\frac{db}{dt} = k_f(a_0 - \frac{b}{2}) - k_r b^2$
$A \rightleftharpoons B + C$	$\frac{dc}{dt} = k_f(a_0 - c) - k_r(b_0 + c)(c_0 + c)$
$A + B \rightleftharpoons C$	$\frac{dc}{dt} = k_f(a_0 - c)(b_0 - c) - k_r c$
$A + B \rightleftharpoons C + D$	$\frac{db}{dt} = k_f(a_0 - c)(b_0 - c) - k_r(c_0 + c)(d_0 + c)$

x_0 : initial concentration for x

Lynch, Dynamical Systems with Applications using Mathematica, Birkhauser (2007)

Biochemical Models

Chemical Reactions

$$X_i(t) = \frac{N_i(t)}{\Omega}$$

X_i : ith species concentration
Ω : system size $= N_A \times$ Volume
N_i : copy number
N_A : Avogadro's constant

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Chemical Reactions

Reaction scheme R_j

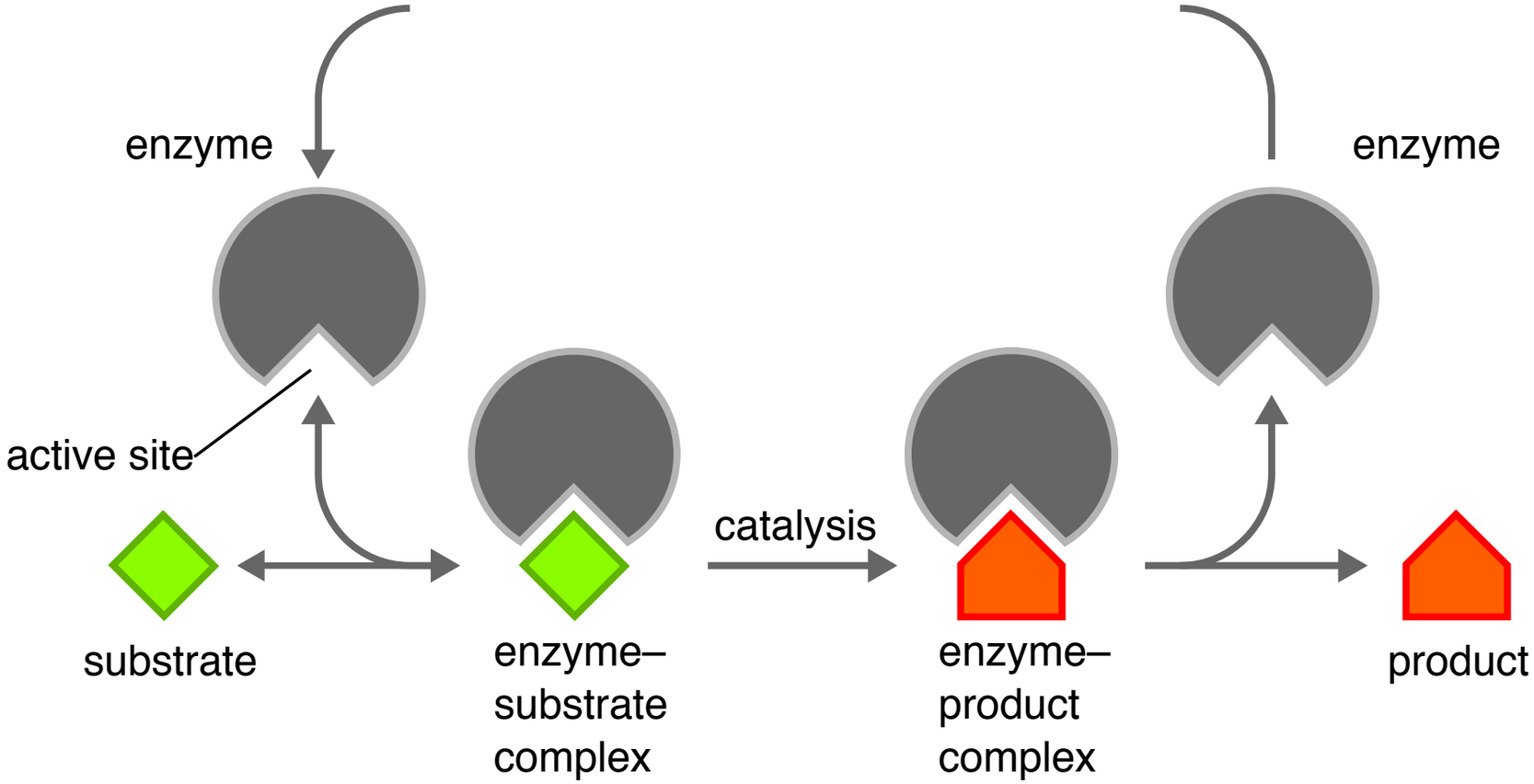
$$\underline{S}_{1j} X_1 + ... + \underline{S}_{sj} X_s \rightarrow \bar{S}_{1j} X_1 + ... + \bar{S}_{sj} X_s$$

S_{ij} stoichiometric coefficient indicates participation of X_i as a reactant

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Chemical Reactions

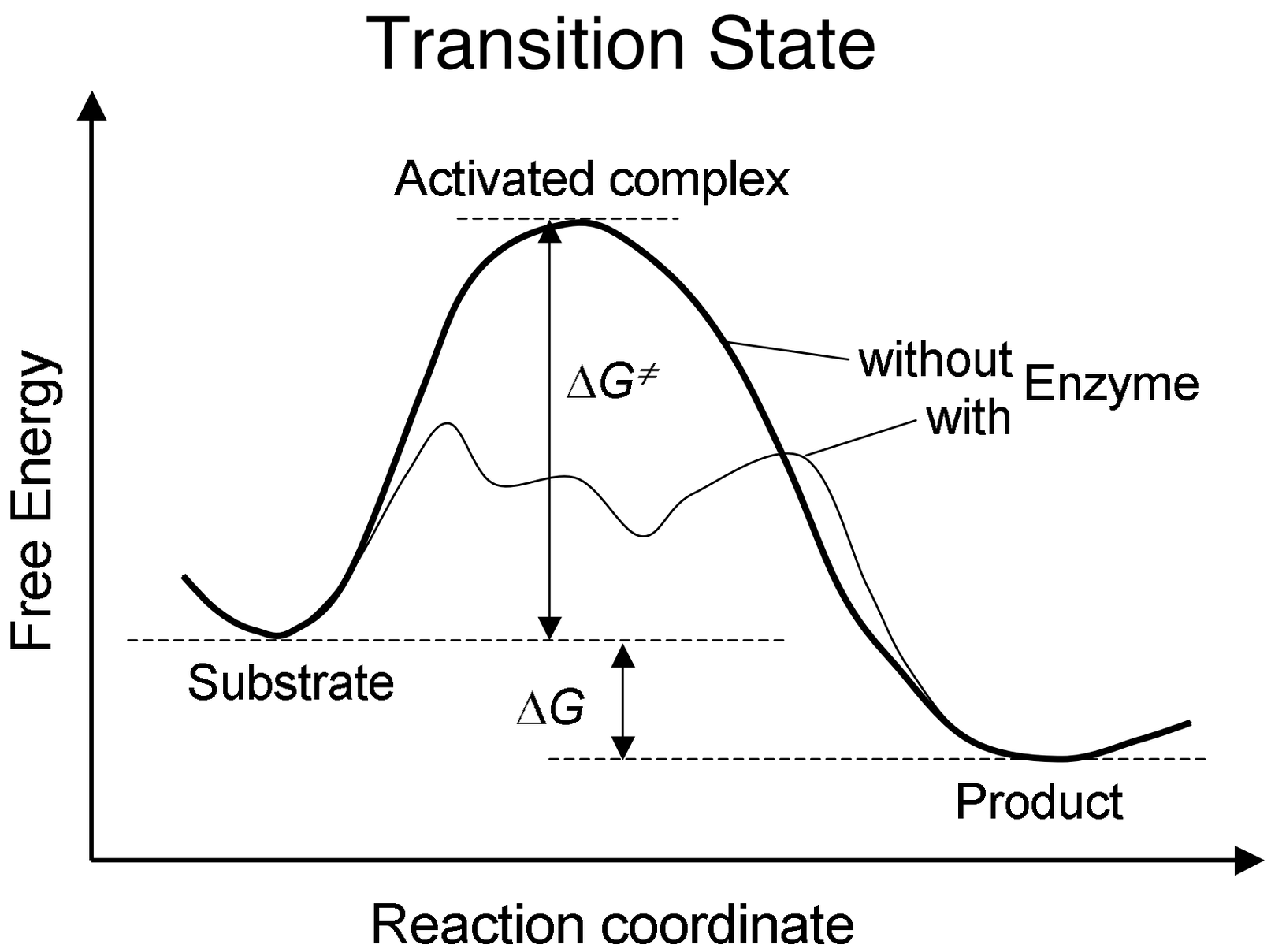


energy

Enzyme Catalyzed Conversion of Substrate to Product

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models



Klipp, Systems Biology, Wiley (2010)

Biochemical Models

Chemical Reactions

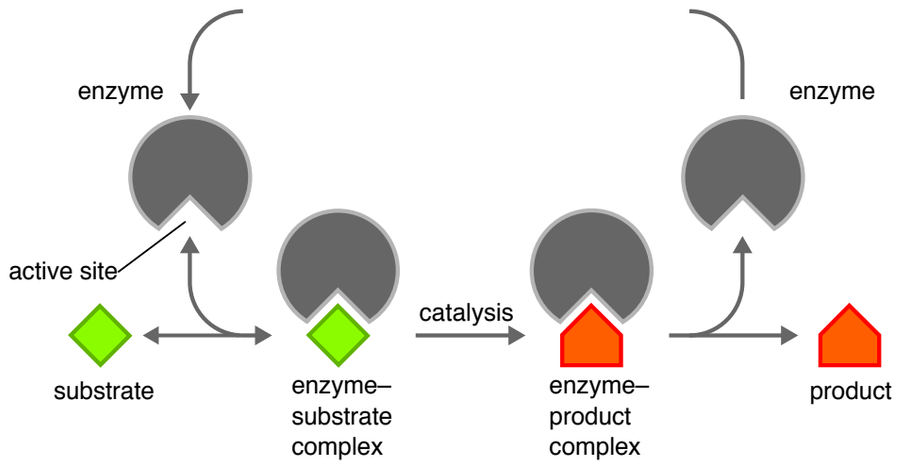
$$E + S \rightarrow E + P$$

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \rightarrow E + P$$

- 3 elementary reactions

$$S \xrightarrow{k_{eff}} P$$

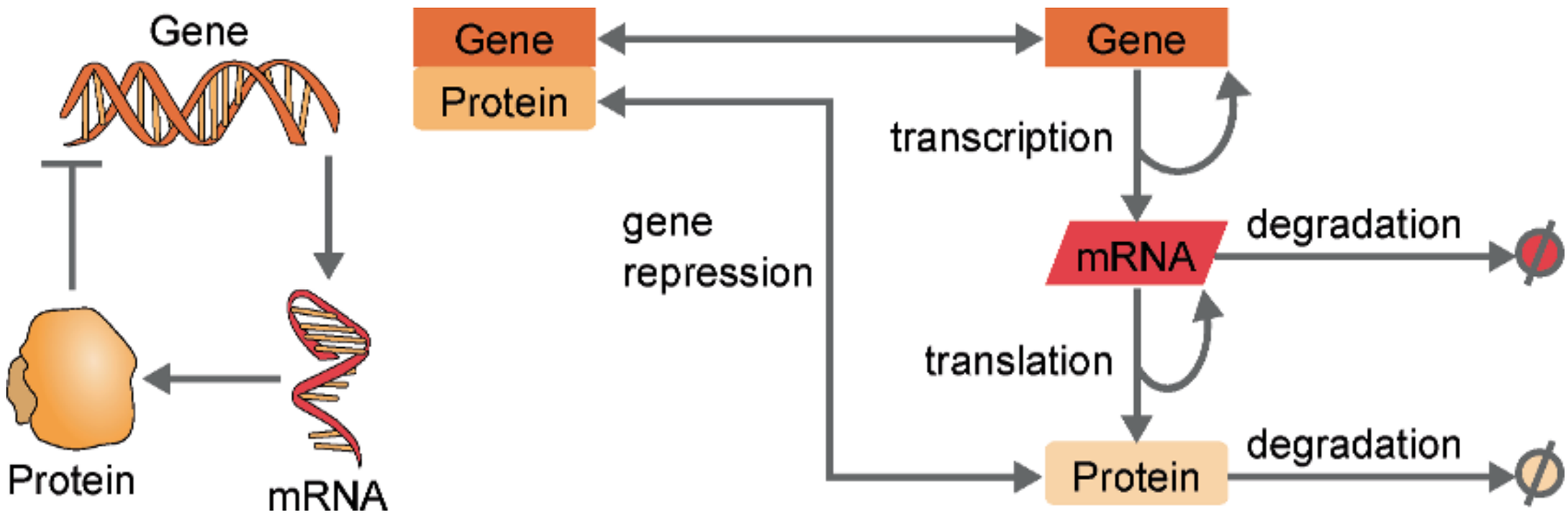
- Can approximate to single step



M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation



M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

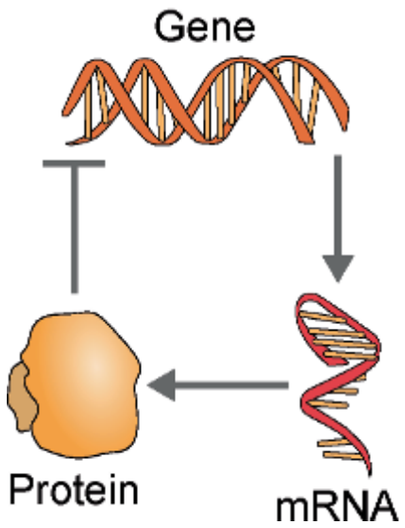
Simplified Gene Regulation

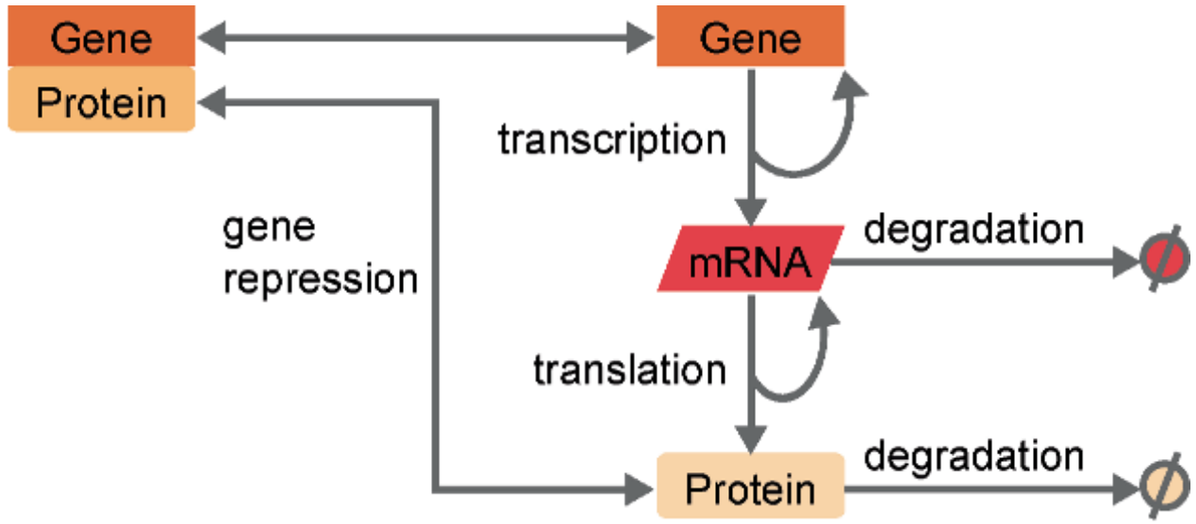
- Transcription

$$G \xrightarrow{k_m} G + M$$

- Translation

$$M \xrightarrow{k_p} M + P$$





- Binding/unbinding

$$G + P \underset{k_u}{\overset{k_b}{\rightleftharpoons}} GP$$

- Degradation

$$M \xrightarrow{k_m^-} \varnothing \qquad P \xrightarrow{k_p^-} \varnothing$$

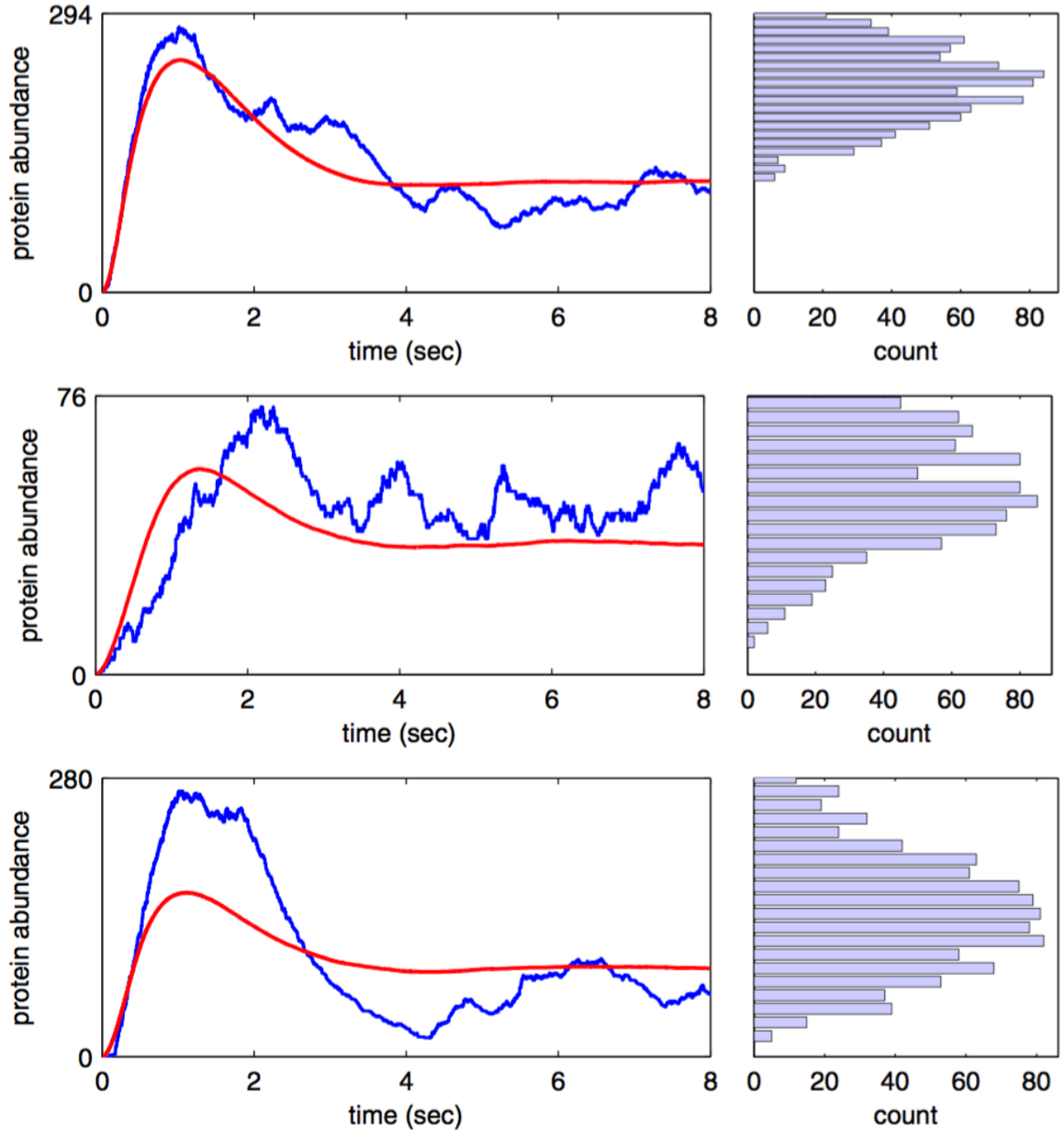
G: gene
M: mRNA
P: protein

M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- Add stochasticity
- Time courses of a single stochastic simulation algorithm run (blue)
- Mean over 1000 runs (red curve).
- Initial conditions
  - 10 copies gene G
  - 0 copies of other species
- Endpoint histogram shows empirical probability that a cell will have a given protein abundance.

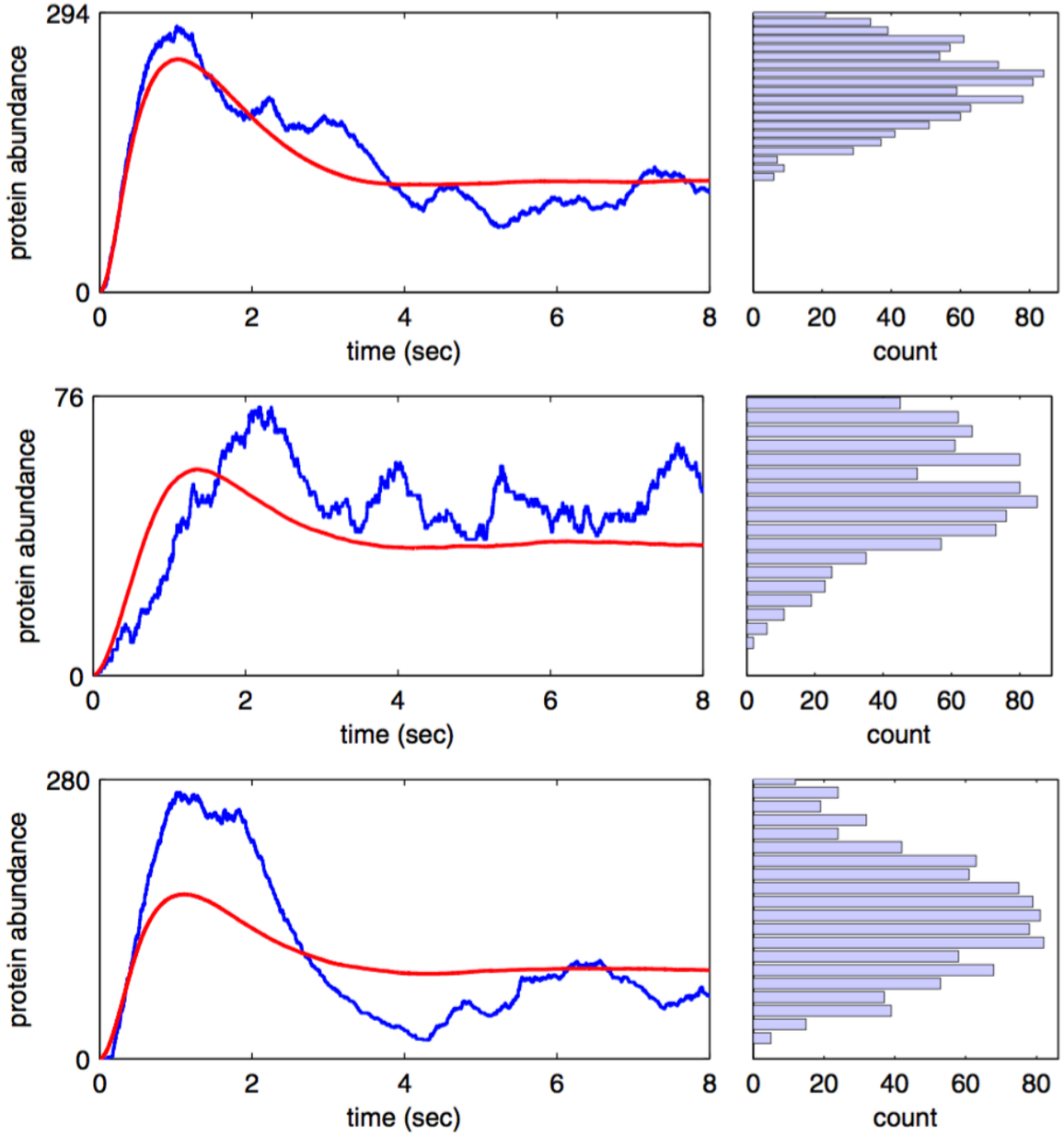


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- Top: Small fluctuations with high copy numbers of expressed mRNA and protein.

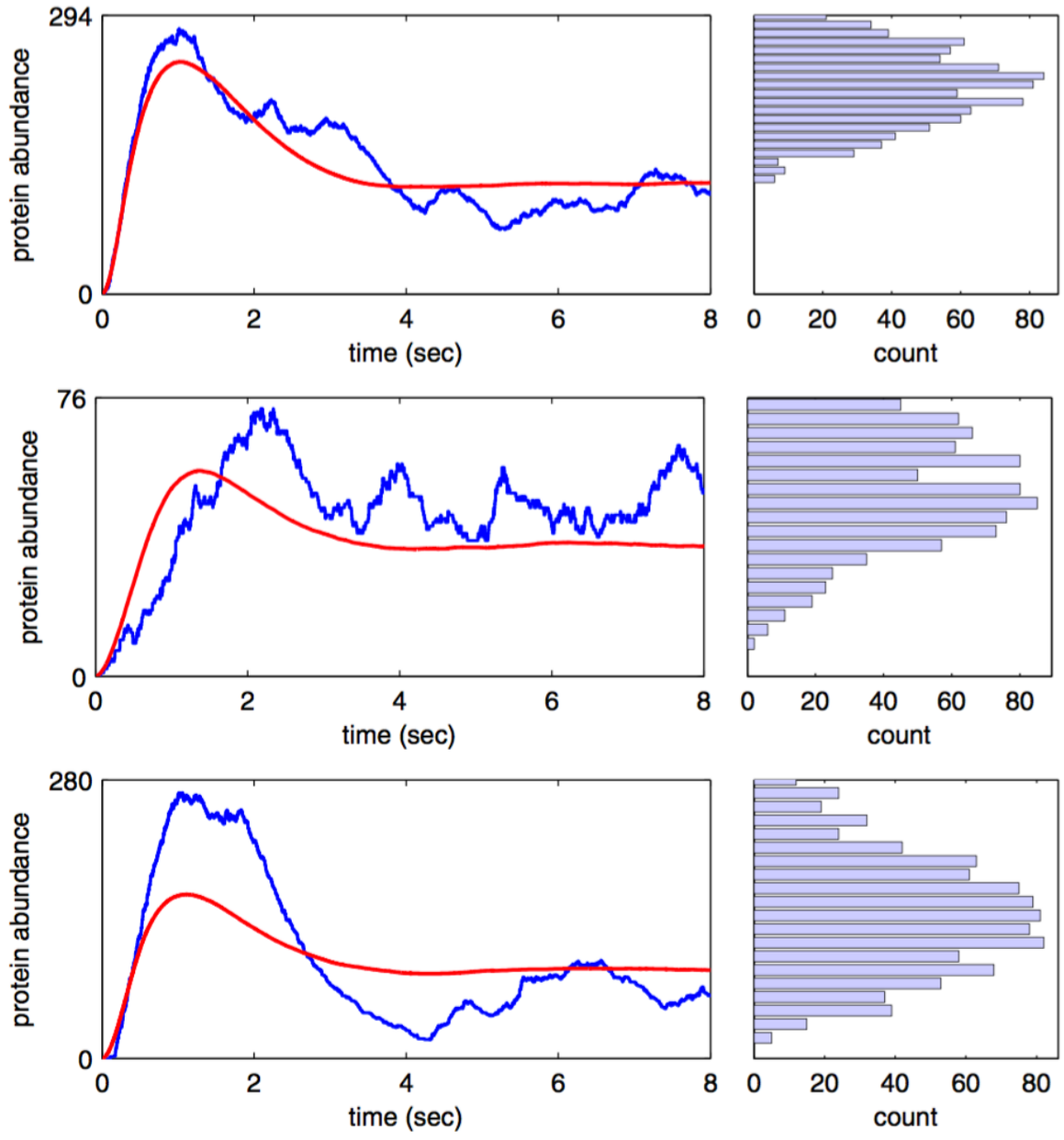


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- Middle: A tenfold decrease in the transcription rate k_m leads to
  - Decrease in the expressed mRNA abundance
  - An associated decrease in protein abundance
  - Large fluctuations in the protein abundance.

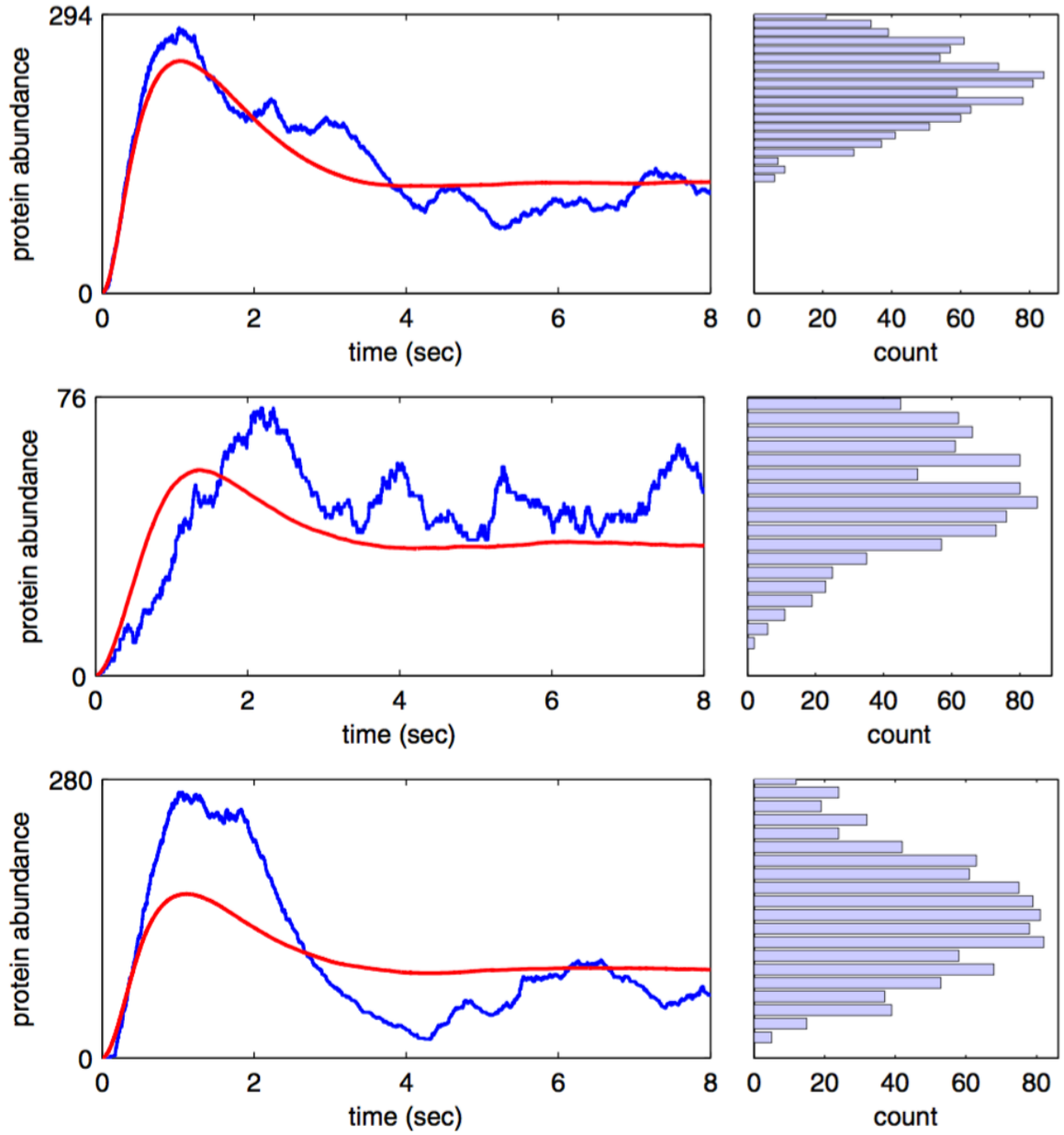


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- Bottom: Fluctuations in mRNA abundance at the transcription level are a second important factor contributing to gene-expression noise.
  - Tenfold decrease rate of transcription
  - Rate of translation is increased fivefold to keep the protein abundance more or less the same as in the first case

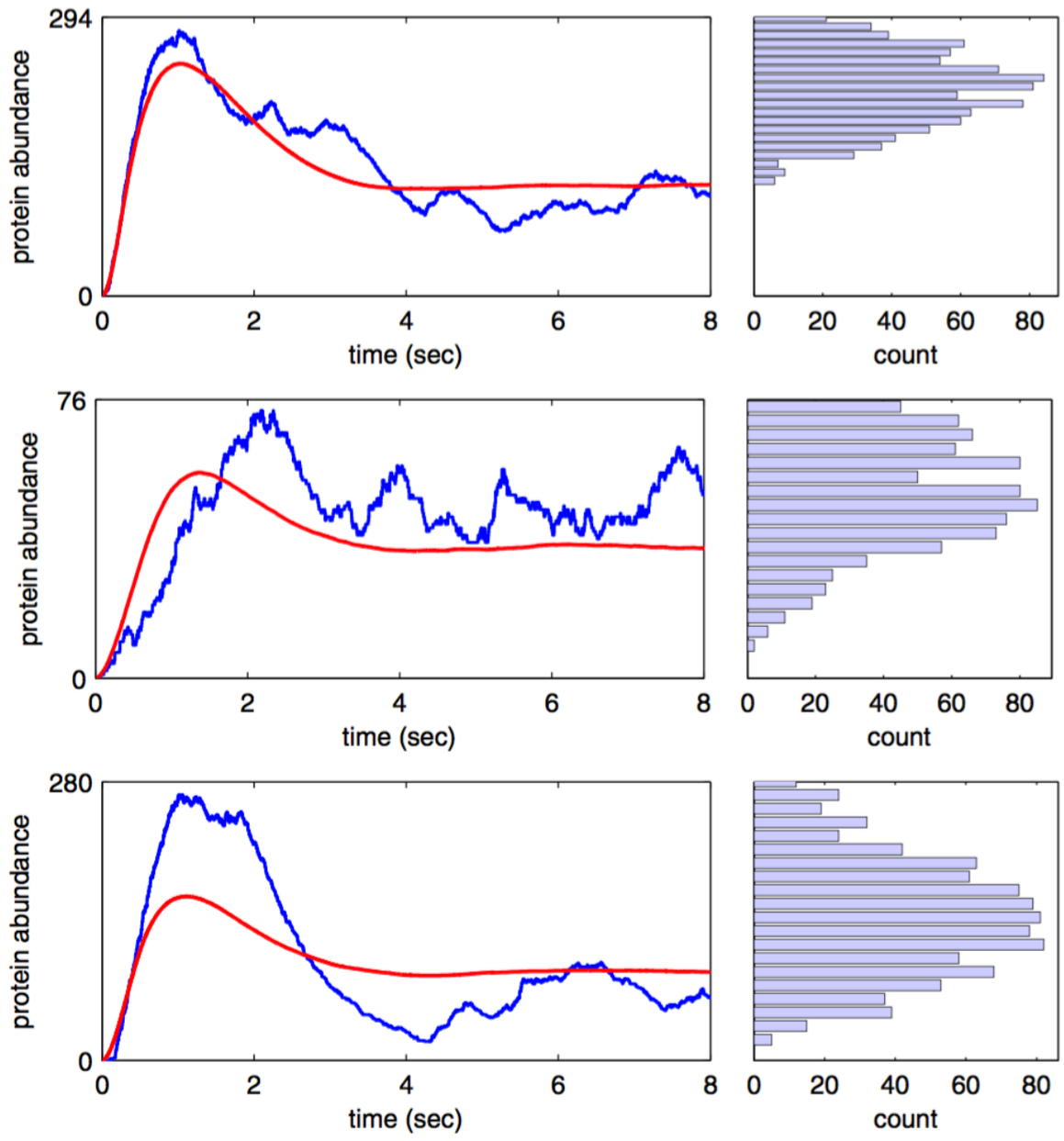


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- Bottom:
  - increased gene-expression noise in spite of large protein abundance
  - Noise attributable to increased fluctuations in mRNA abundance
    - Causing increased fluctuations in the rate of protein synthesis.

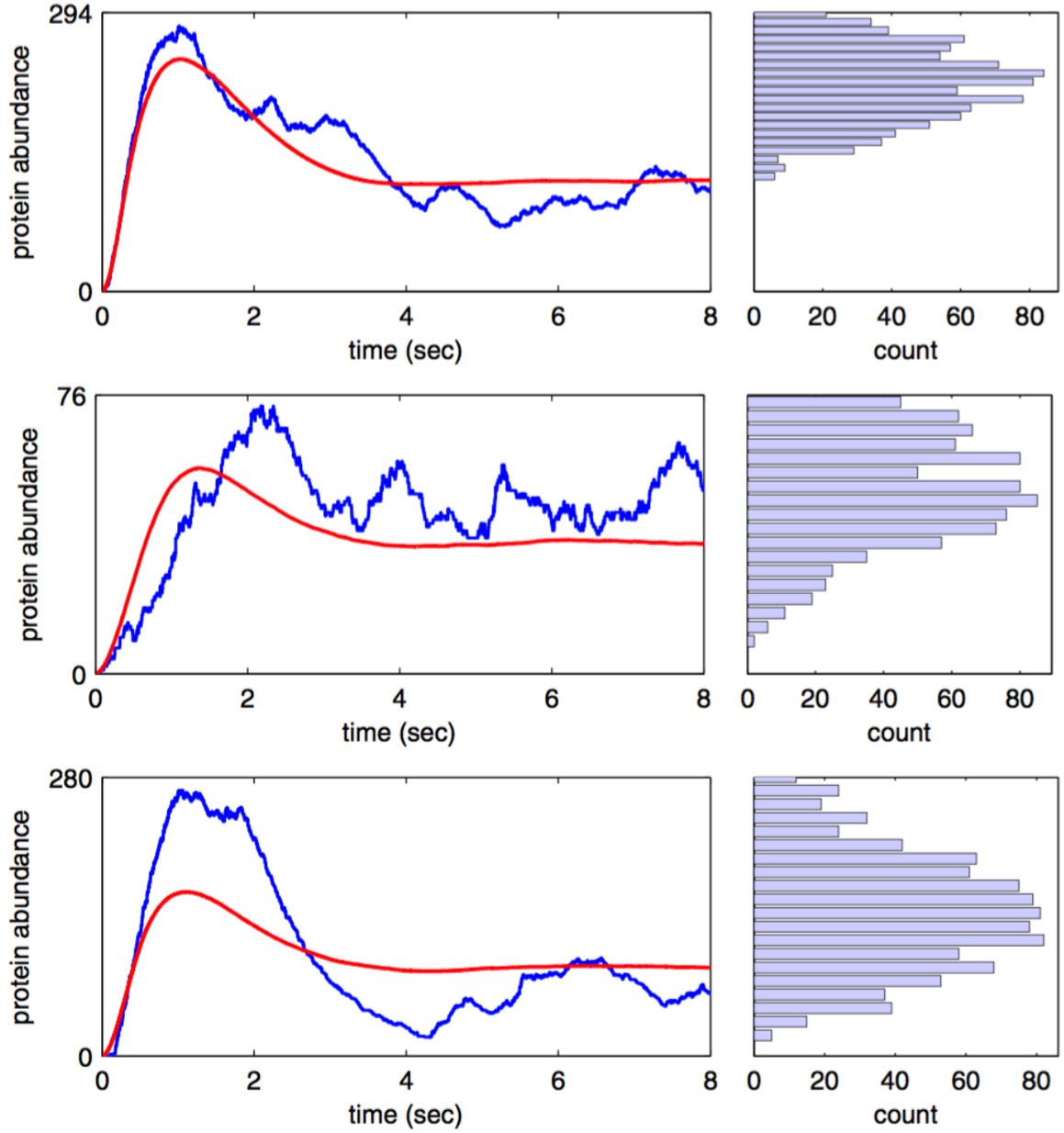


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Biochemical Models

Simplified Gene Regulation

- **Noise is propagated from transcription to translation.**

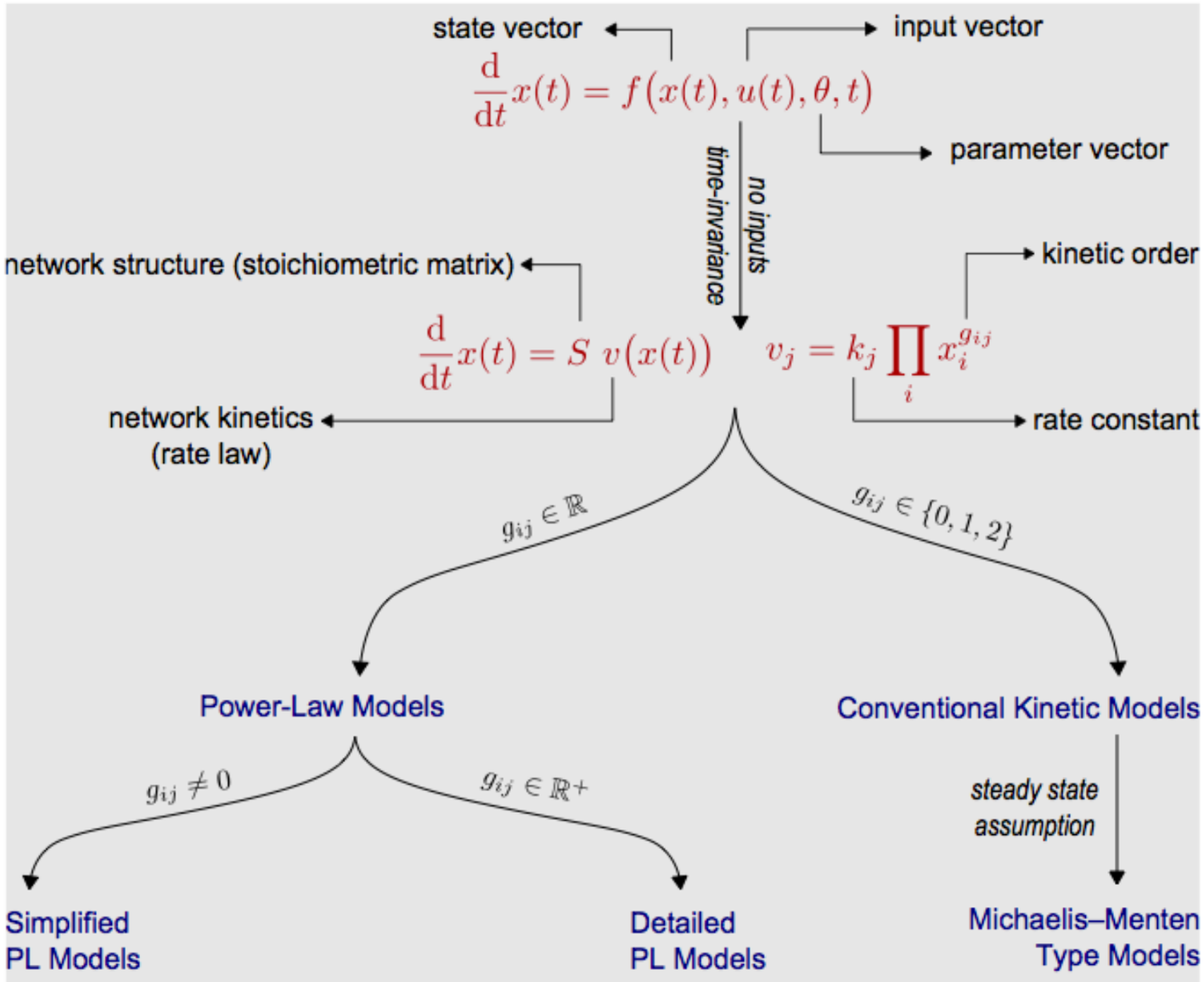


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Deterministic Models

Chemical Reaction Networks

- Differential Equations
  - Stoichiometry matrix
    - (network structure)
  - rate law
    - kinetics
    - collision theory
    - transition state theory

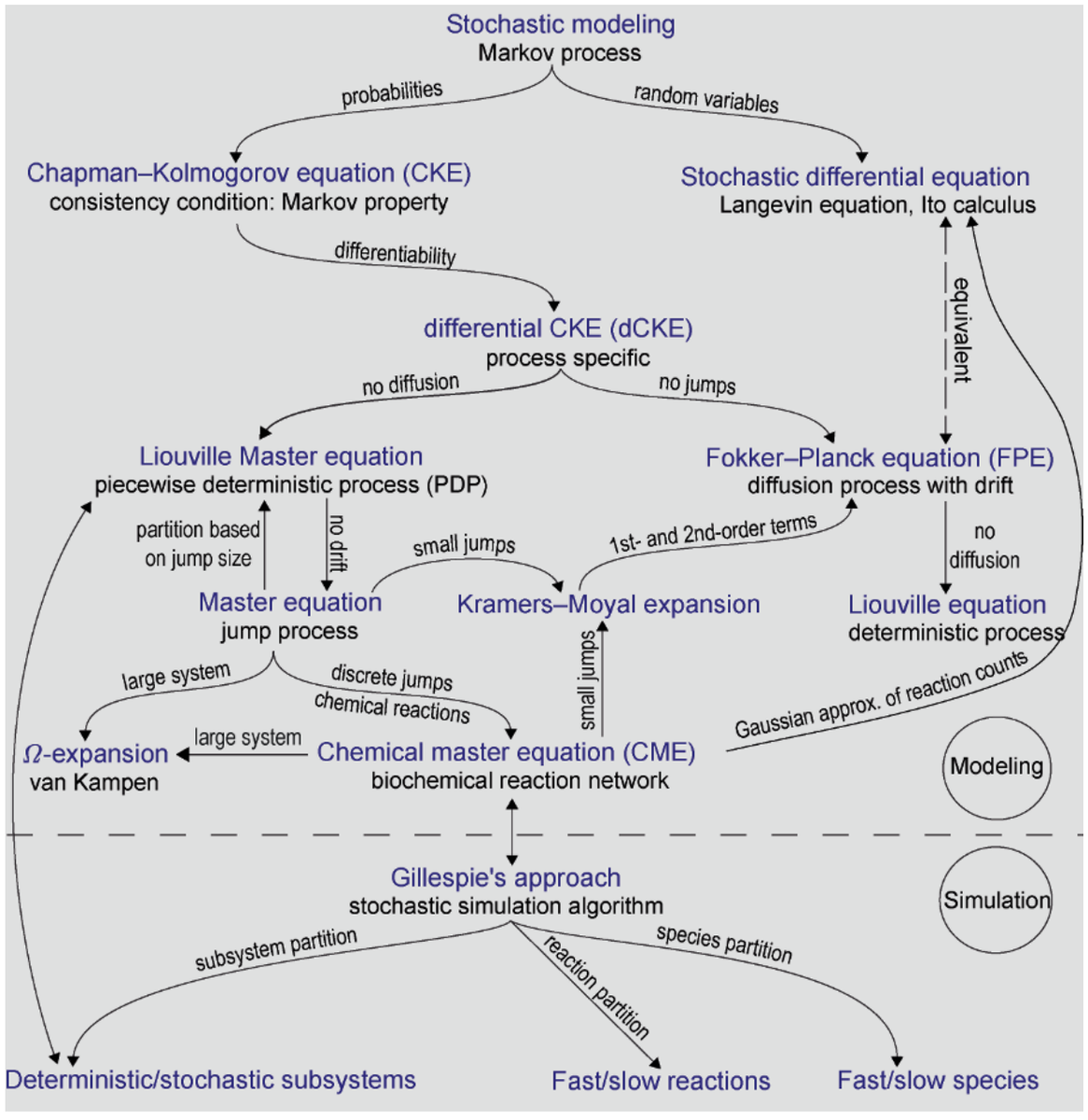


M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Stochastic Models

Chemical Reaction Networks

- Stochastic
  - statistical
  - fluctuations
  - noise



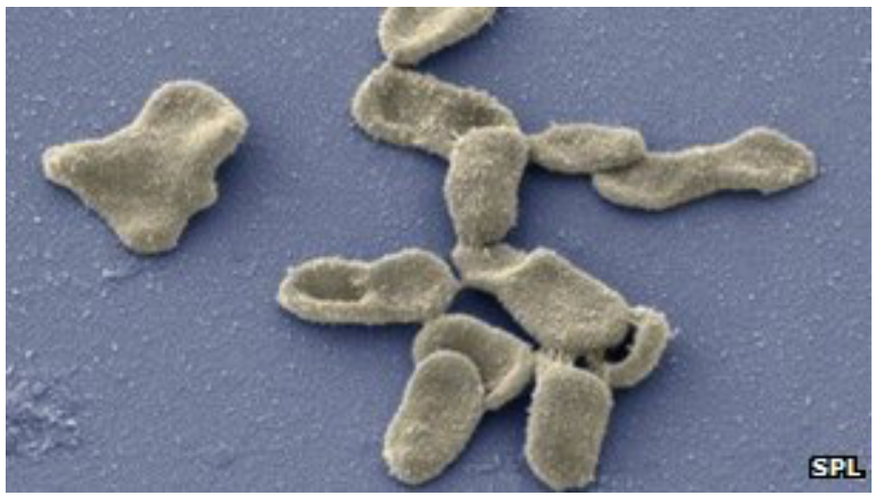
M. Ullah and O. Wolkenhauer, Stochastic Approaches for Systems Biology, Springer Science+Business Media, LLC (2011)

Example: Whole Cell Model

- *Mycoplasma genitalium*

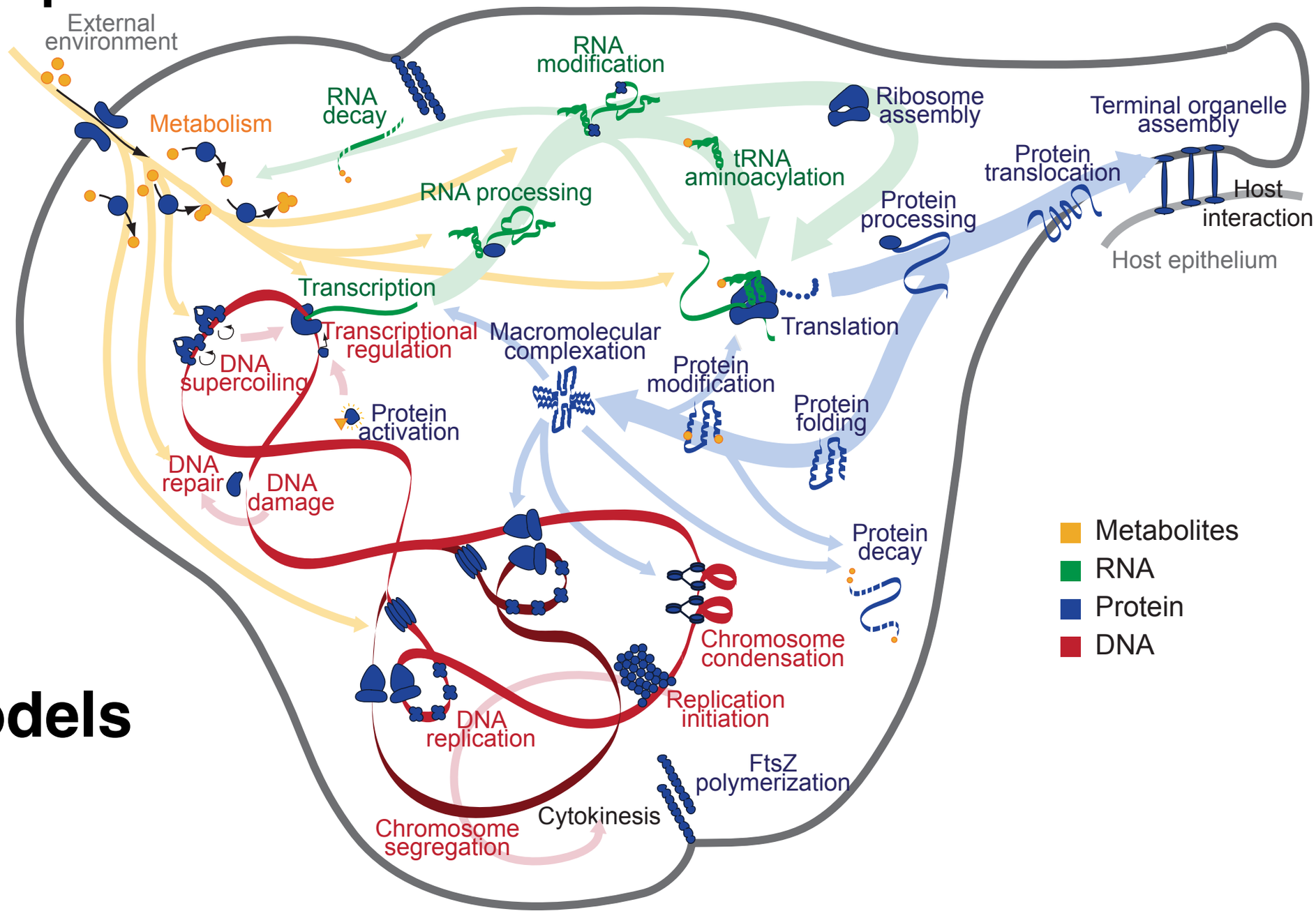
  parasitic bacterium
  synthetic genome 2008 (J. Craig Venter Institute)

- 525 Genes
- 580.07 kb pairs
- 2nd smallest



(left) Tully et al, International Journal of Systematic Bacteriology 33 (2): 387 (1983).
(right) http://www.bbc.com/news/science-environment-19016772
The virtual cell that simulates life

Example: Whole Cell Model



28 Submodels

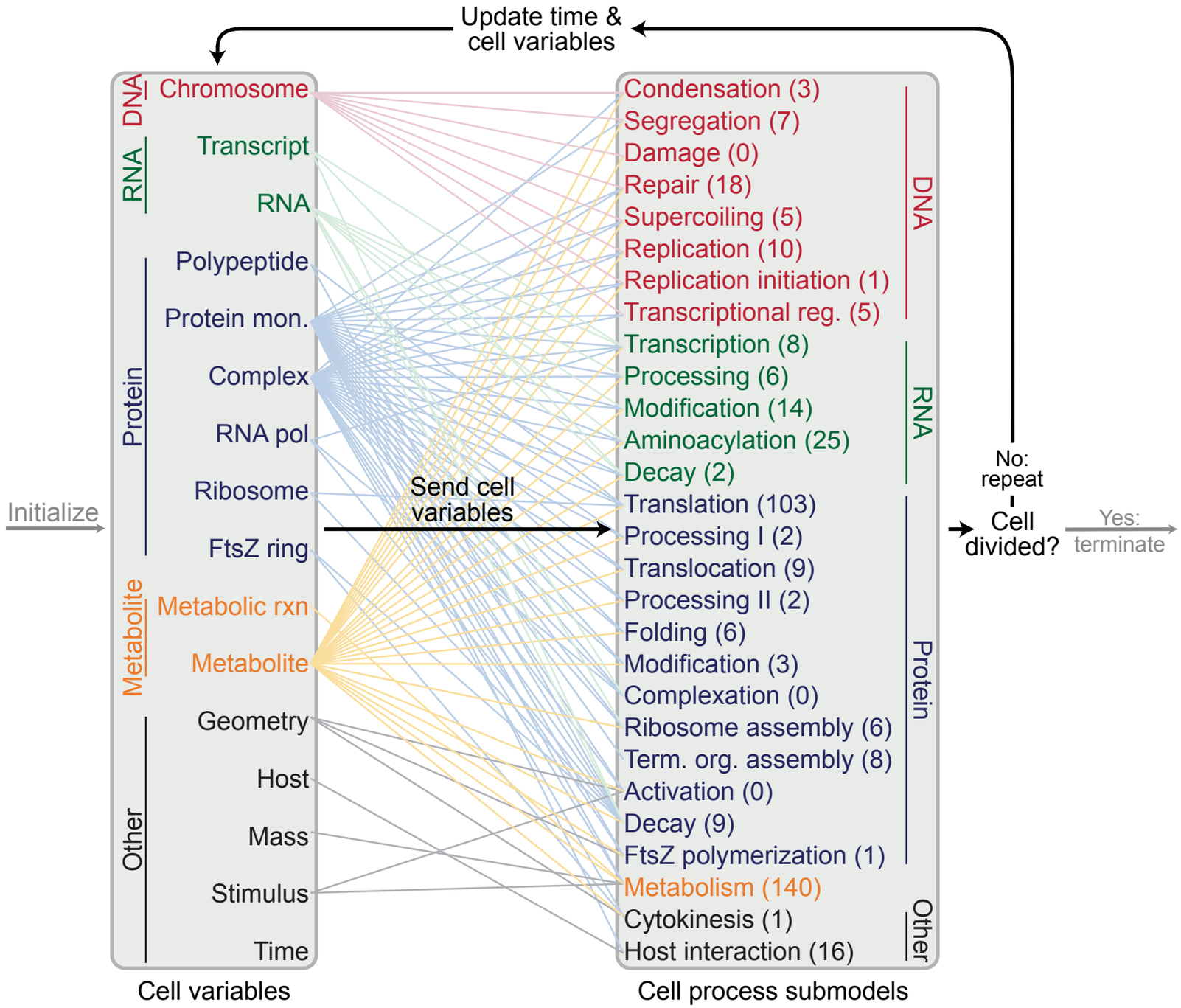
Karr et al., Cell 150, p 389 (2012)

Example: Whole Cell Model

16 Variables

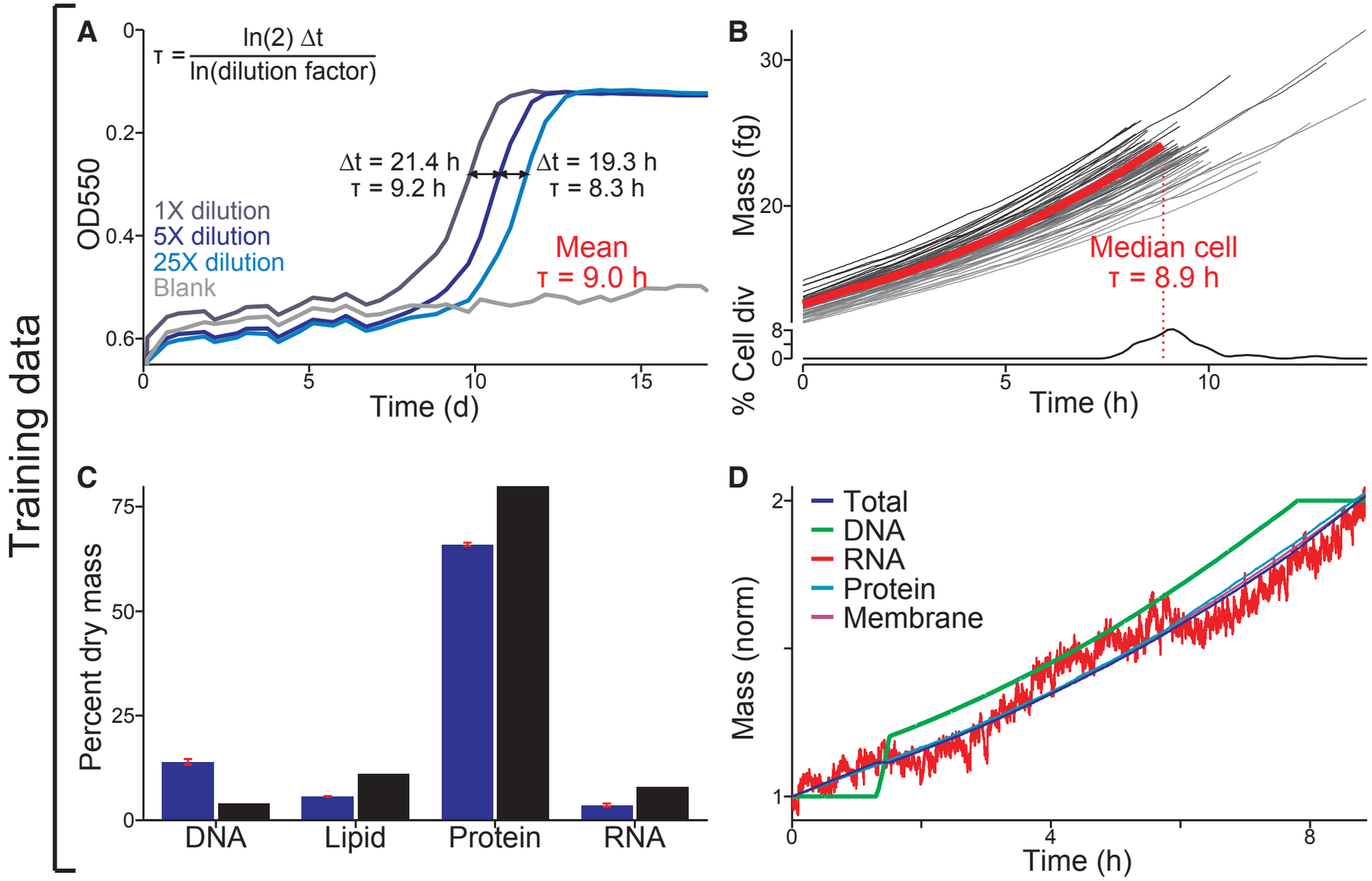
a. Random initialize
b. 1s time step
c. Repeat
d. Terminate on cell division

- Metabolites
- RNA
- Protein
- DNA



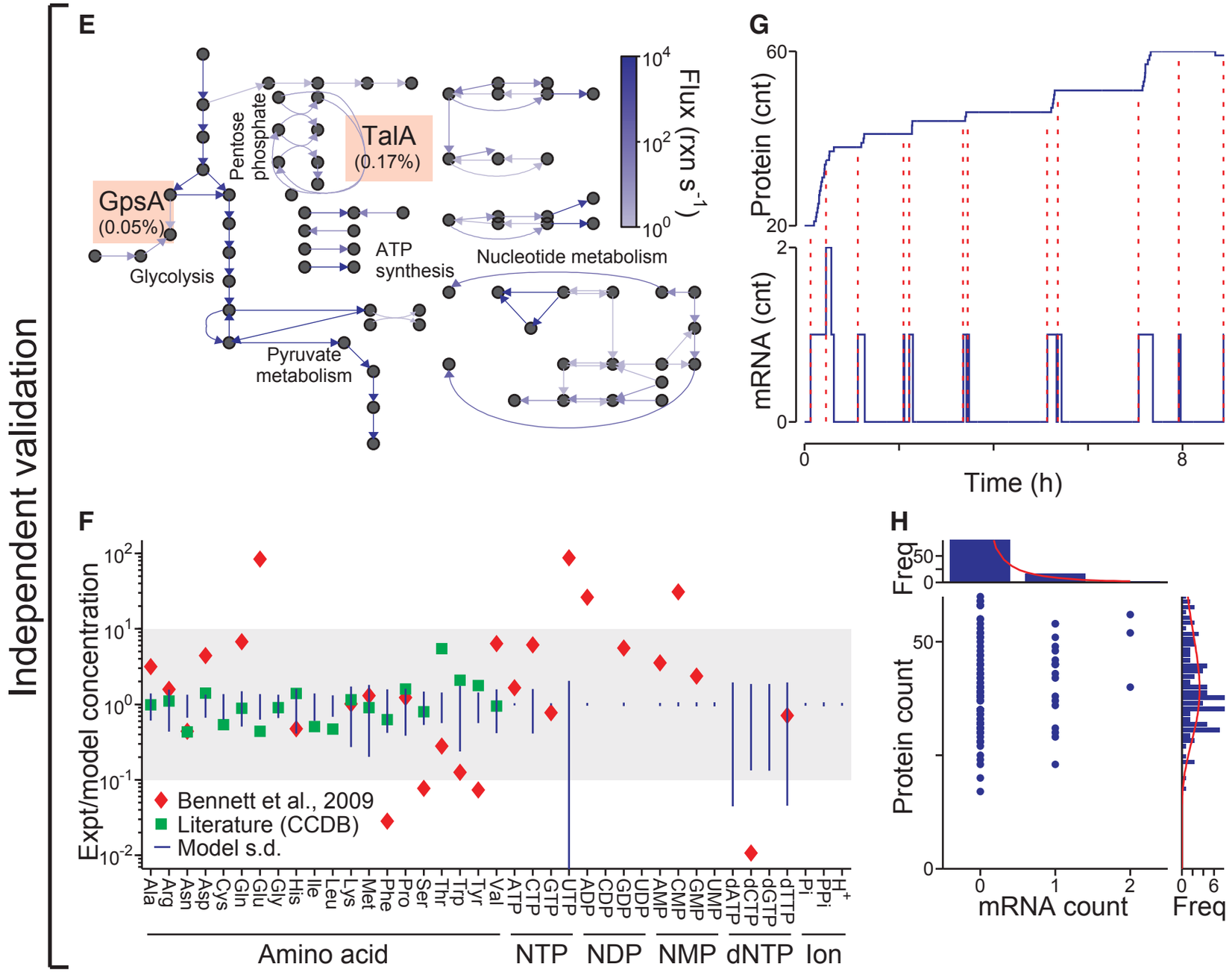
Karr et al., Cell 150, p 389 (2012)

Example: Whole Cell Model



Karr et al., Cell 150, p 389 (2012)

Example: Whole Cell Model



Karr et al., Cell 150, p 389 (2012)

Math tools

- Exploratory data analysis and Data Mining
    - Statistics
        - Distributions
        - Correlations
    - Clustering/Visualization
    - Dimensional Reductions
        - Principal Components Analysis
        - Neural Networks
    - Mathematics
        - Systems of differential Equations
    - different parts require different models

Math tools

http://www.r-project.org

http://www.mathworks.com/products/matlab/

https://www.gnu.org/software/octave/

http://www.wolfram.com/mathematica/

http://www.sagemath.org

home made varieties

(not an exhaustive list)

Representation

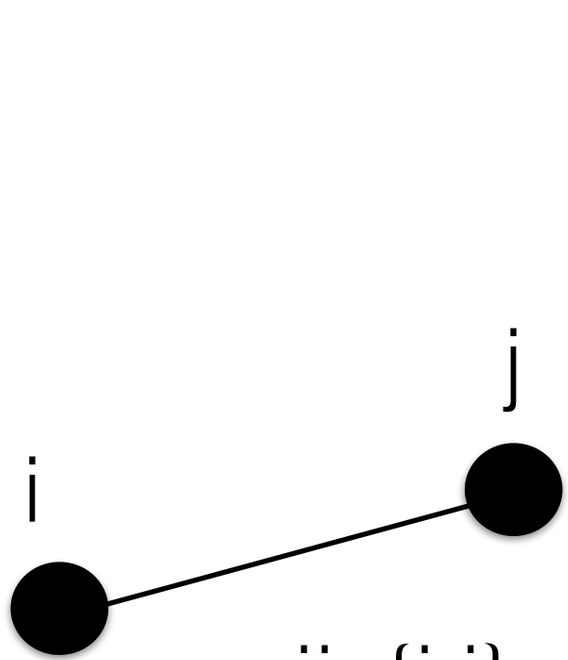
- **Common languages**
  - XML (extensible markup language)
    - <tag attribute=“value> content </tag>
    - SBML (Systems Biology Markup Language)
    - CellML
      - SED-ML (Simulation Experiment Description Markup Language)
  - BioPAX (Biological Pathway Exchange - Resource Description Framework [RDF])
- **Databases**
  - KEGG, Gene Ontology (GEO), WikiPathways, IPA, Reactome
- **Networks** (next)
- **Multi-model systems**

Networks & Pathways

Graphs

Vertices (nodes) {V}

Edges (links) {E}

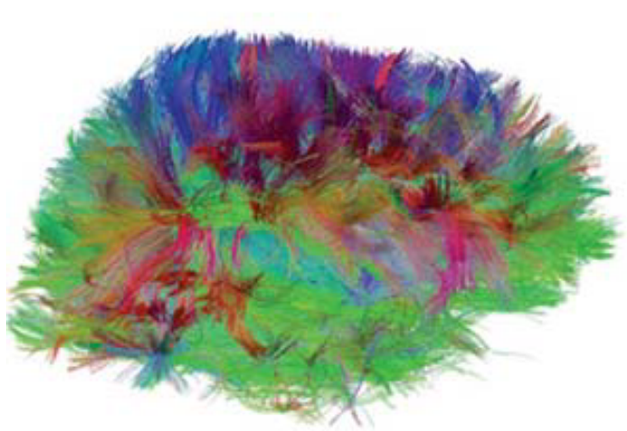


ij, {i,j}, <i,j>, $e_{ij,}$

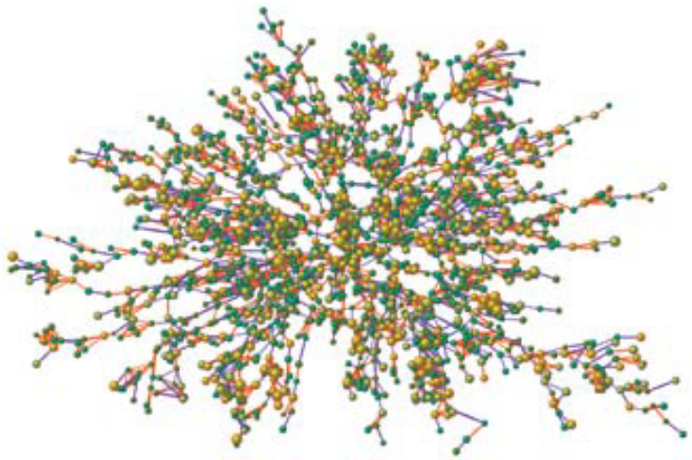
graph G = (V, E)
V(G) ={set of vertices}
E(G) = {set of edges}

Motter & Albert, Phys. Today 65(4), 43 (2012)

Examples



Adult Human Brain

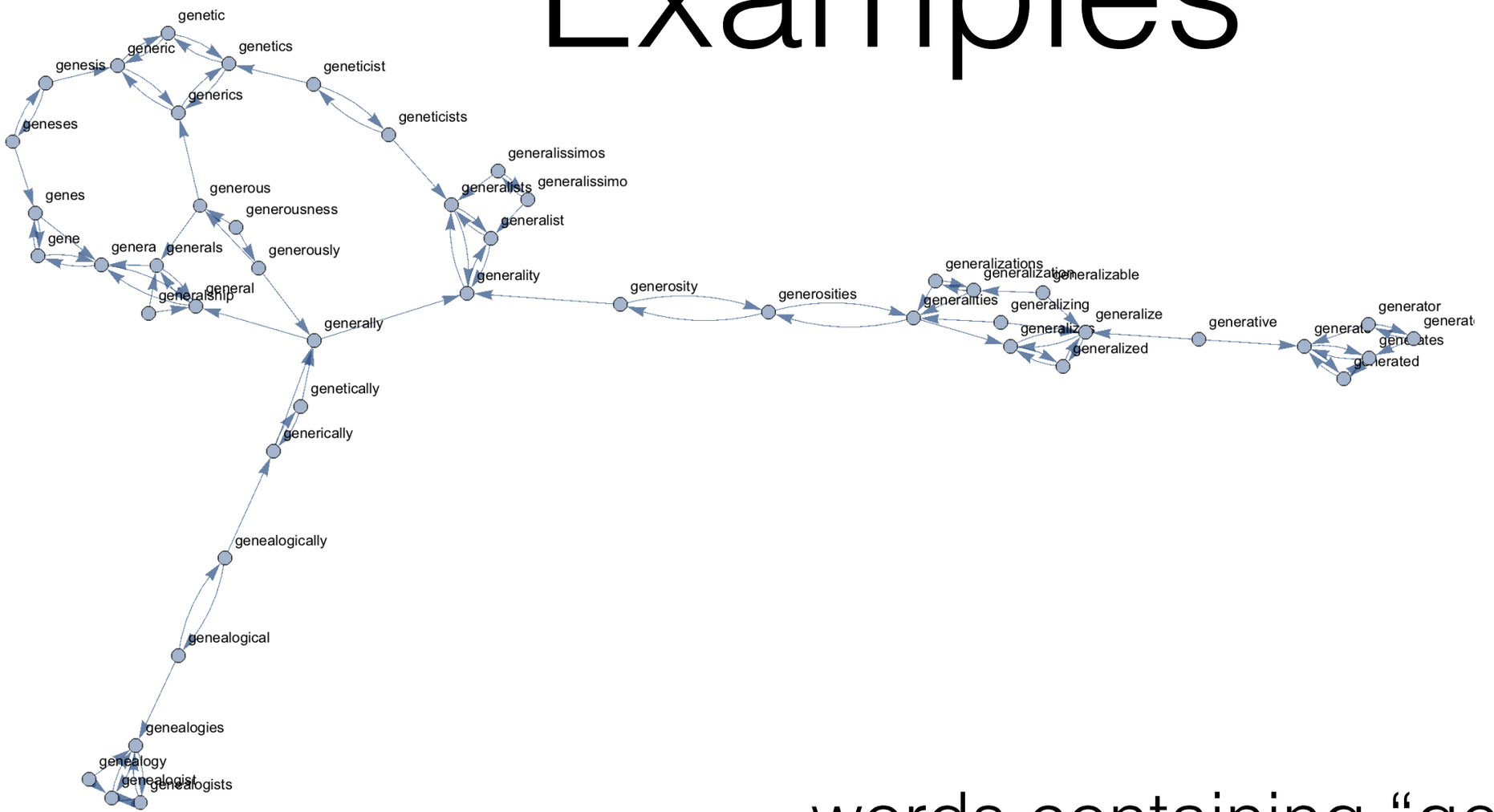


Social Network Influence of Body Weight

Internet Service Providers

Motter & Albert, Phys. Today 65(4), 43 (2012)

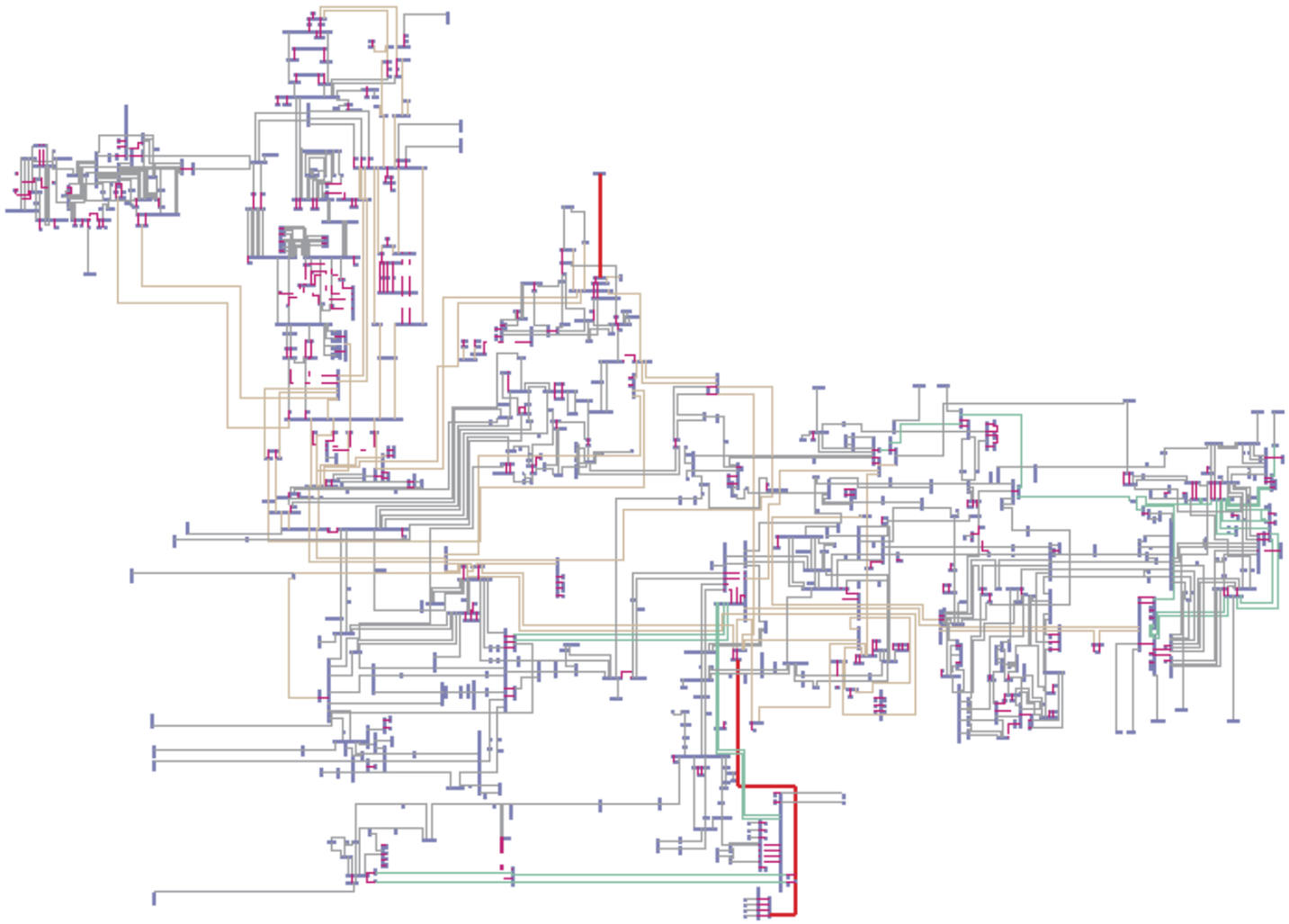
Examples



words containing “gene”

Mathematica, Wolfram Research Inc. (2006-16)

Examples



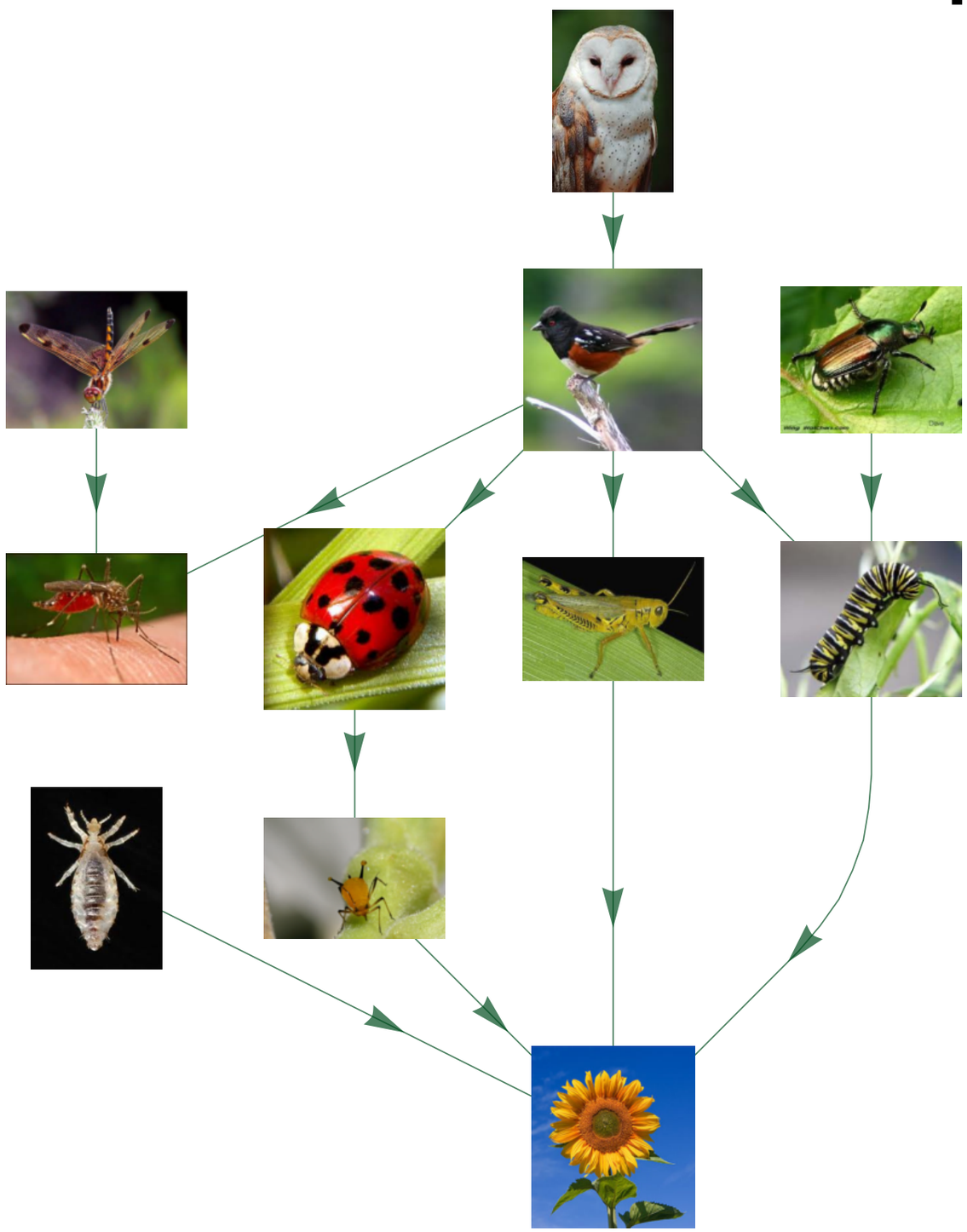
Nodes: Generators and substations

Edges: Transmission lines and transformers. (Line thickness and color indicate the voltage level)

NY Power Grid

Strogatz, Nature 410**,** p268 (2001)

Examples

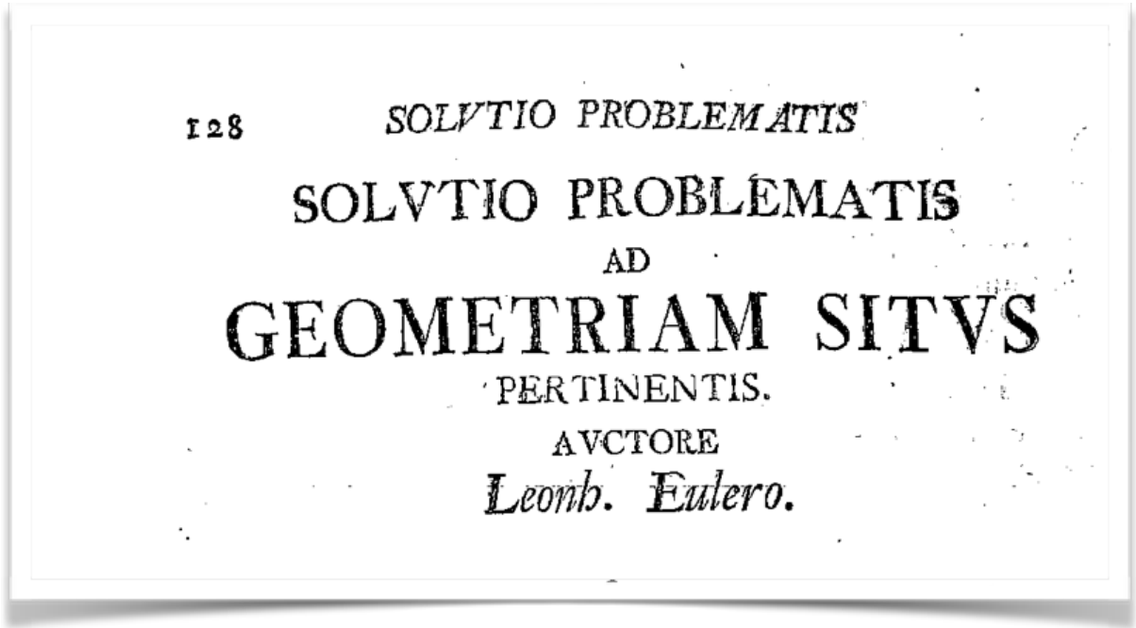


Food Web Network

Mathematica, Wolfram Research Inc. (2006-16)

Bridges of Königsberg

- Leonhard Euler
  - 1736
  - 7 bridges across Pregolya River
  - Prussian Königsberg (now Kaliningrad, Russia)
  - Is there a route to cross each bridge exactly once?

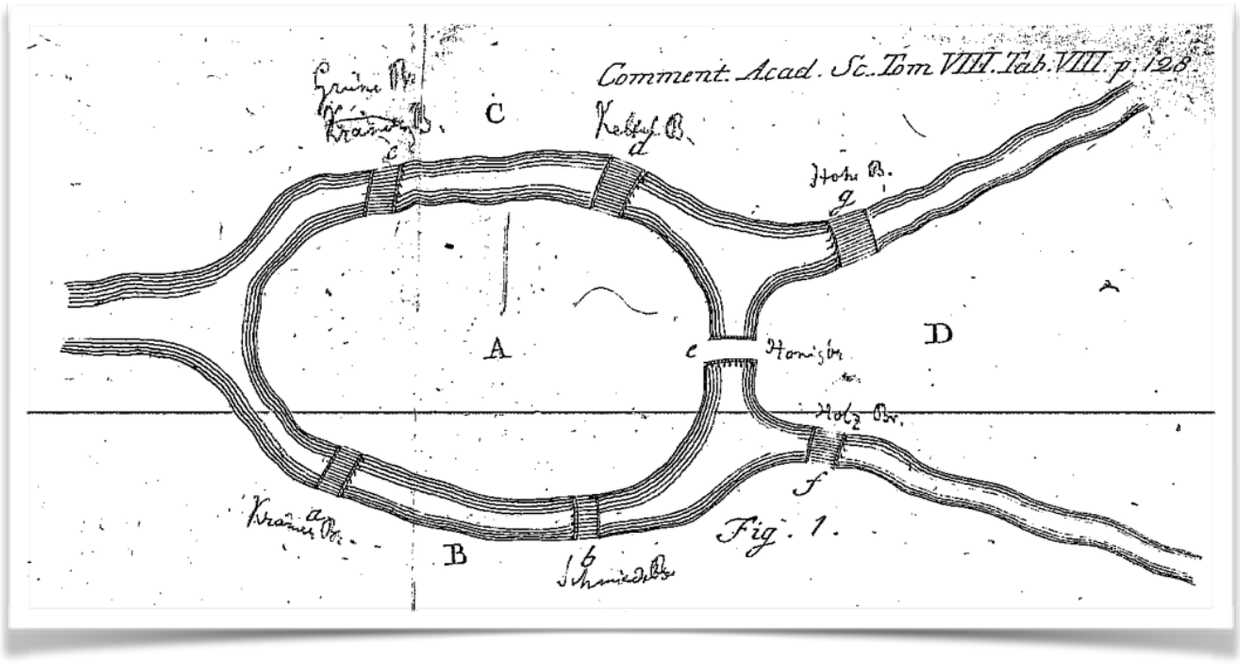

128 SOLVTIO PROBLEMATIS

SOLVTIO PROBLEMATIS
AD
GEOMETRIAM SITVS
PERTINENTIS.
AVCTORE
Leonh. Eulero.

The Euler Archive http://eulerarchive.maa.org

Bridges of Königsberg

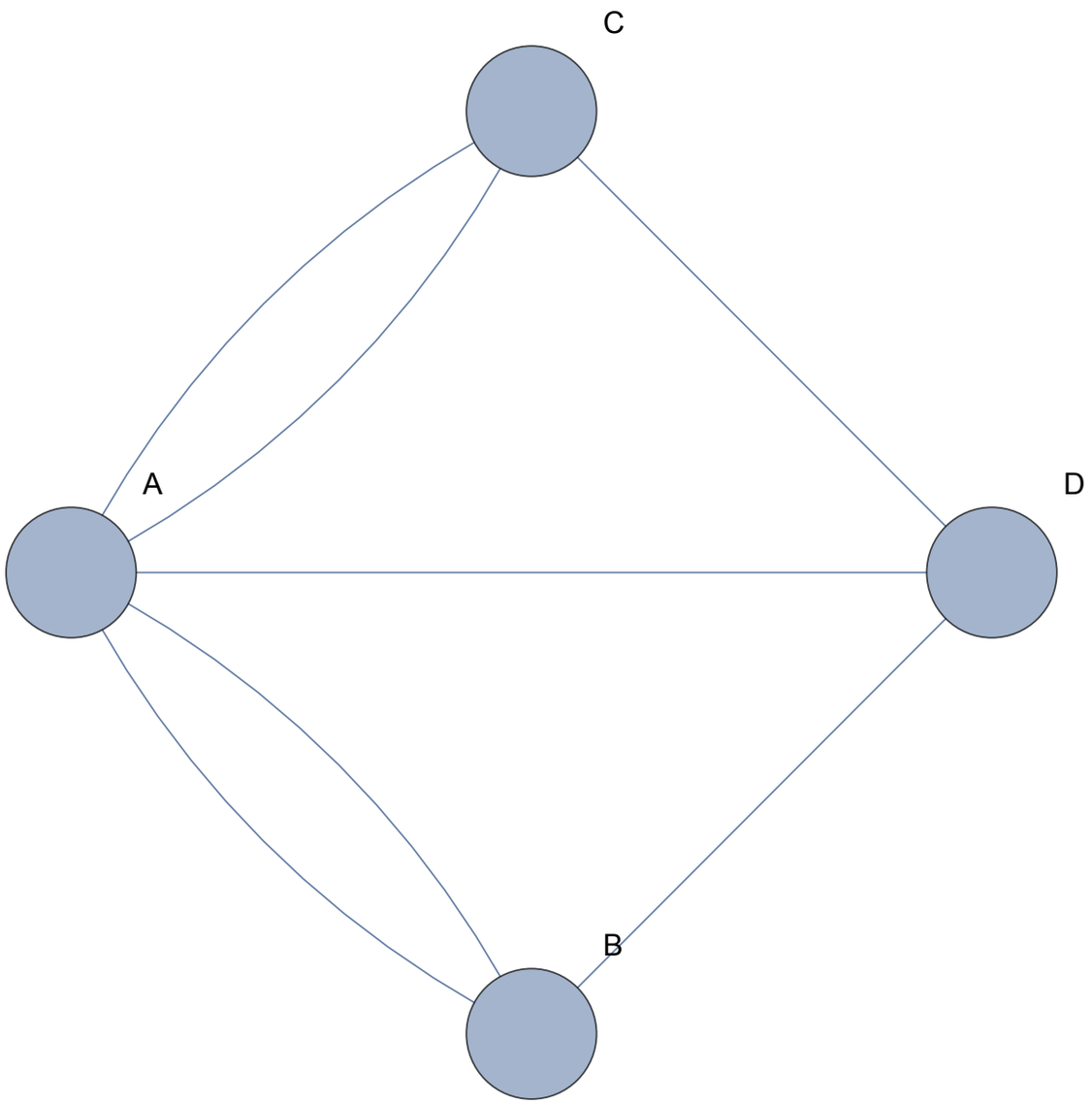
- Leonhard Euler
  - 1736
  - 7 bridges across Pregolya River
- Prussian Königsberg (now Kaliningrad, Russia)
- Is there a route to cross each bridge exactly once?

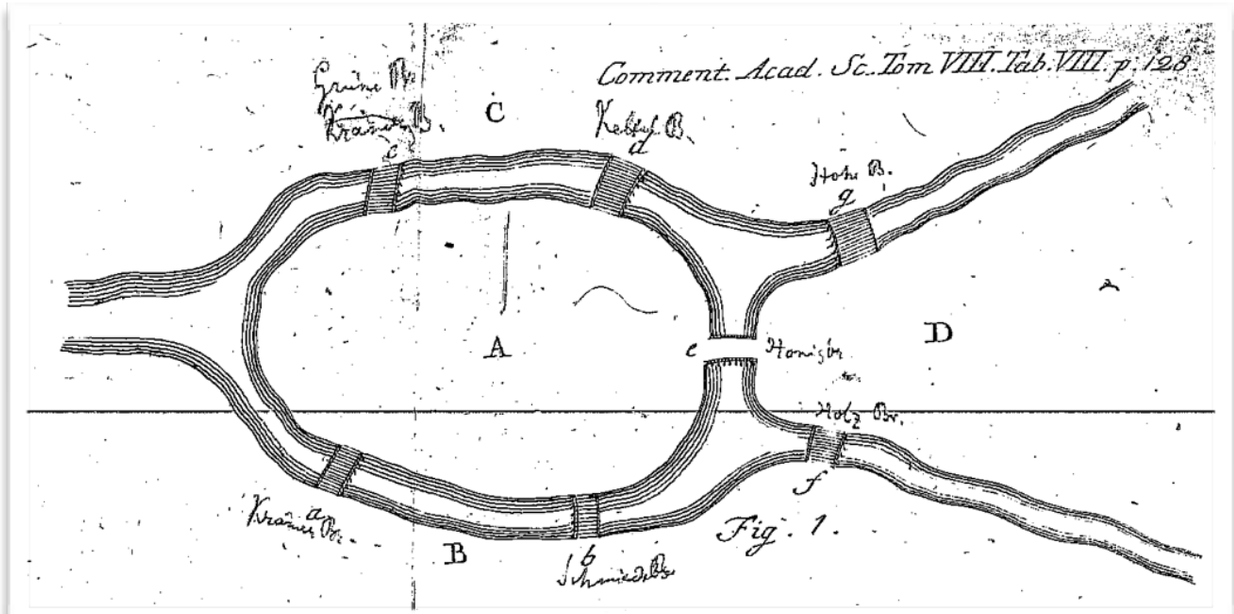


e.g. AB, BA, AC, CA, AD, …,?

The Euler Archive http://eulerarchive.maa.org

Bridges of Königsberg

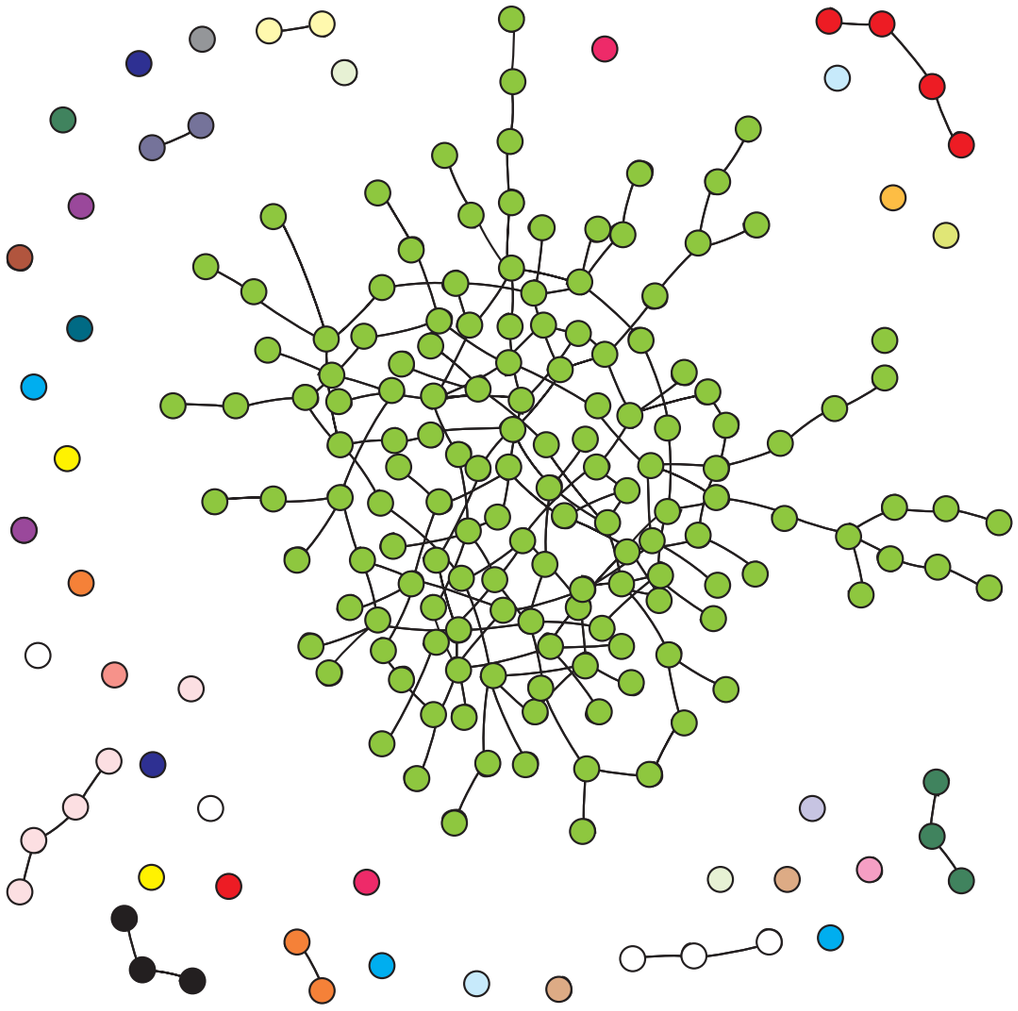


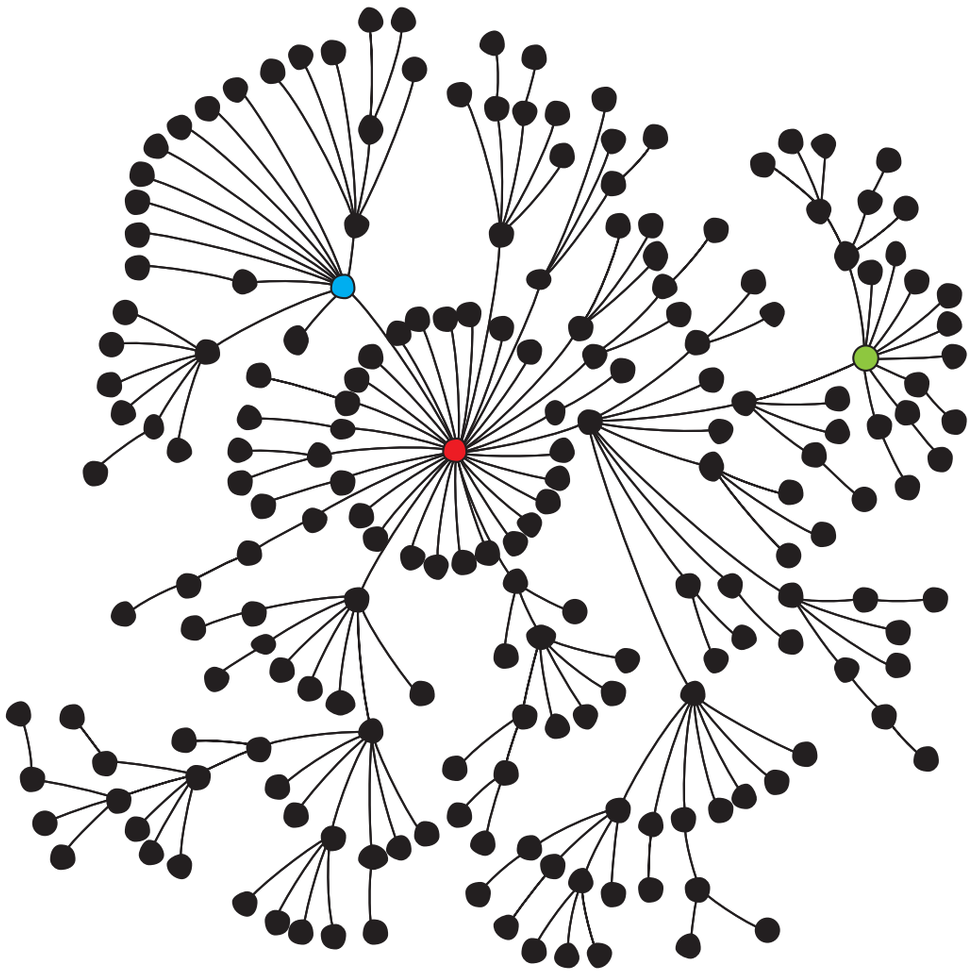


e.g. AB, BA, AC, CA, AD, …,?

The Euler Archive http://eulerarchive.maa.org

Random Networks





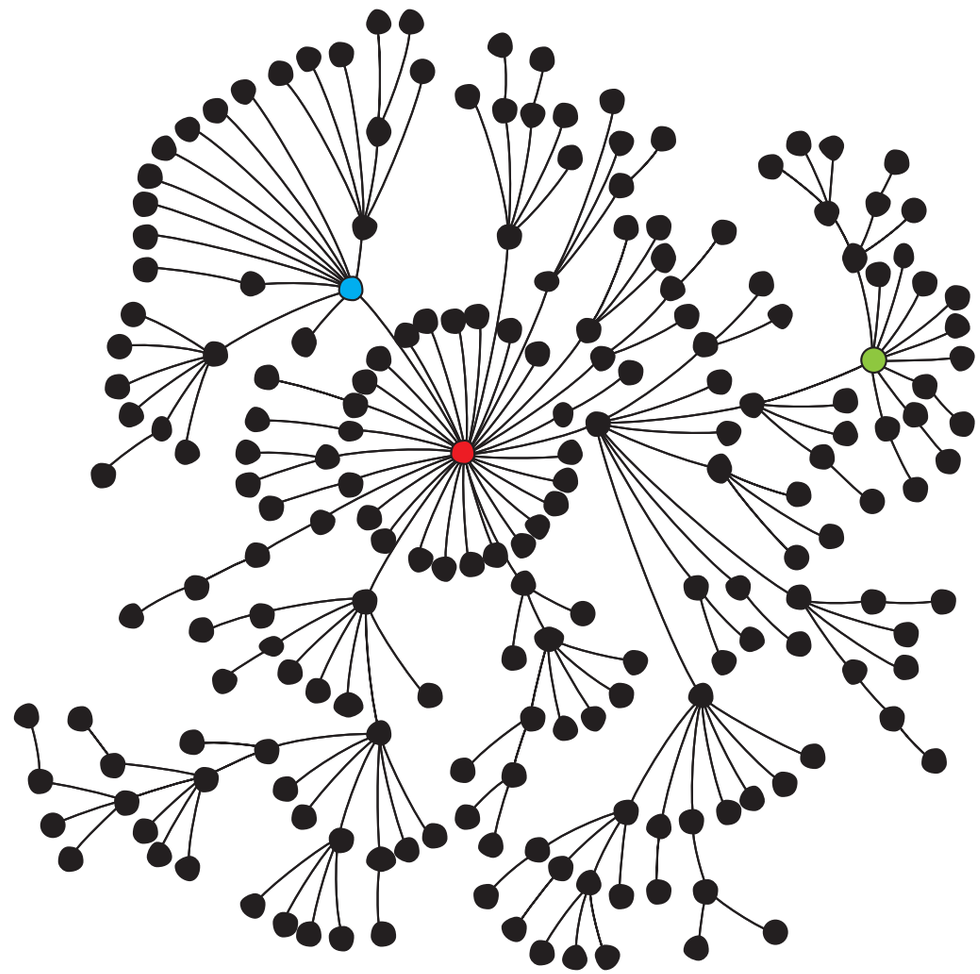
Scale Free Graph

Strogatz, Nature 410**,** p268 (2001)

Properties

Scale Free Networks

- Robust
  - if vertices removed network still connected
- Vulnerable
  - can target hubs



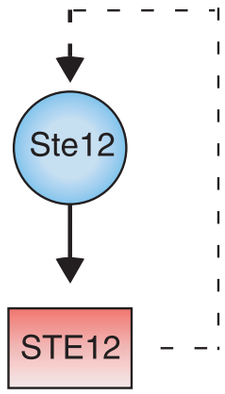
Strogatz, Nature 410**,** p268 (2001)
Albert et al., Nature 406 p.378 (2000)

Community Structure

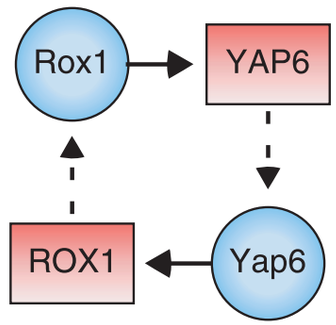
- **Social Networks**
  - groups of different people
- **internet pages**
  - various topics
- **Metabolites**
  - various functions
- **genes**
  - various processes

Gene Regulation

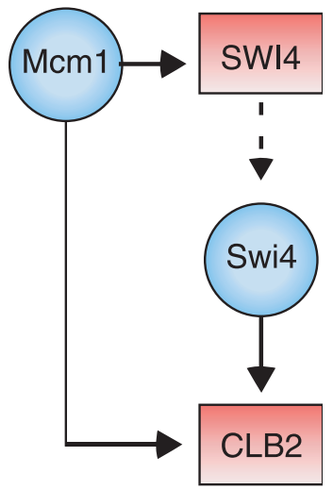
Autoregulation



Multi-Component Loop

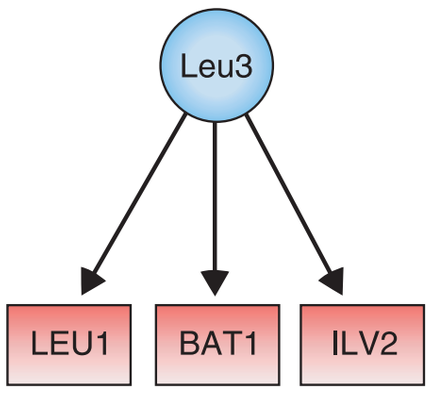


Feedforward Loop

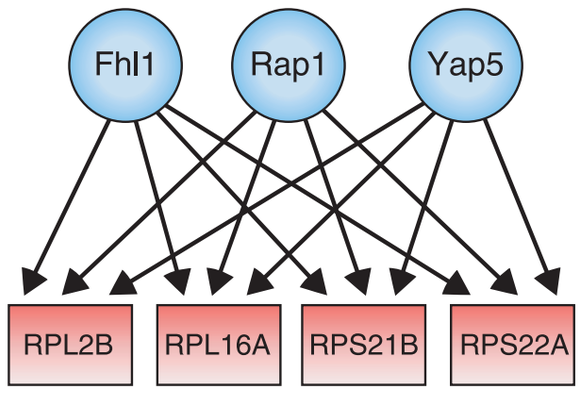


- *Saccharomyces cerevisiae*
- Activators (increase)
- Repressor (decrease)
- Feedback loops
- Motifs

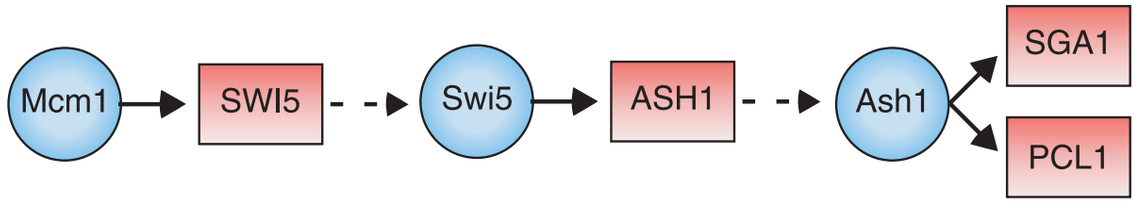
Single Input Motif



Multi-Input Motif

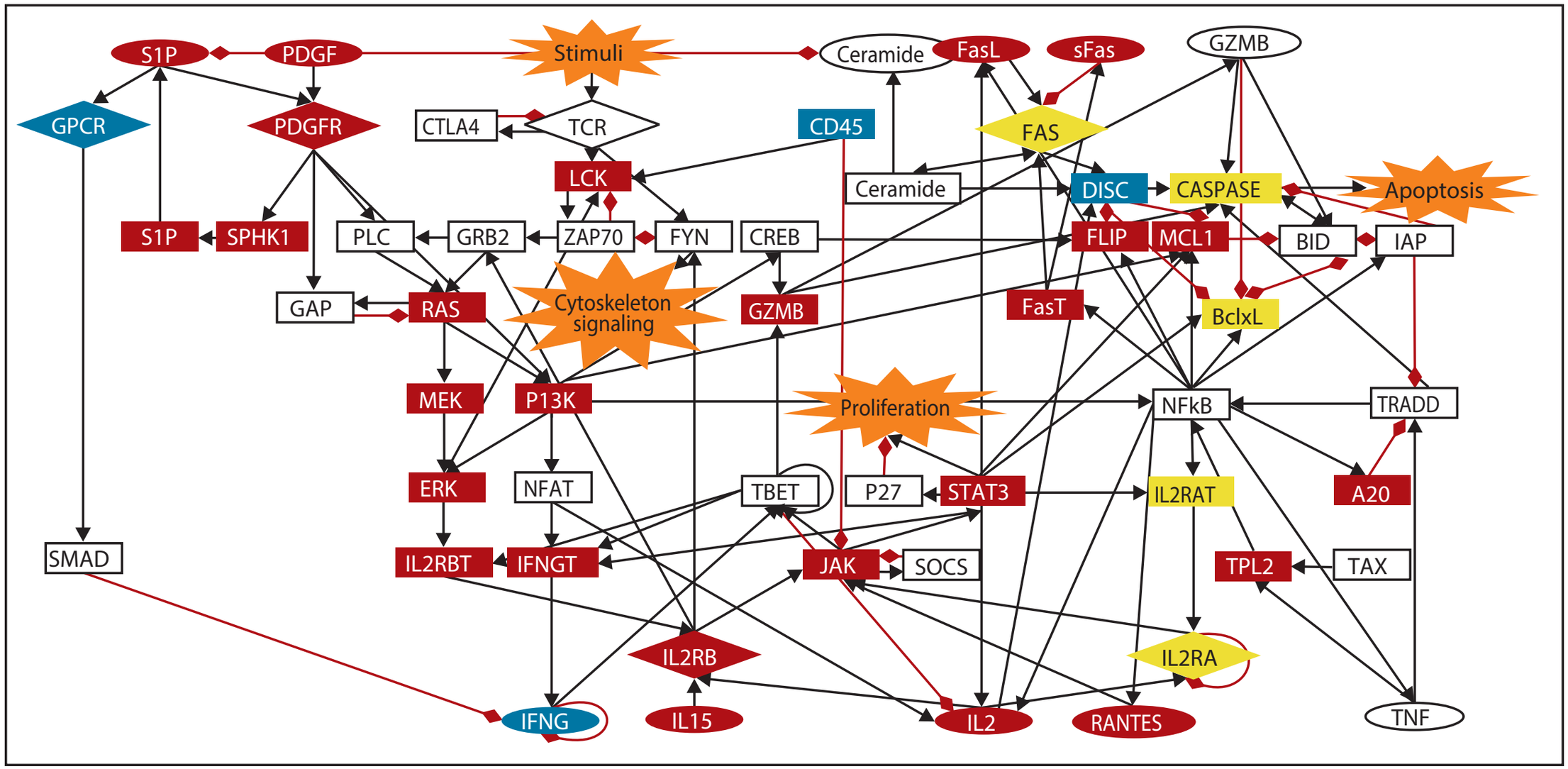


Regulator Chain



Lee et al., Science 298, p.799 (2002)

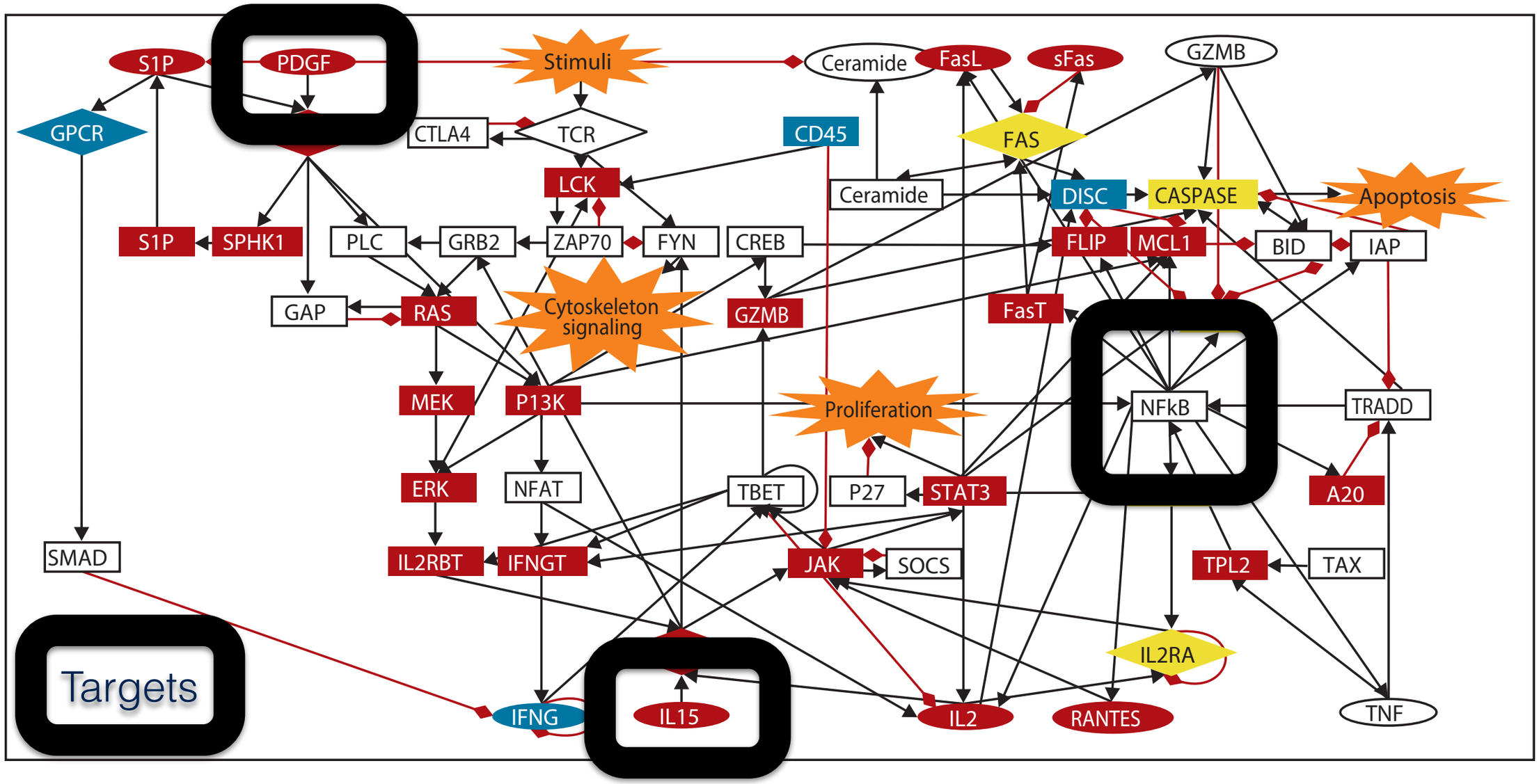
Examples



T cell signaling (Apoptosis) Protein Network

Motter & Albert, Phys. Today 65(4), 43 (2012)

Examples

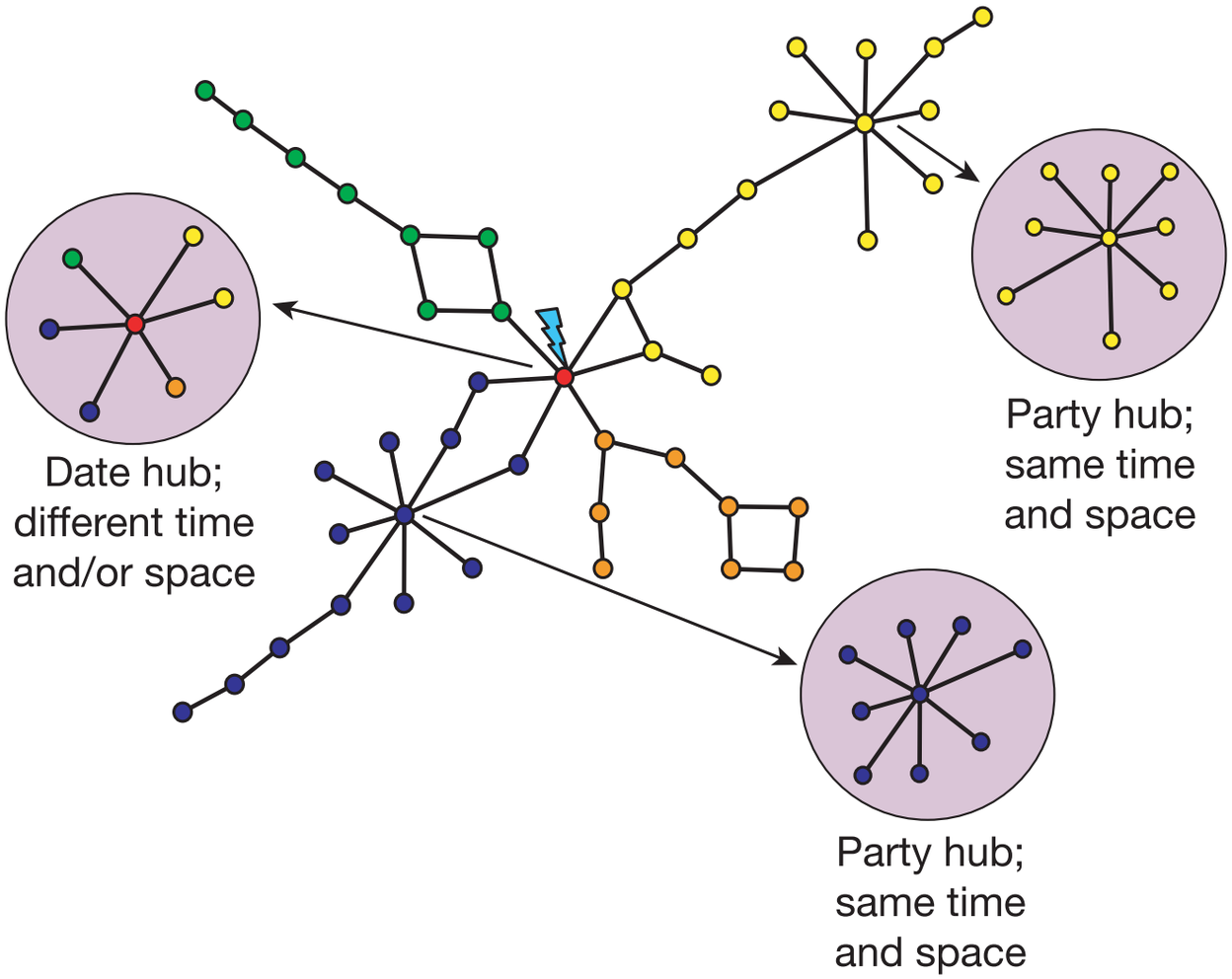


T cell signaling (Apoptosis) Protein Network

Motter & Albert, Phys. Today 65(4), 43 (2012)

Examples

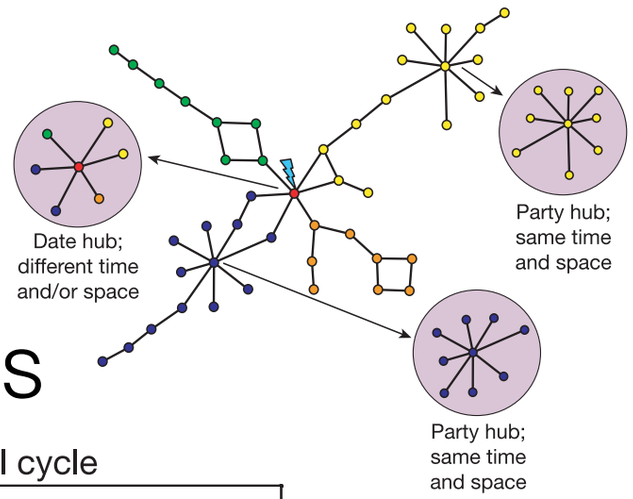
modularity in Yeast Protein-Protein Interactions

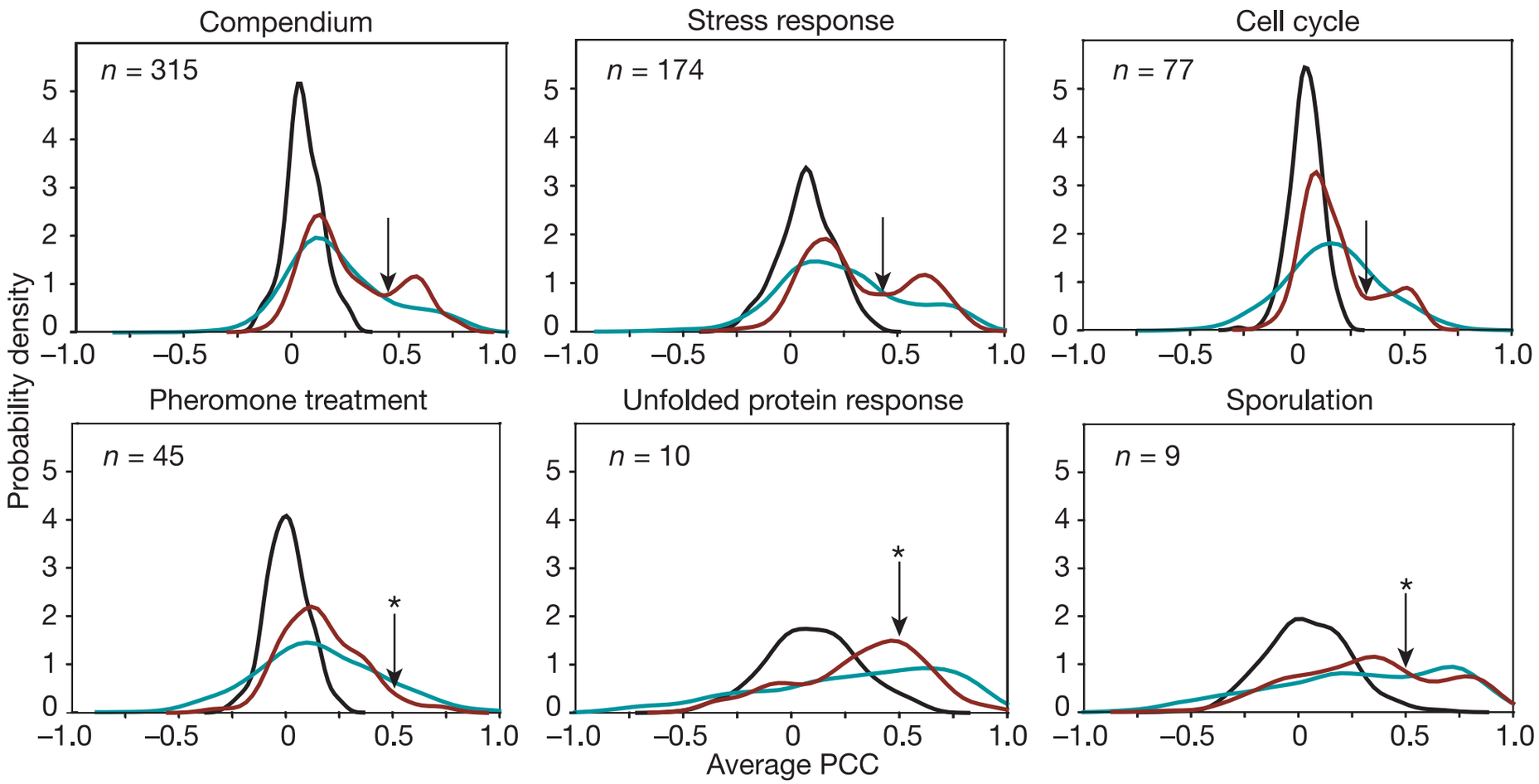


Han et al., Nature 430, p. 88 (2004)

Examples

modularity in Yeast Protein-Protein Interactions



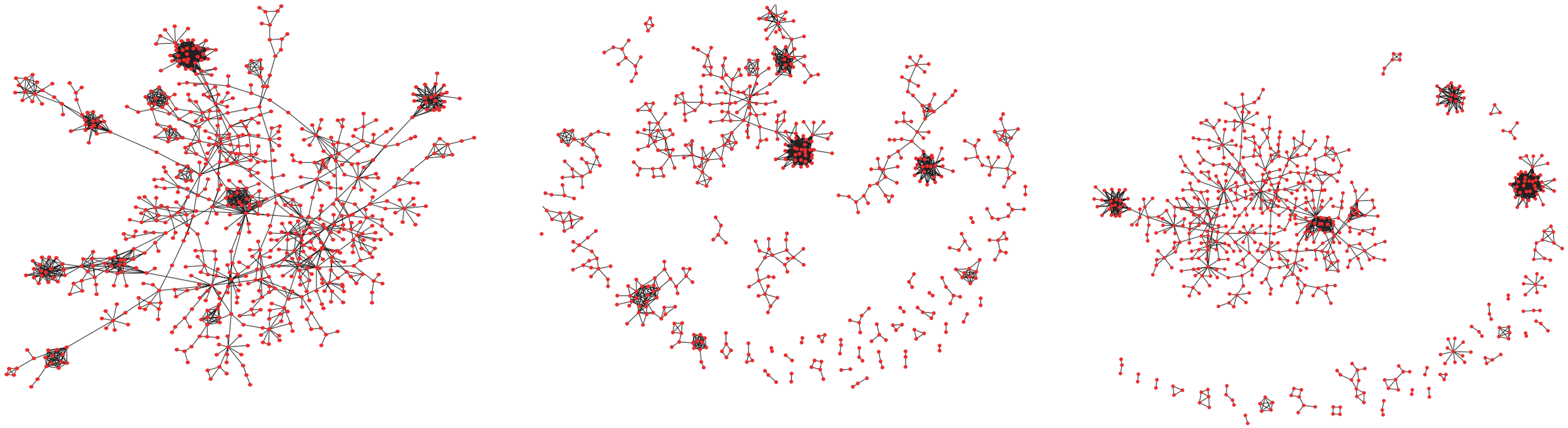


PCC: average Pierson Correlation Coefficient between the hub and each of its respective partners for mRNA expression

Han et al., Nature 430, p. 88 (2004)

Examples

modularity in Yeast Protein-Protein Interactions



main component

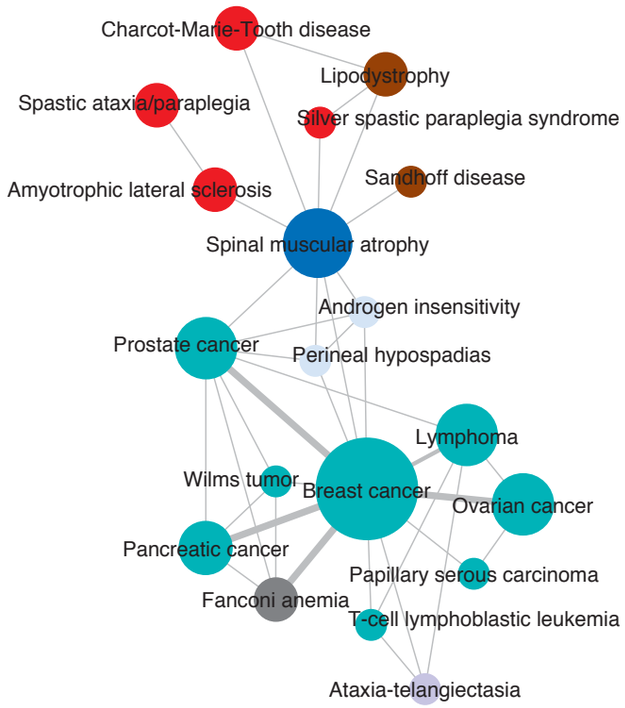
removal of date hubs
(small subnetworks)

removal of party hubs
(intact)

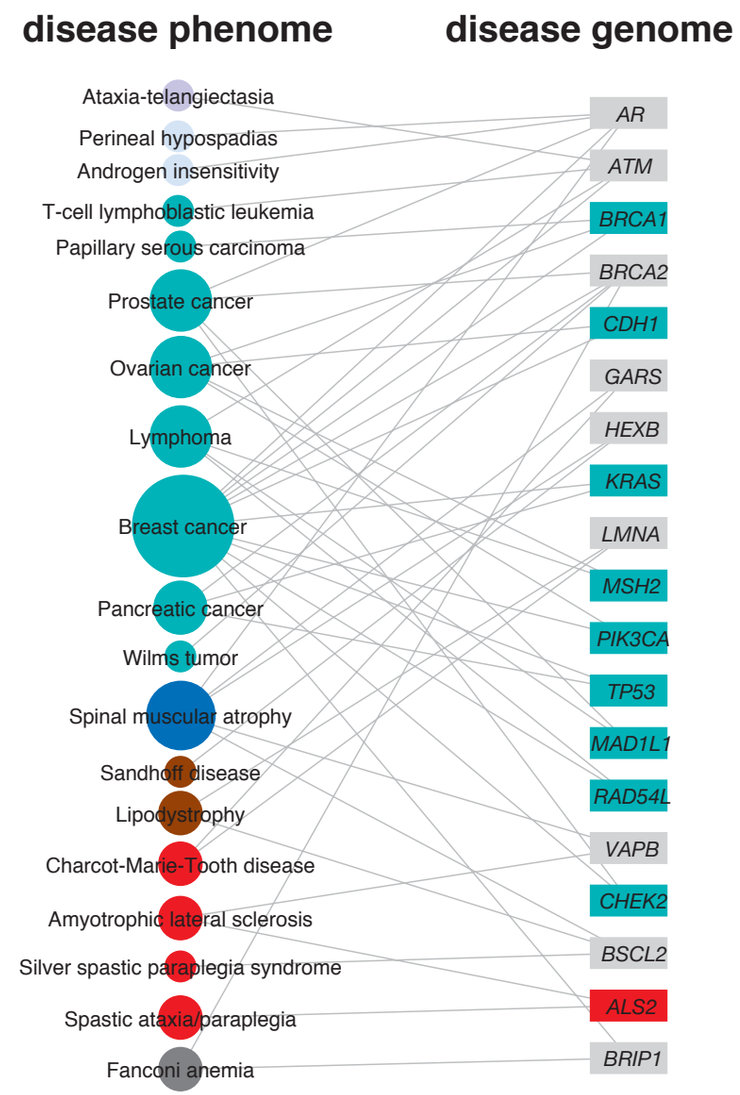
Han et al., Nature 430, p. 88 (2004)

Examples

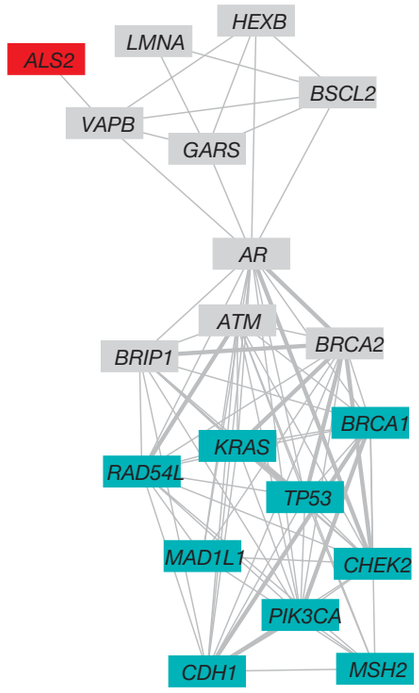
Human Disease Network (HDN)



DISEASOME



Disease Gene Network (DGN)

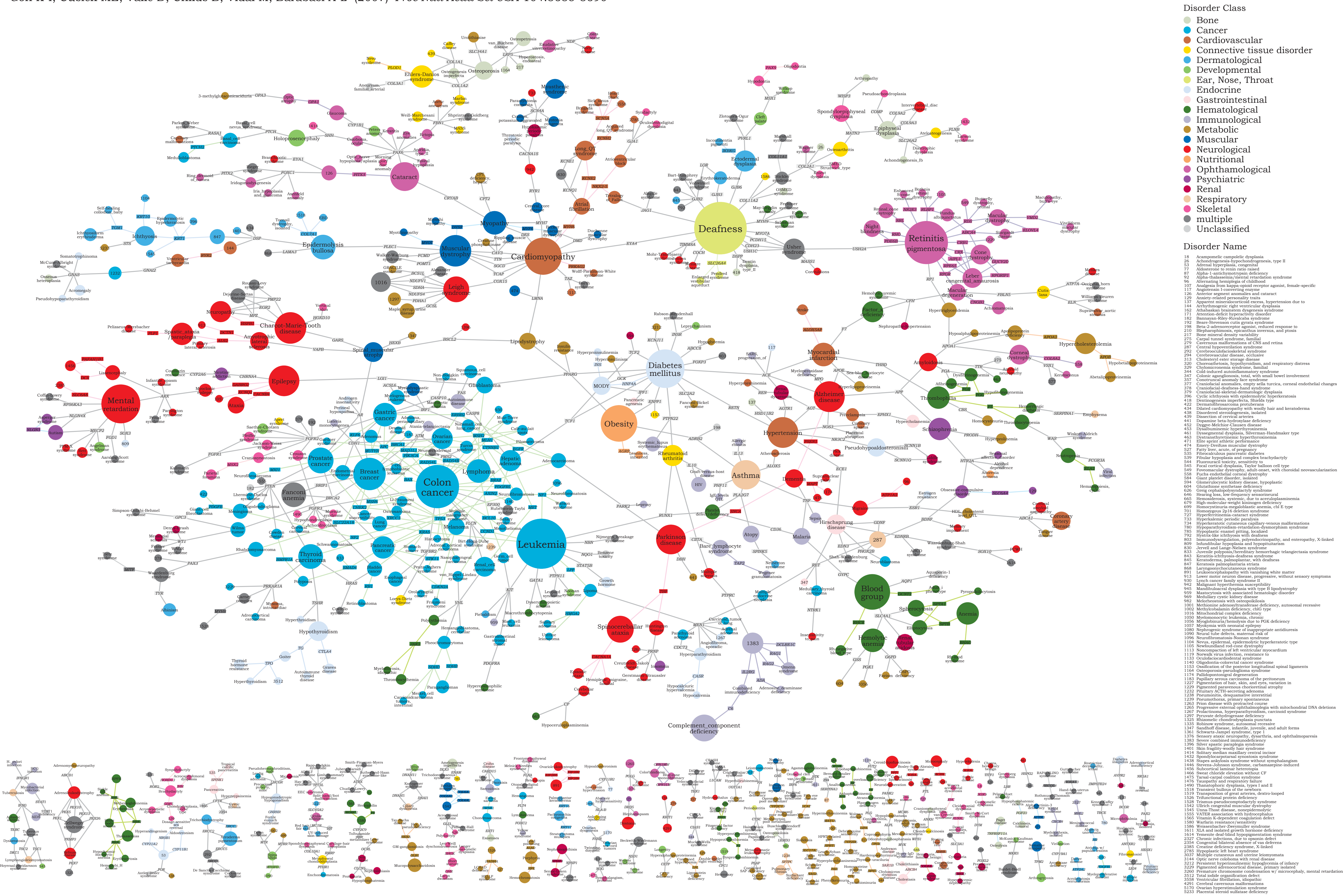


bipartite
OMIM based

Goh et al., PNAS 104(21) p.8685 (2007)

The human disease network

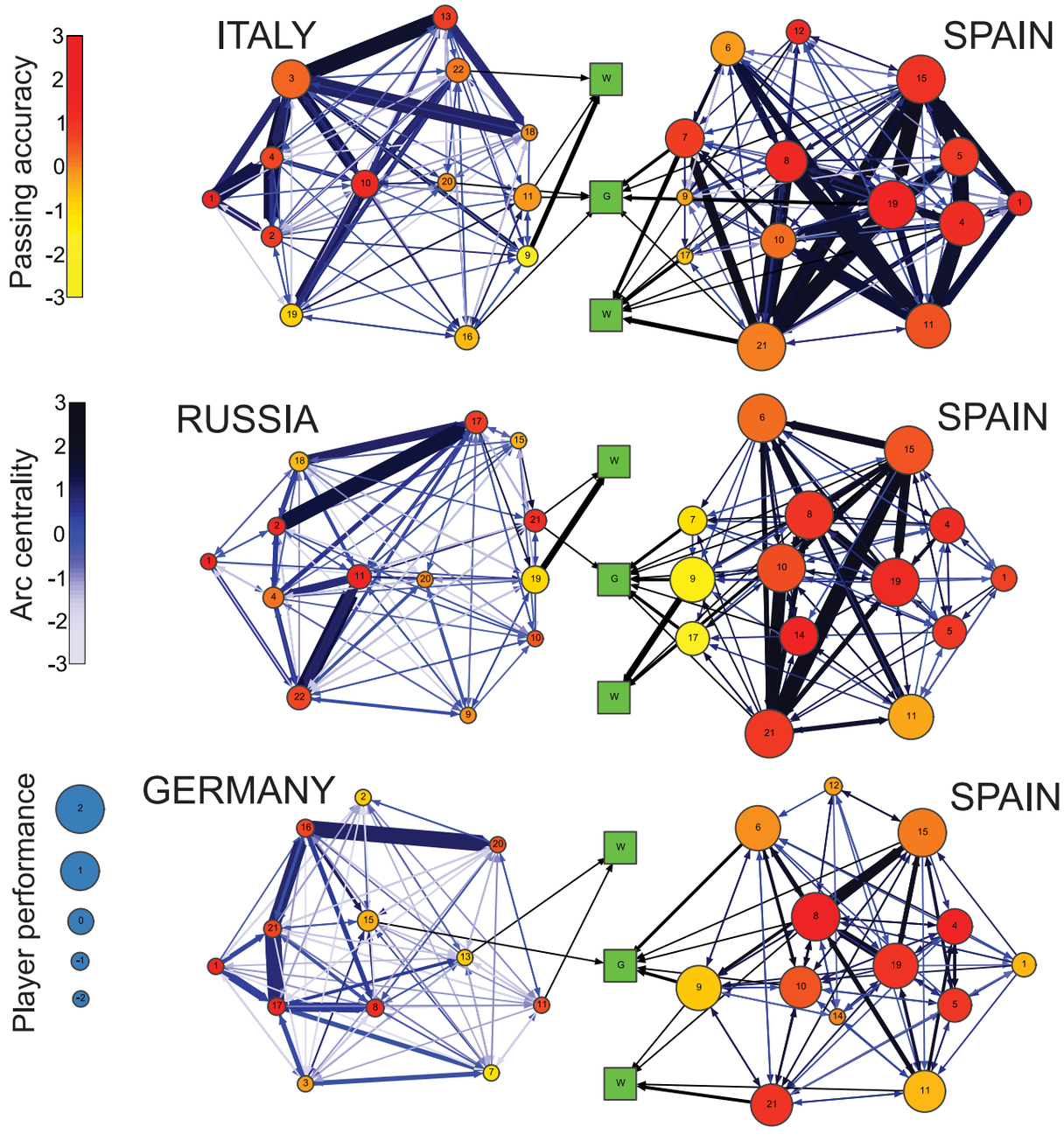
Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L (2007) *Proc Natl Acad Sci USA* 104:8685-8690



Supporting Information Figure 13 | Bipartite-graph representation of the diseasome. A disorder (circle) and a gene (rectangle) are connected if the gene is implicated in the disorder. The size of the circle represents the number of distinct genes associated with the disorder. Isolated disorders (disorders having no links to other disorders) are not shown. Also, only genes connecting disorders are shown.

Other Applications

Quantifying the Performance of Individual Players in a Team Activity (Euro 2008)



Duch, Waitzman1, Amaral, PLoS ONE 5(6) e10937

Gene Ontology

Ontology

Definition

- “A set of concepts and categories in a subject area or domain that shows their properties and the relations between them.” *New Oxford American Dictionary, Oxford University Press 2013*

Gene Ontology

Gene Ontology (GO)

“a computational representation of our evolving knowledge of how genes encode biological functions at the molecular, cellular and tissue system levels”

- Controlled vocabulary of terms
  - Describe gene product characteristics
  - Gene product annotation data
- Tools for GO

http://geneontology.org

Gene Ontology

Controlled vocabularies

- **Cellular Component (CC),** the parts of a cell or its extracellular environment
- **Molecular Function (MF)**, the elemental activities of a gene product at the molecular level, such as binding or catalysis
- **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.

http://geneontology.org (4/2016)

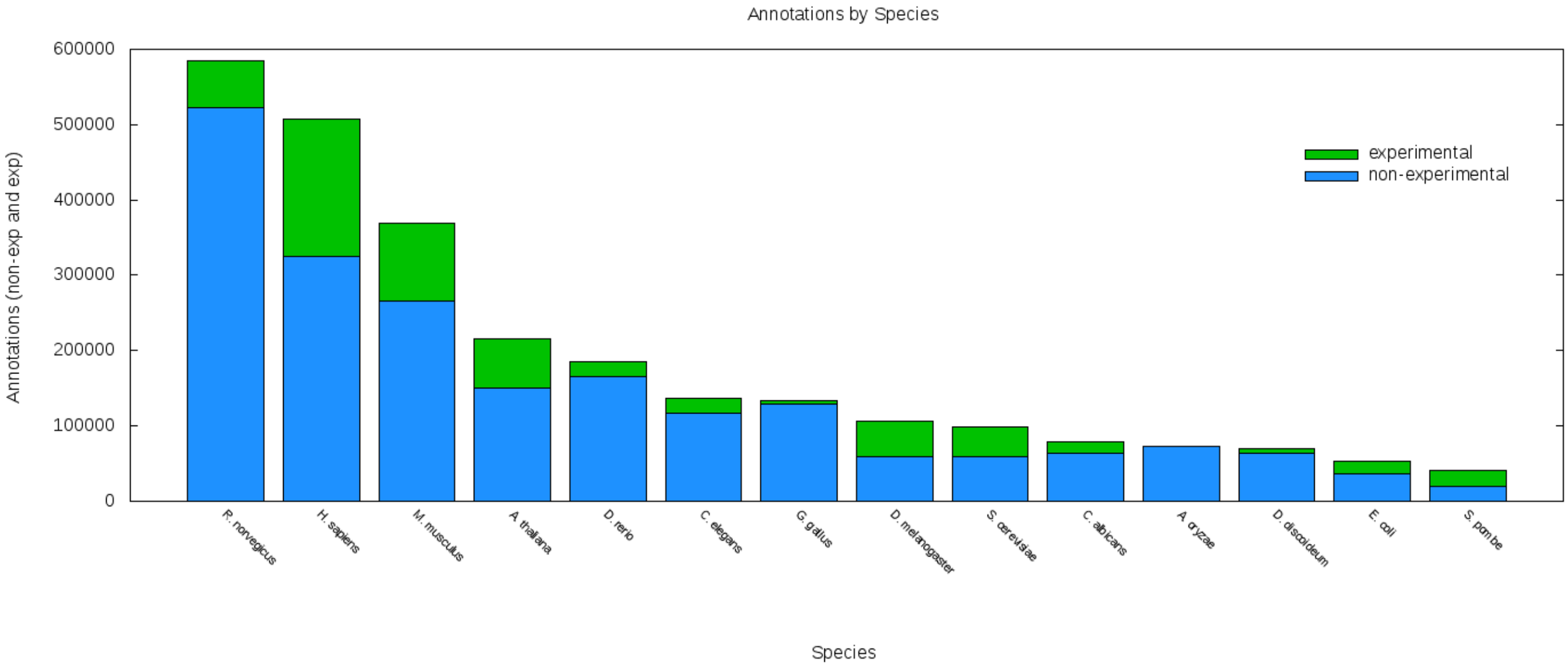
Gene Ontology

Example

- gene product "cytochrome c"
  - **Molecular Function**
    - "oxidoreductase activity”
  - **Biological Process**
    - "oxidative phosphorylation"
    - "induction of cell death”
  - **Cellular Component**
    - “mitochondrial matrix"
    - "mitochondrial inner membrane".

http://geneontology.org (4/2016)

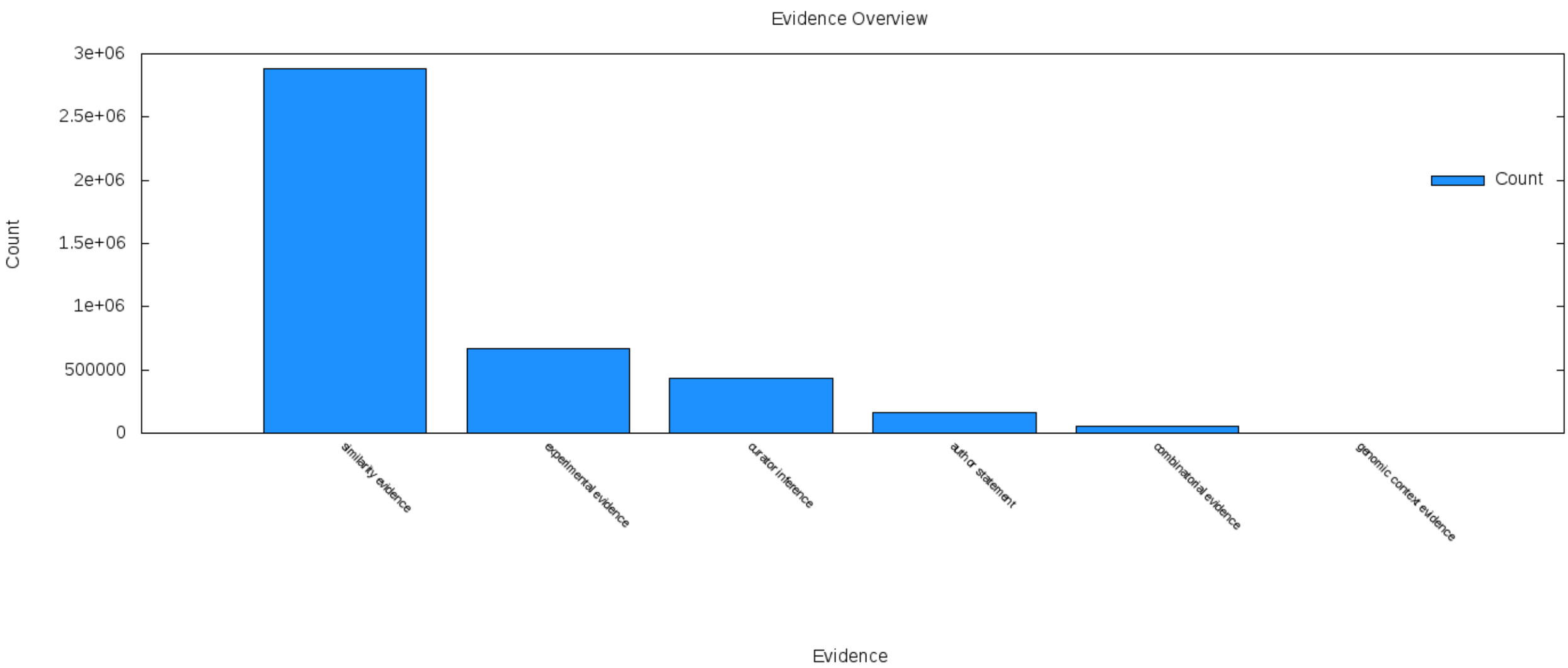
Gene Ontology



Note: GO vocabulary is designed to be species-agnostic, and includes terms applicable to prokaryotes and eukaryotes, and single and multicellular organisms.

http://geneontology.org (4/2016)

Gene Ontology



http://geneontology.org (4/2016)

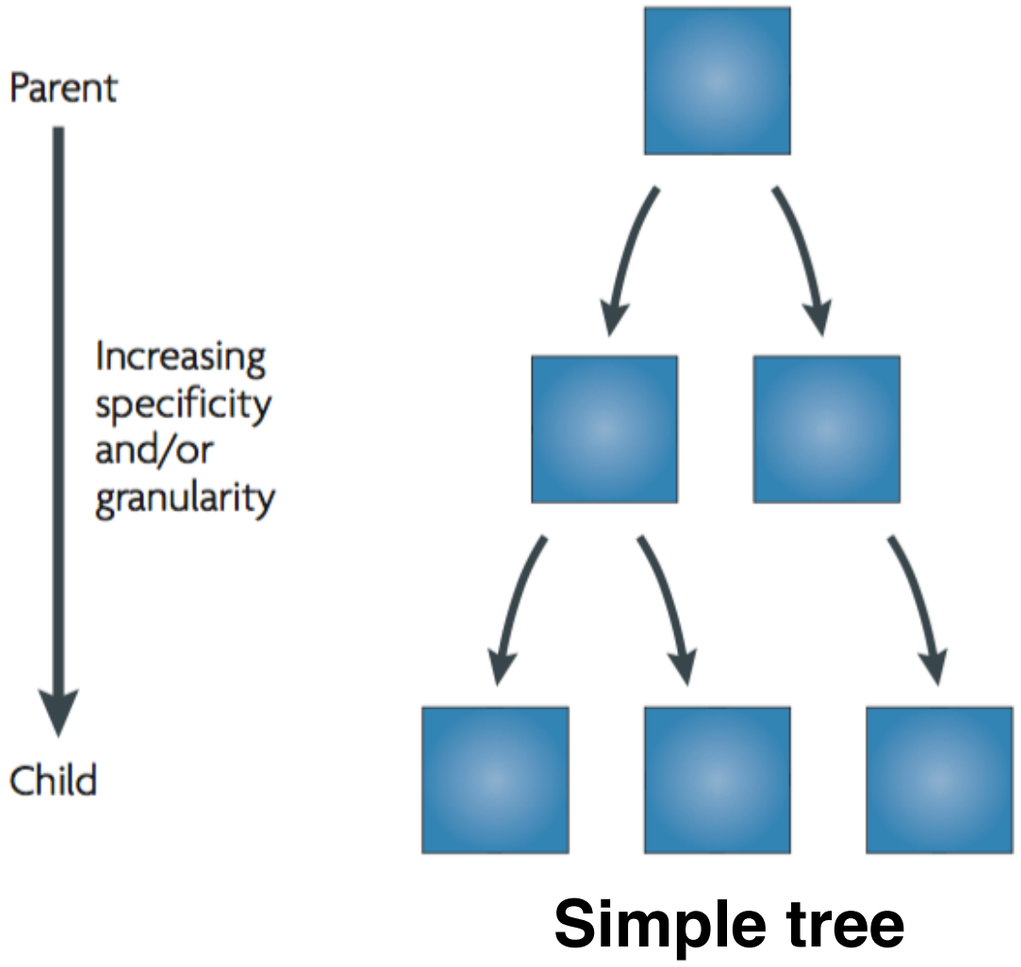
Gene Ontology

evidence code	evidence code description	source of evidence	Manually checked
IDA	Inferred from direct assay	Experimental	Yes
IEP	Inferred from expression pattern	Experimental	Yes
IGI	Inferred from genetic interaction	Experimental	Yes
IMP	Inferred from mutant phenotype	Experimental	Yes
IPI	Inferred from physical interaction	Experimental	Yes
ISS	Inferred from sequence or structural similarity	Computational	Yes
RCA	Inferred from reviewed computational analysis	Computational	Yes
IGC	Inferred from genomic context	Computational	Yes
IEA	Inferred from electronic annotation	Computational	No
IC	Inferred by curator	Indirectly derived from experimental or computational evidence made by a curator	Yes
TAS	Traceable author statement	Indirectly derived from experimental or computational evidence made by the author of the published article	Yes
NAS	Non-traceable author statement	No ‘source of evidence’ statement given	Yes
ND	No biological data available	No information available	Yes
NR	Not recorded	Unknown	Yes

Rhee et al Nature Reviews Genetics 9, 509-515 (2008)

Gene Ontology

GO ontology structured as directed acyclic graph



Simple tree

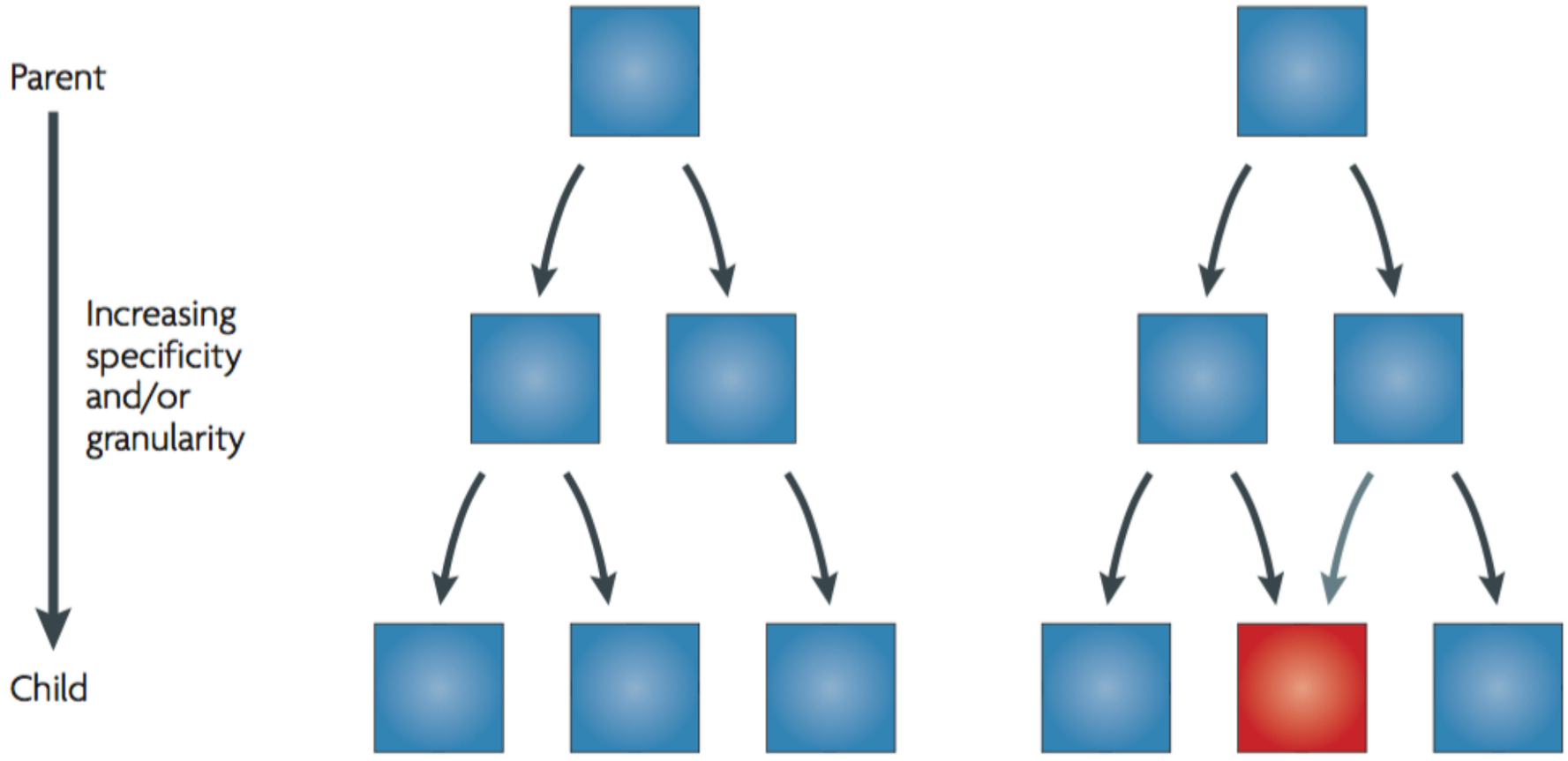
Simple tree

- each child has only one parent
- the edges are directed

Rhee et al Nature Reviews Genetics 9, 509-515 (2008)
http://geneontology.org

Gene Ontology

GO ontology structured as **directed acyclic graph.**



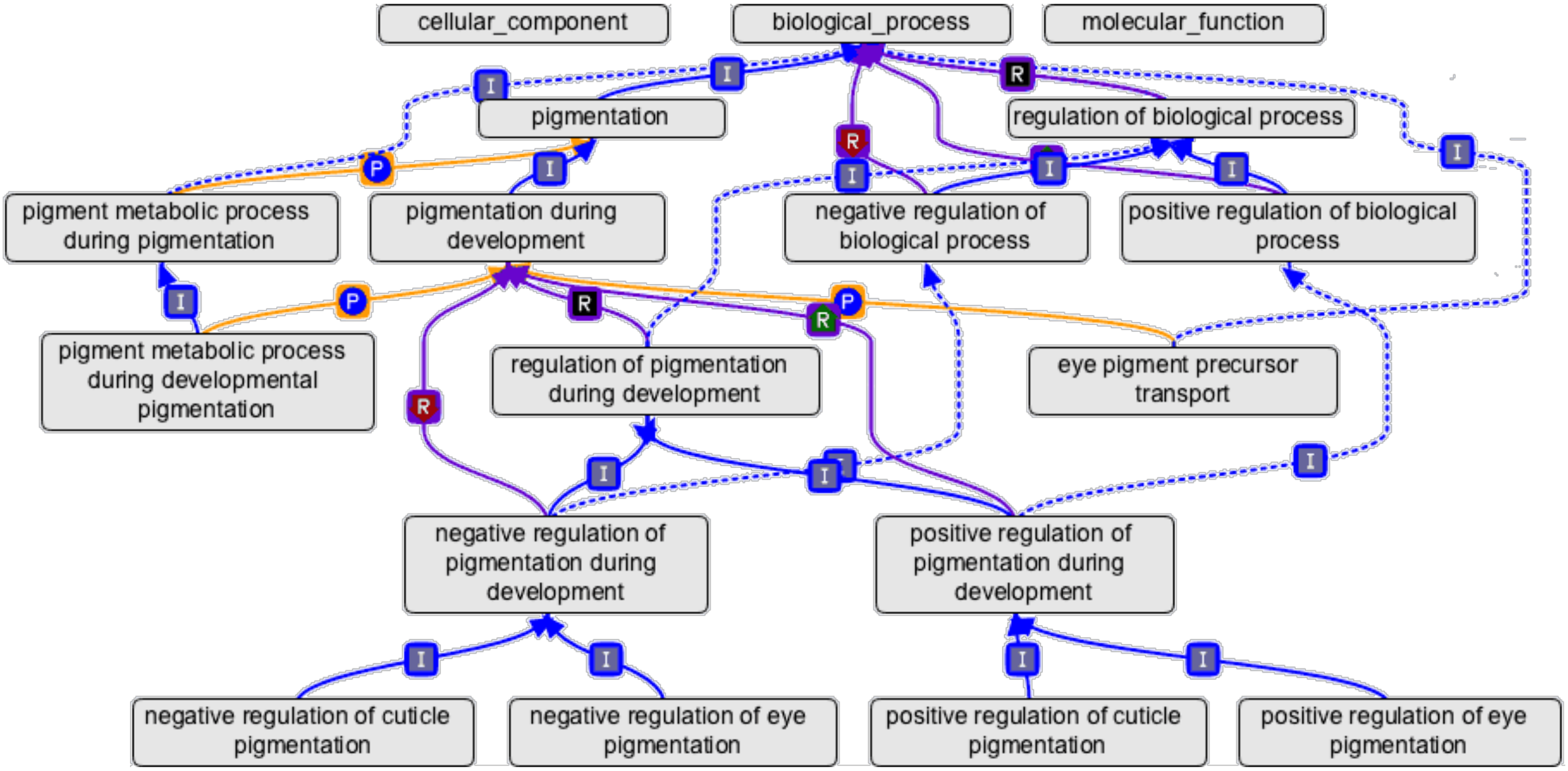
Simple tree

A directed acyclic graph (DAG)

each child can have one or more parents.

Rhee et al Nature Reviews Genetics 9, 509-515 (2008)
http://geneontology.org

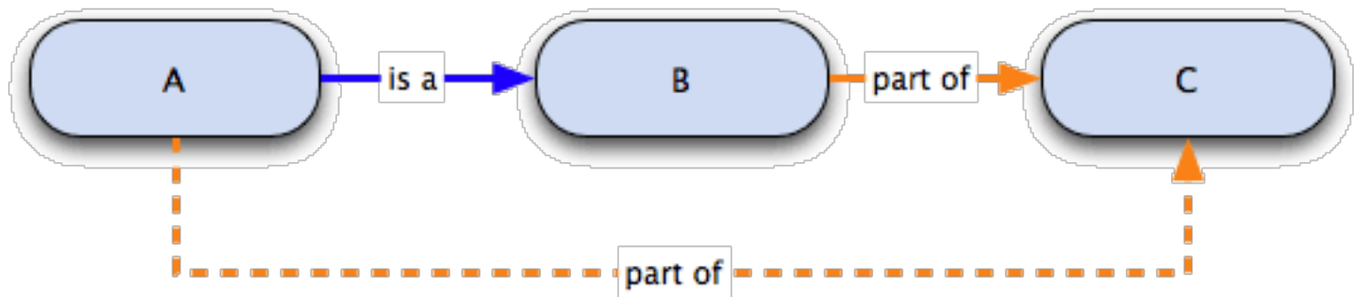
Gene Ontology



http://geneontology.org

Each term can have relationships to one or more other terms in the same or other domains

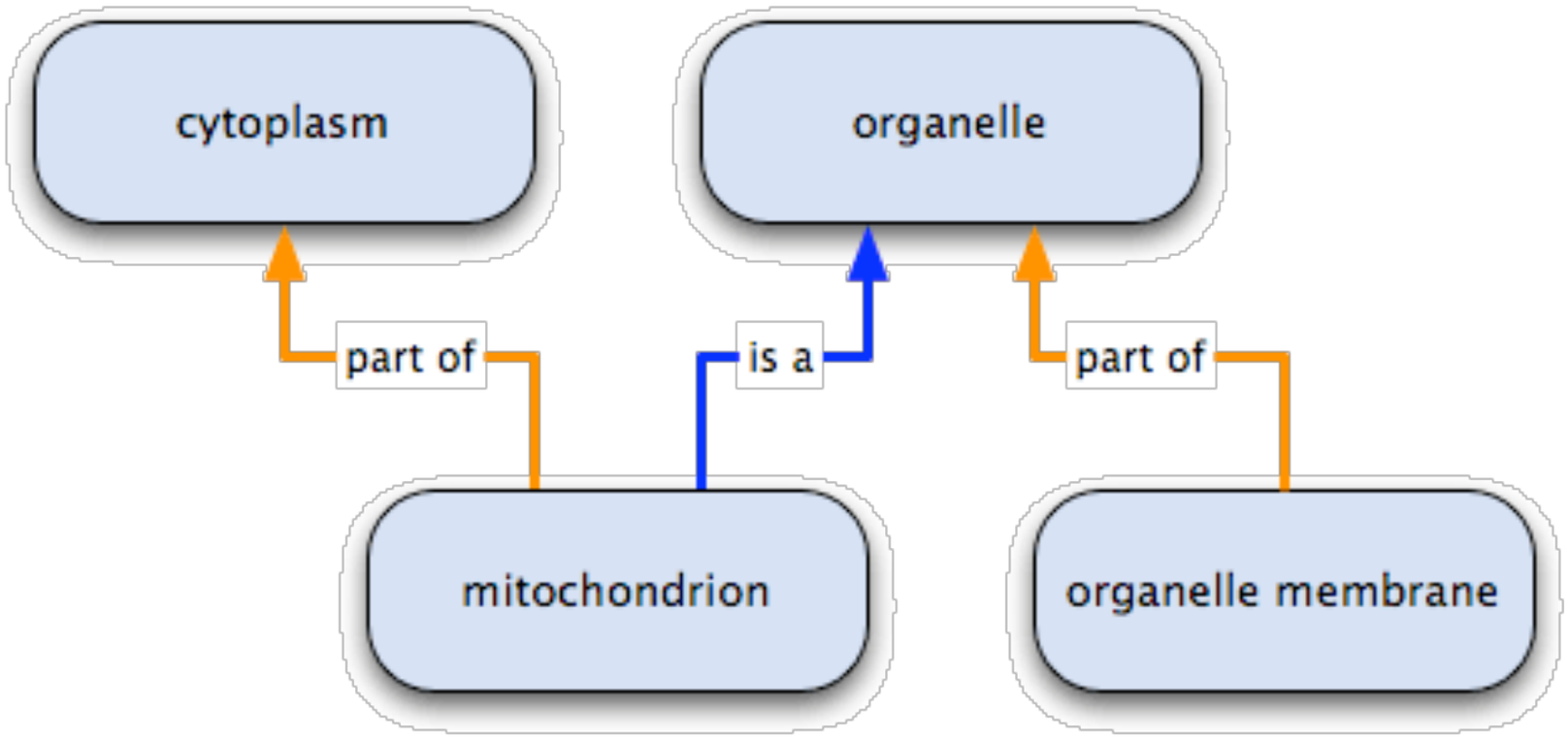
Gene Ontology



- A *is a* B
- B is *part of* C
- we can infer that A is *part of* C

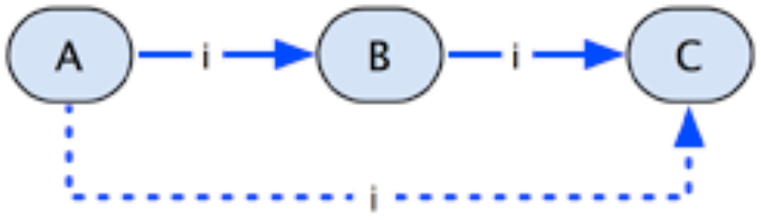
http://geneontology.org (4/2016)

Gene Ontology



http://geneontology.org (4/2016)

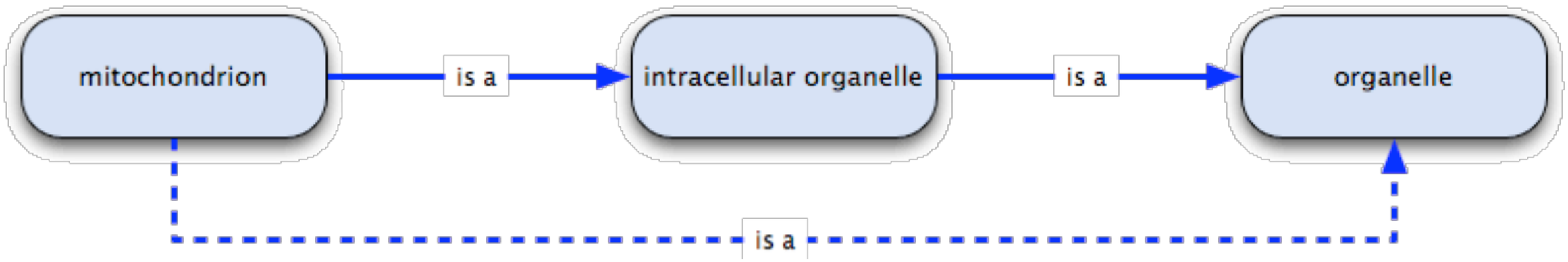
Gene Ontology



The *is a* relation is **transitive:**

If A *is a* B & B *is a* C

We can infer that A *is a* C

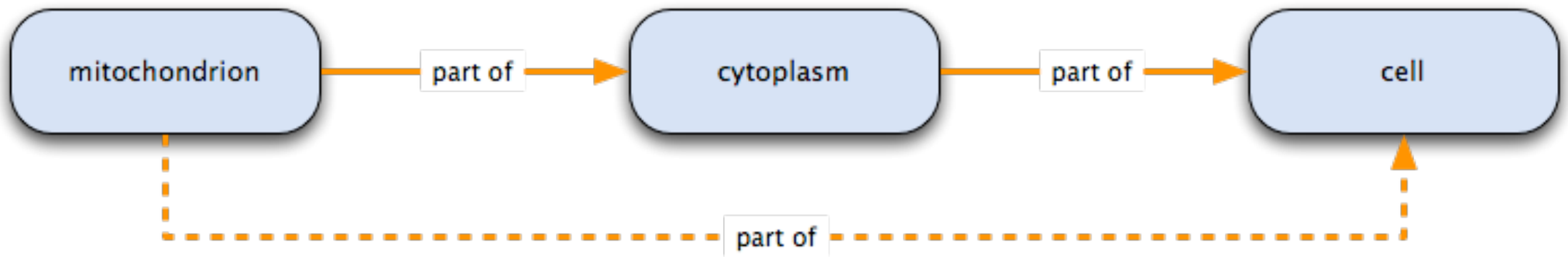


mitochondrion *is an* intracellular organelle and intracellular organelle *is an* organelle therefore mitochondrion *is an* organelle.

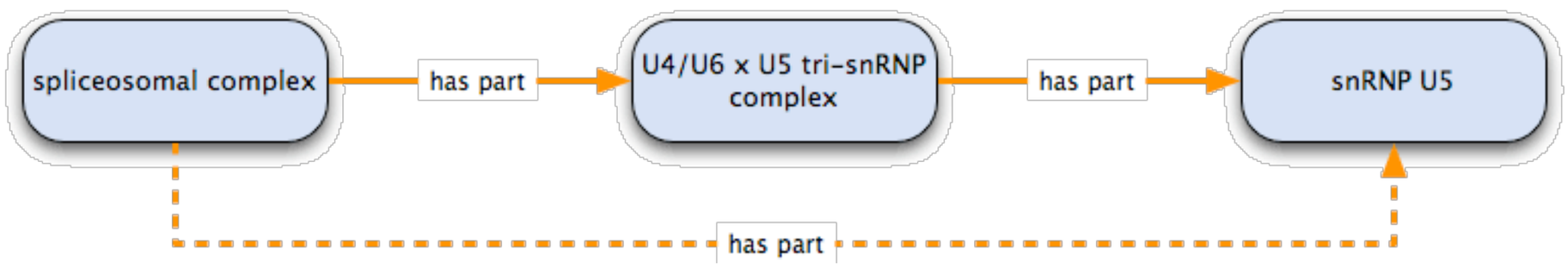
http://geneontology.org (4/2016)

Gene Ontology

part of



has part

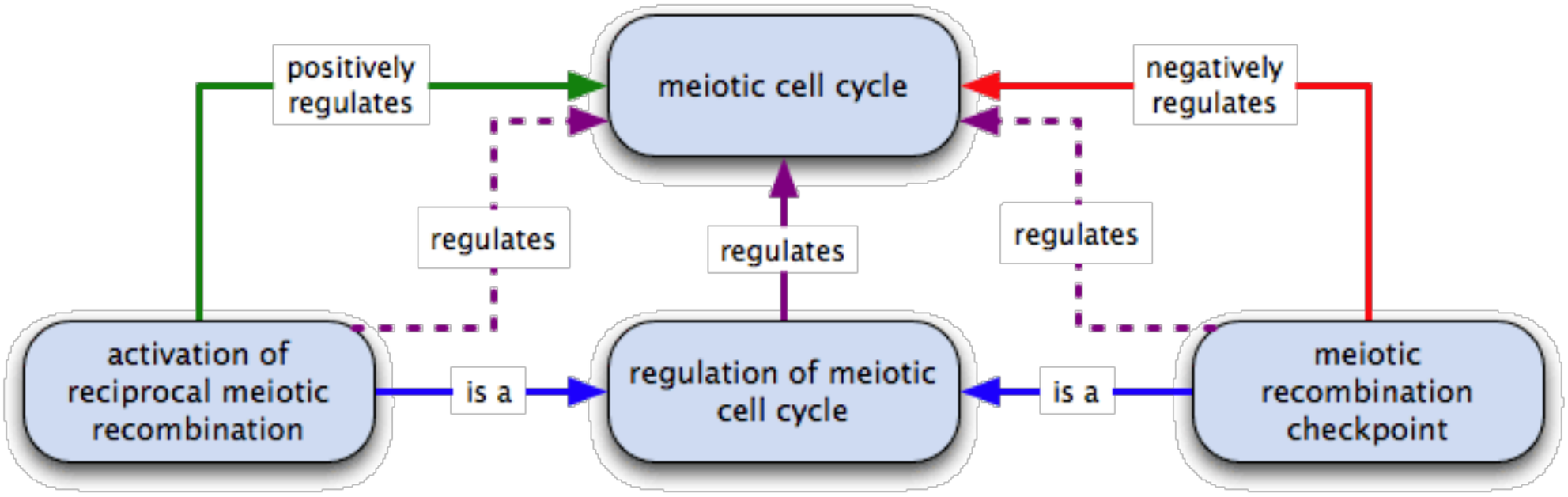


- spliceosomal complex has part U4/U6 x U5 tri-snNRP complex
- U4/U6 x U5 tri-snNRP complex has part snRNP U5
- therefore spliceosomal complex has part snRNP U5

http://geneontology.org (4/2016)

Gene Ontology

regulates

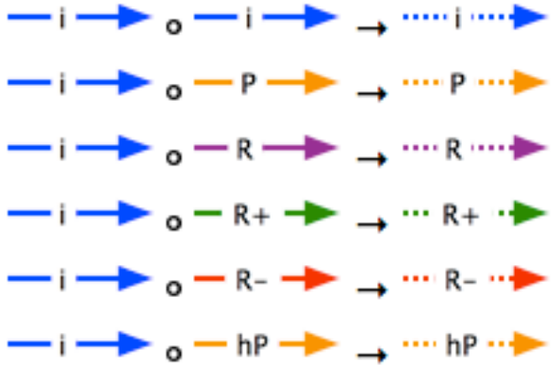


http://geneontology.org (4/2016)

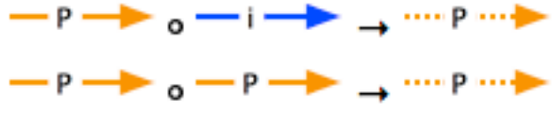
Gene Ontology

Inferences

is a



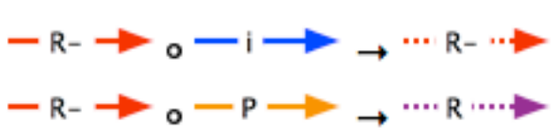
part of



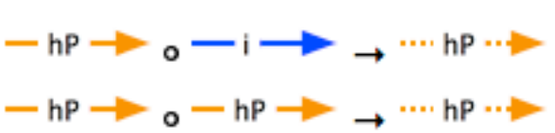
regulates

positively regulates

negatively regulates

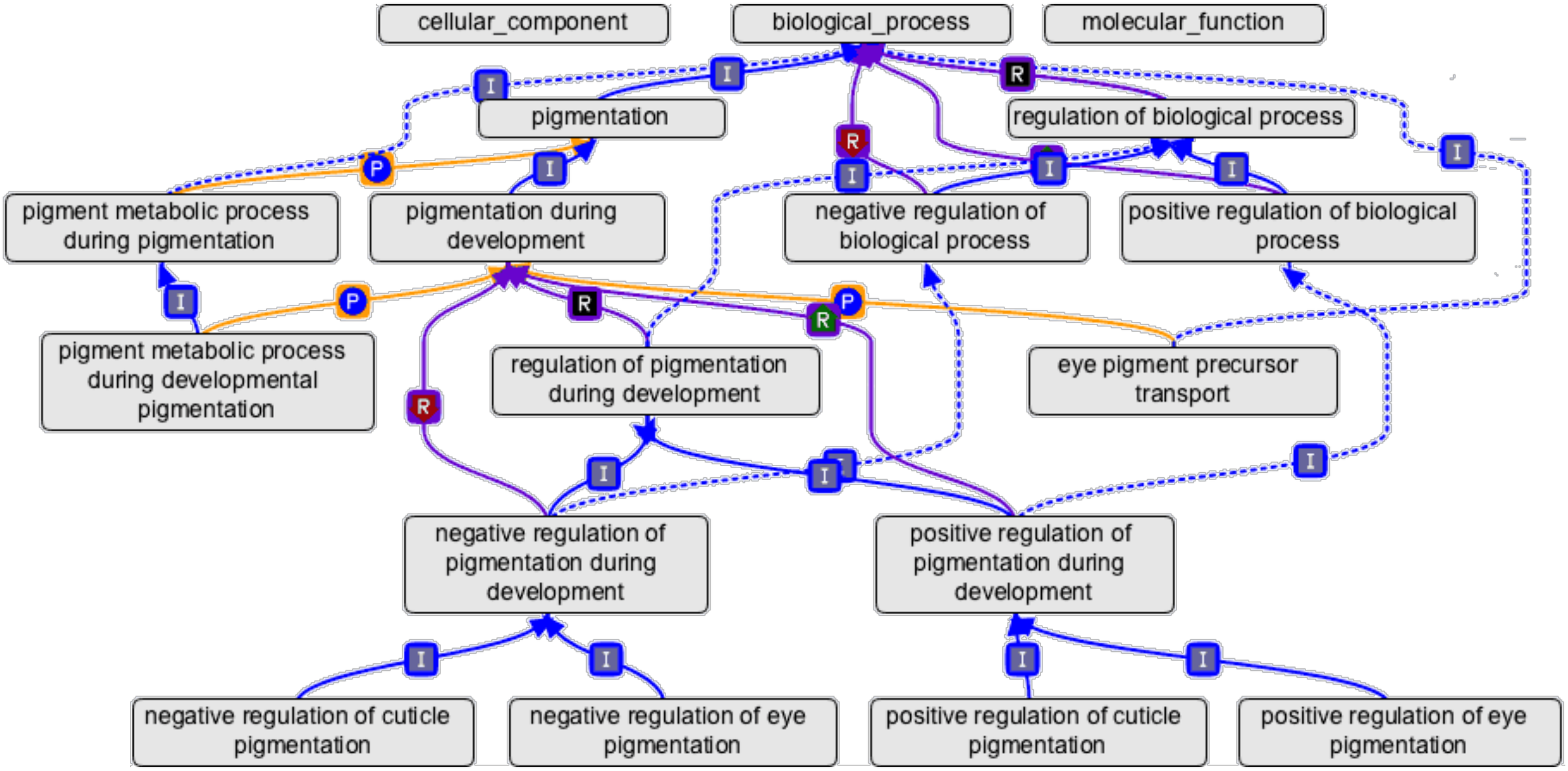


has part



http://geneontology.org (4/2016)

Gene Ontology



http://geneontology.org

Each term can have relationships to one or more other terms in the same or other domains

Gene Ontology

Ten Quick Tips for Using the Gene Ontology

1. Know the Source of the GO Annotations You Use
2. Understand the Scope of GO Annotations
3. Consider Differences in Evidence Codes
4. Probe Completeness of GO Annotations
5. Understand the Complexity of the GO Structure
6. Choose Analysis Tools Carefully
7. Provide the Version of the Data/Tools Used
8. Seek Input from the GOC Community and Make Use of GOC Resources
9. Contribute to the GO
10. Acknowledge the Work of the GO Consortium

Blake PLoS Comput Biol 9(11): e1003343 (2013)

Gene Ontology

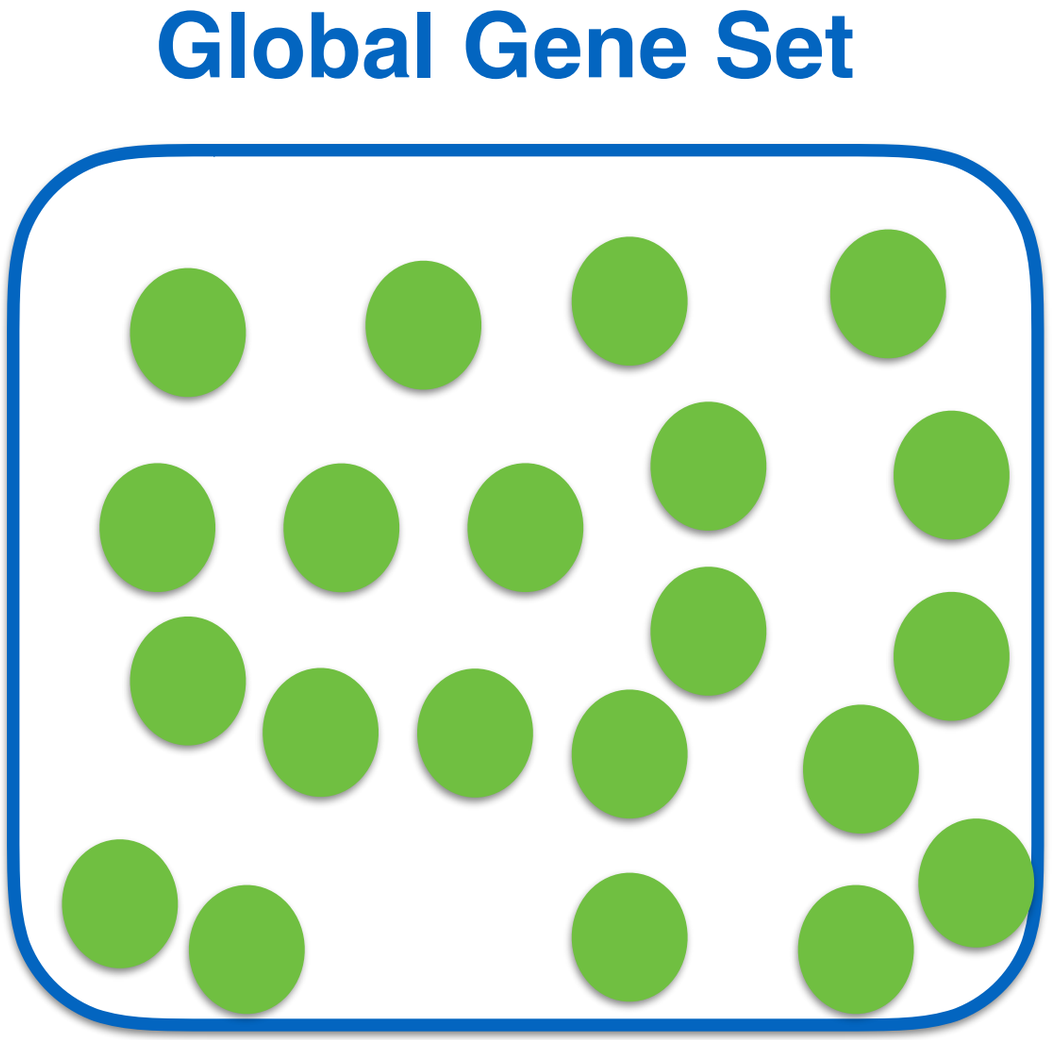
GO Enrichment Analysis

For a set of genes up/down-regulated find which GO terms are over-represented (or under-represented) using annotations for that gene set

http://geneontology.org

Gene Ontology

Basic Enrichment Analysis

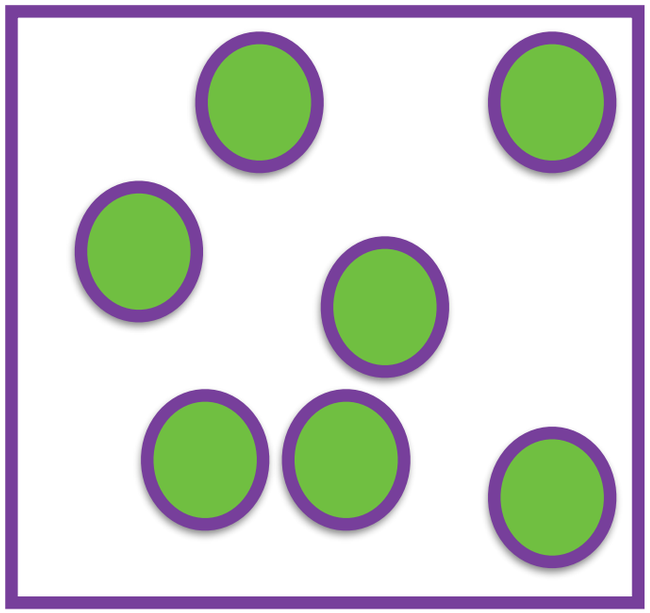


http://geneontology.org

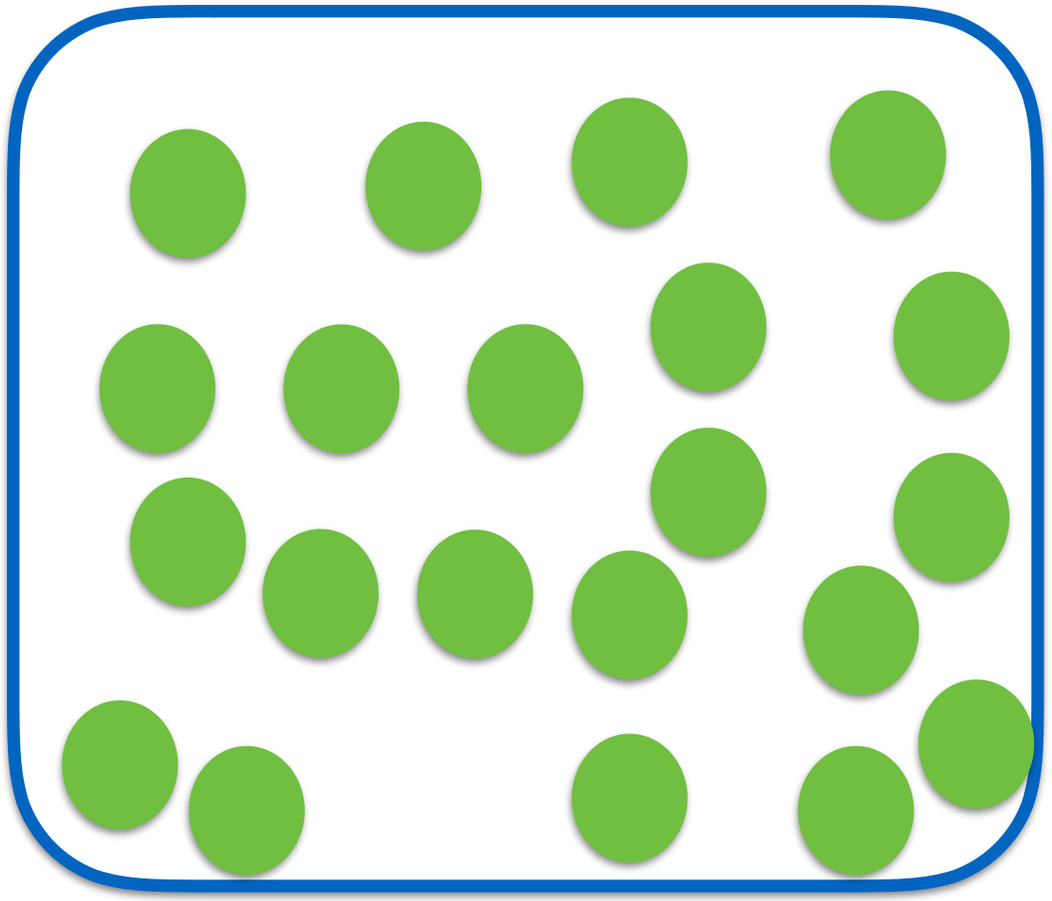
Gene Ontology

Basic Enrichment Analysis

GO Category or Pathway of interest



Global Gene Set

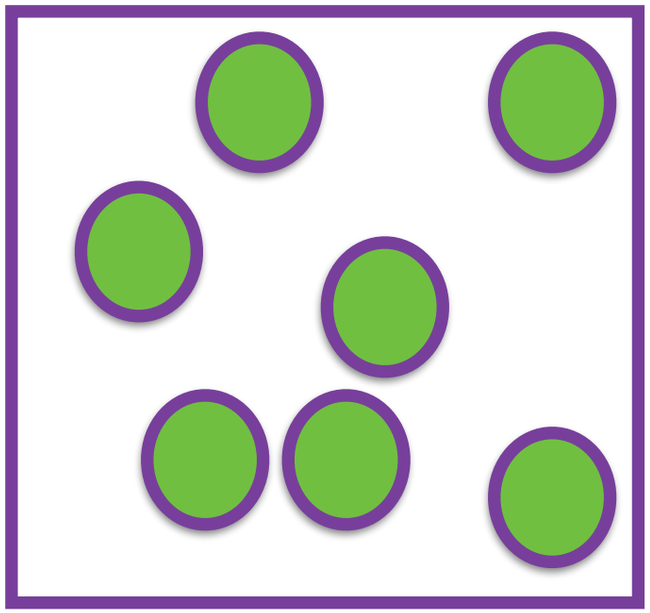


http://geneontology.org

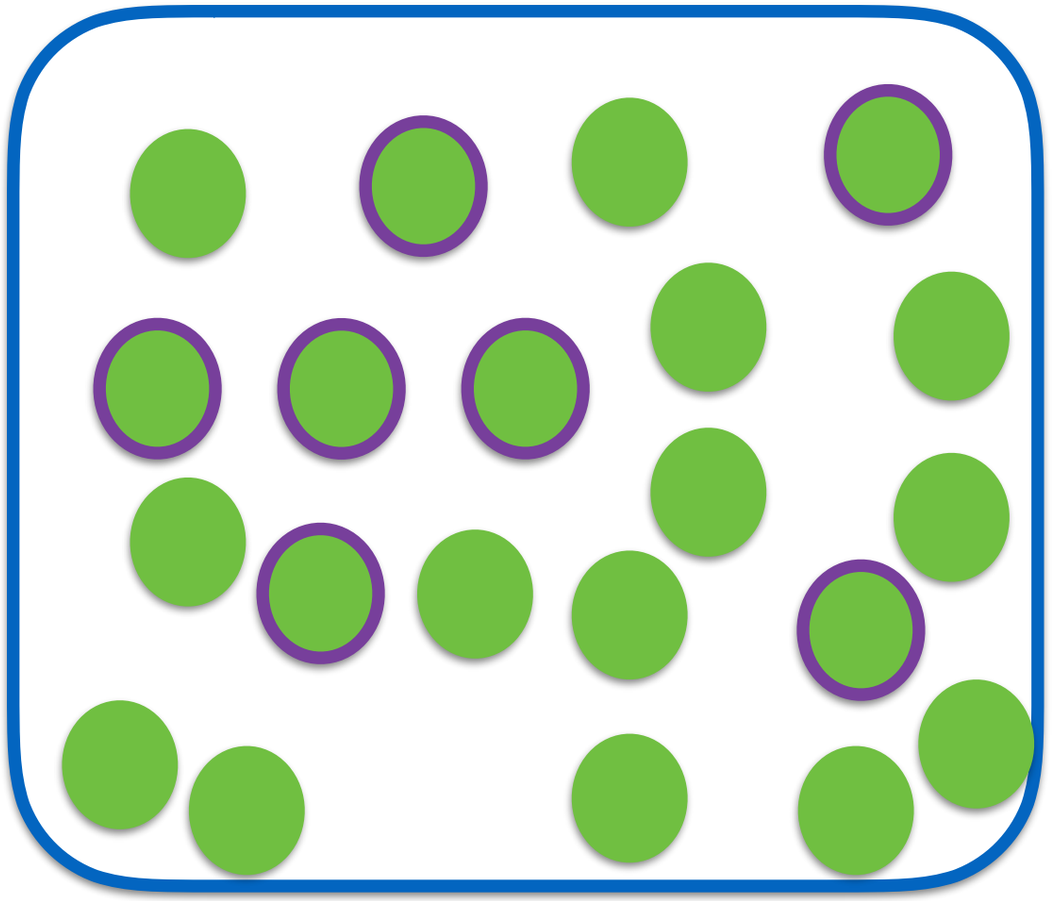
Gene Ontology

Basic Enrichment Analysis

GO Category or Pathway of interest



Global Gene Set

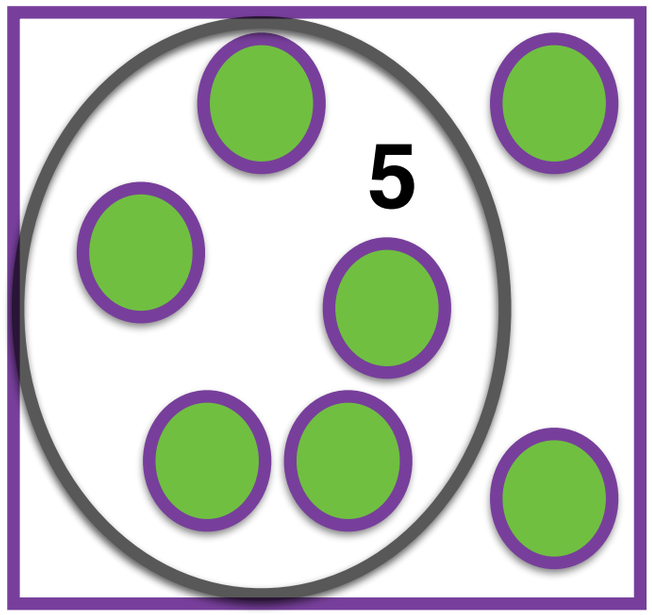


http://geneontology.org

Gene Ontology

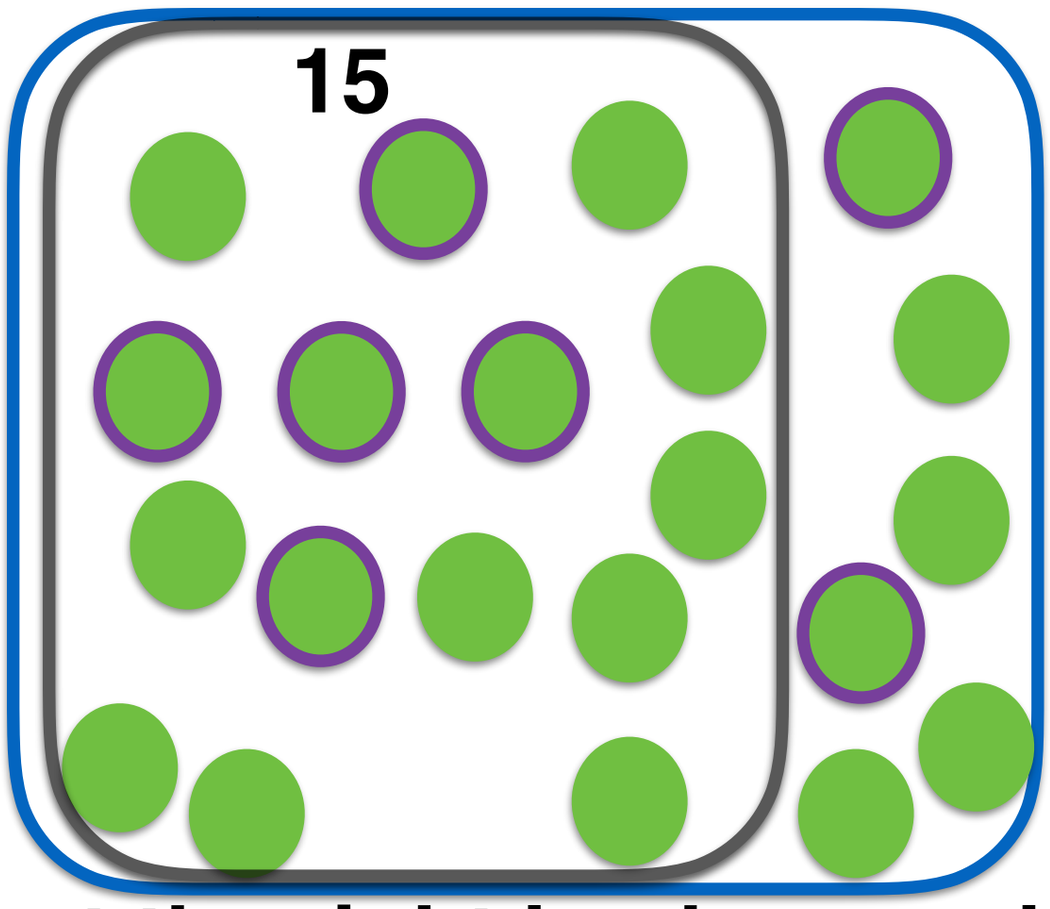
Basic Enrichment Analysis

GO Category or Pathway of interest



query/tested

Global Gene Set



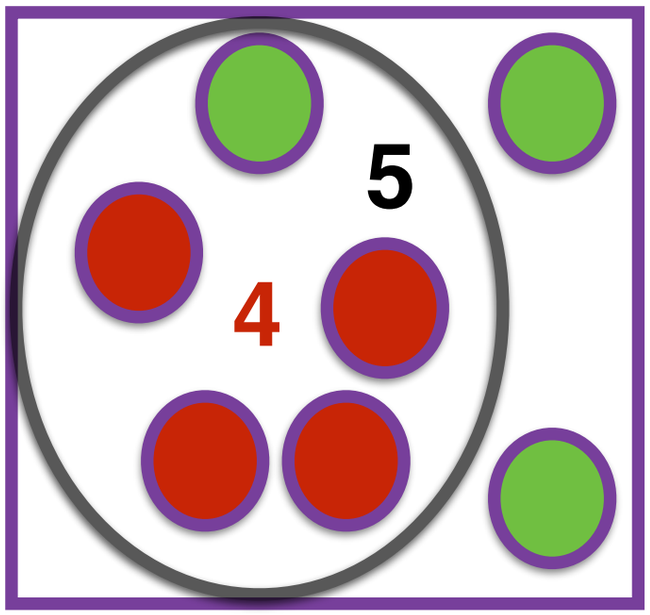
select the right background set
e.g. right organism

http://geneontology.org

Gene Ontology

Basic Enrichment Analysis

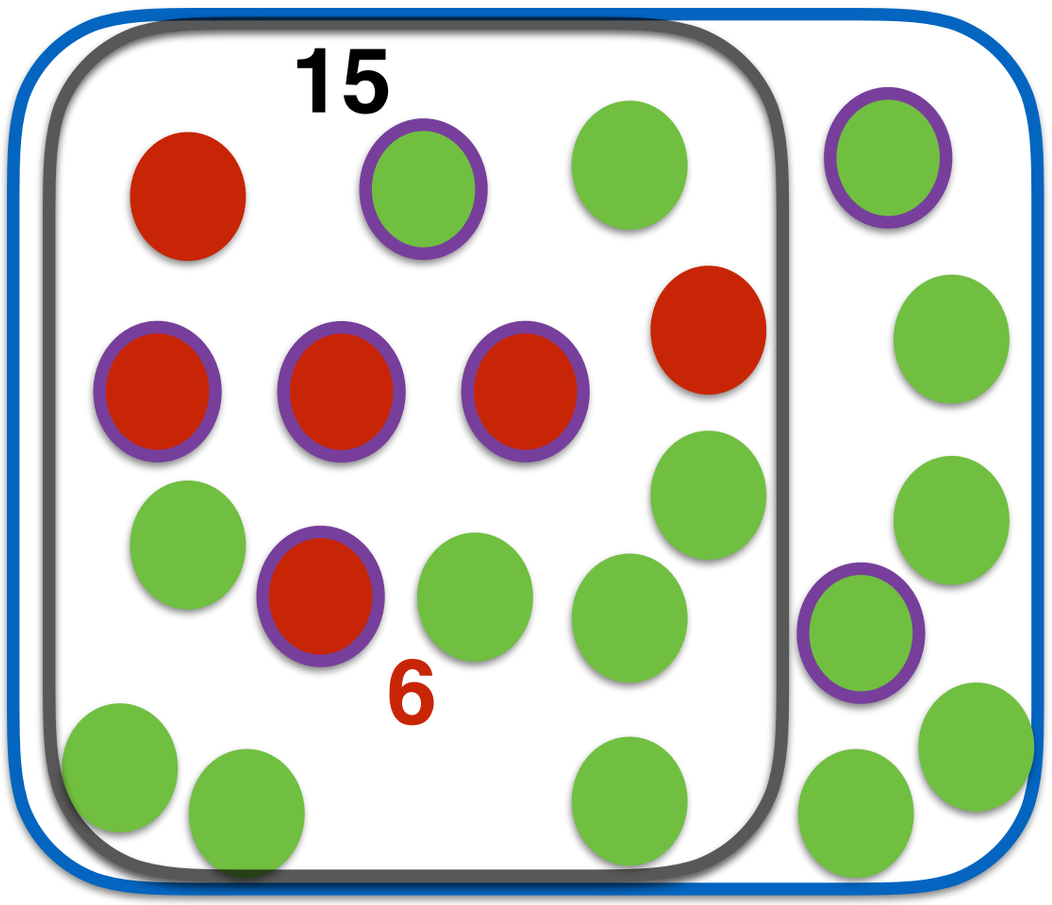
GO Category or Pathway of interest



Experimentally tested

Interesting behavior
E.g. Unregulated

Global Gene Set

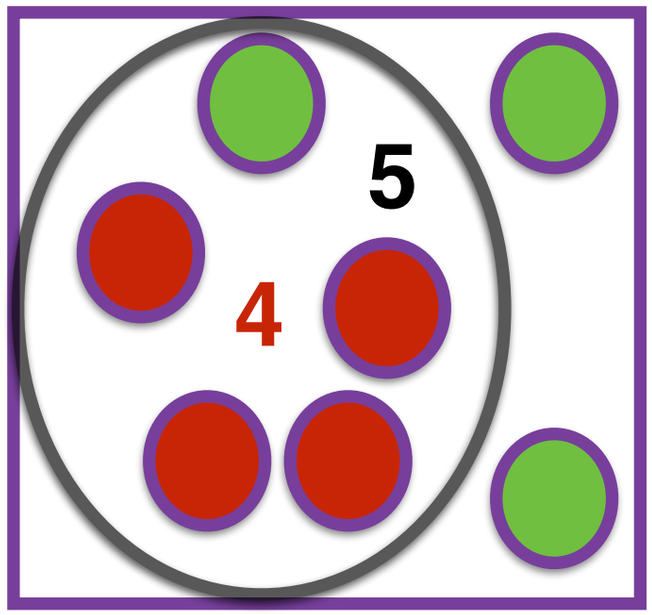


http://geneontology.org

Gene Ontology

Basic Enrichment Analysis

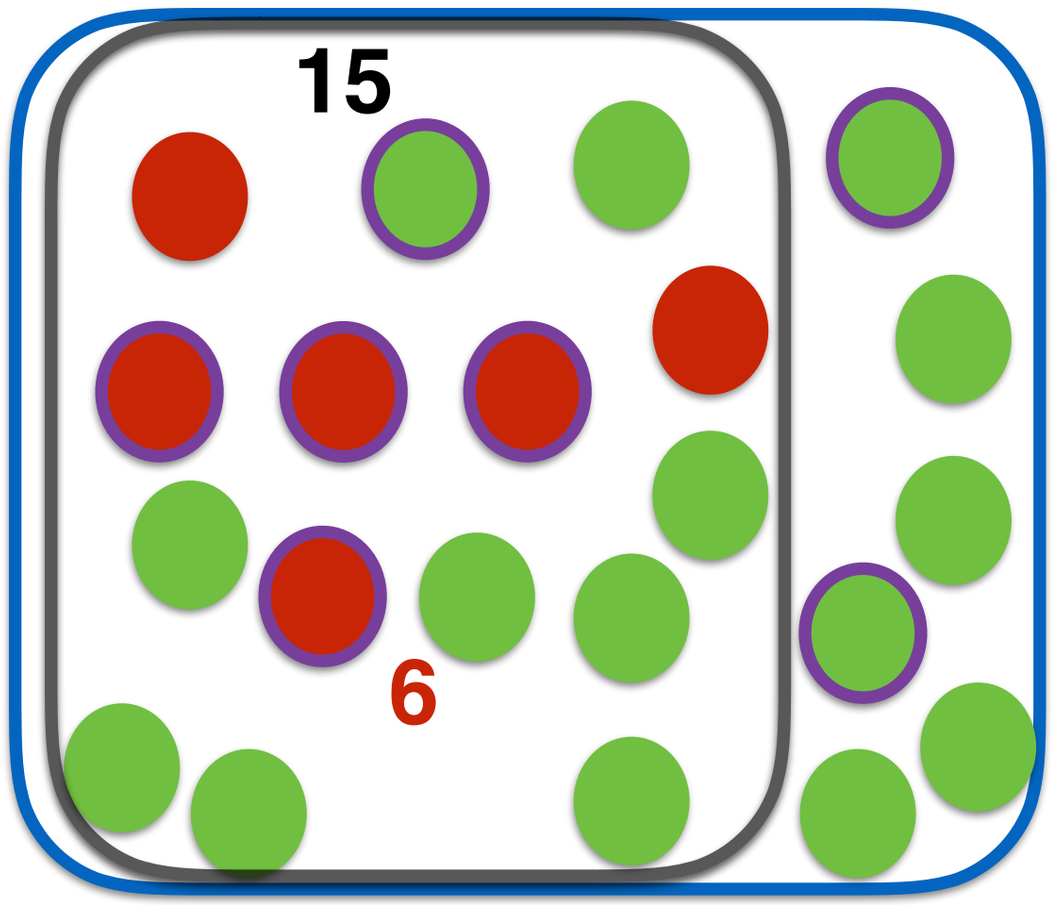
GO Category or Pathway of interest



Experimentally tested

Interesting behavior
E.g. Unregulated

Global Gene Set



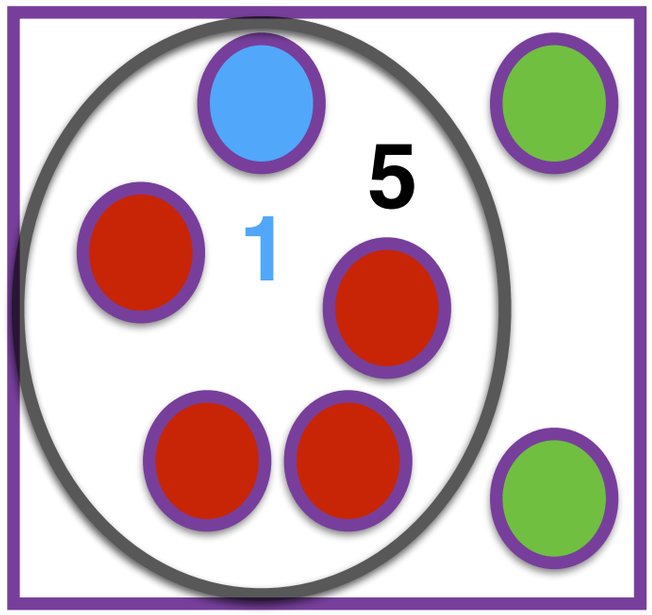
4 out of 6 Vs **5 out of 15**

http://geneontology.org

Gene Ontology

Basic Enrichment Analysis

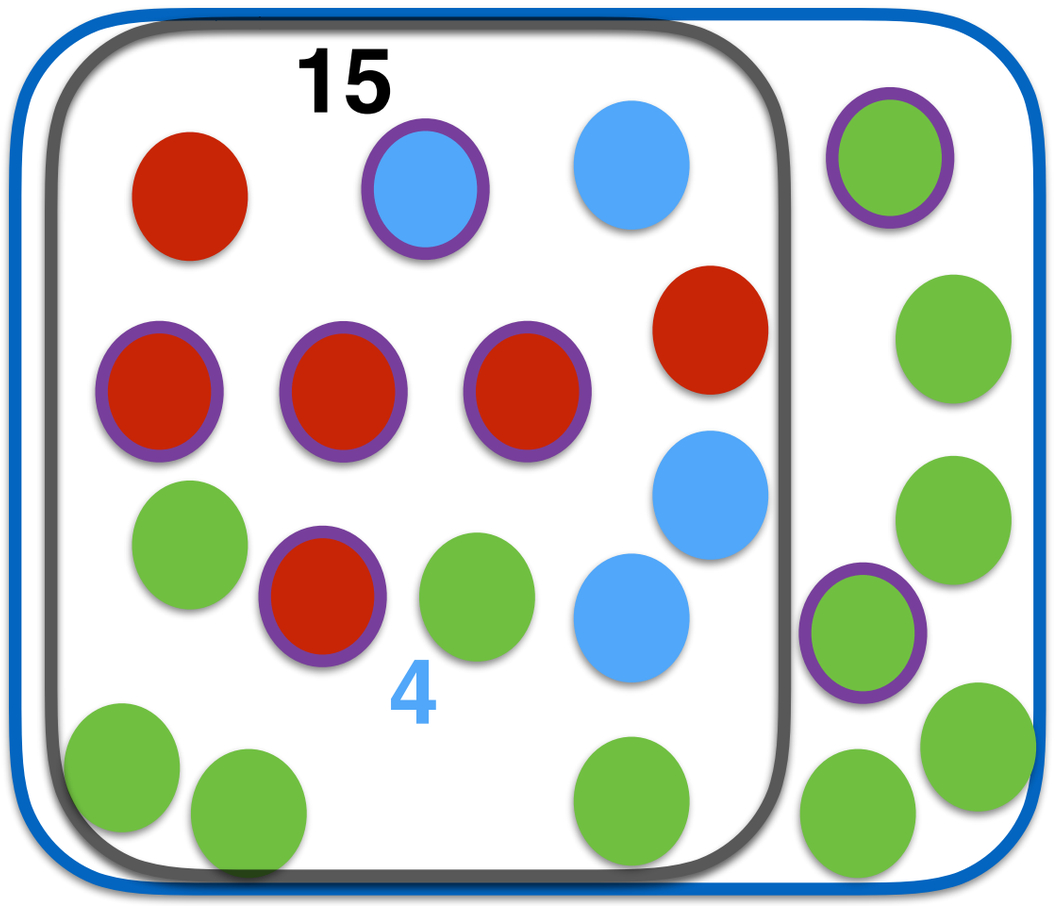
GO Category or Pathway of interest



Experimentally tested

Interesting behavior
E.g. Downregulated

Global Gene Set



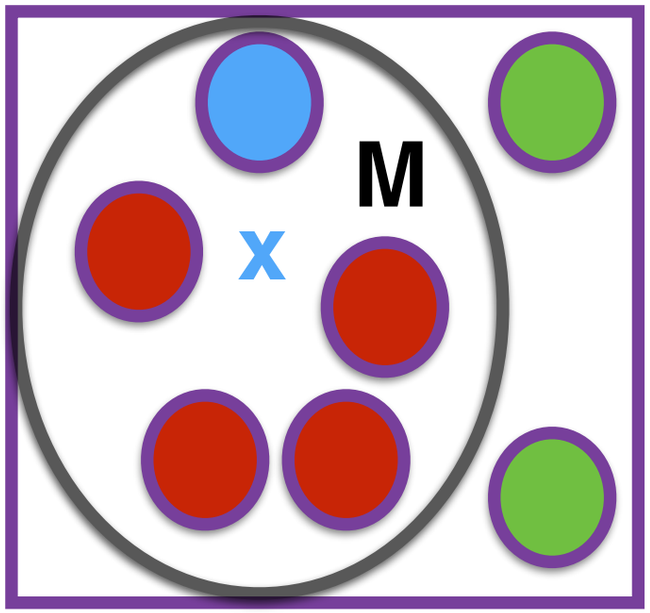
1 out of 4 Vs **5 out of 15**

http://geneontology.org

Gene Ontology

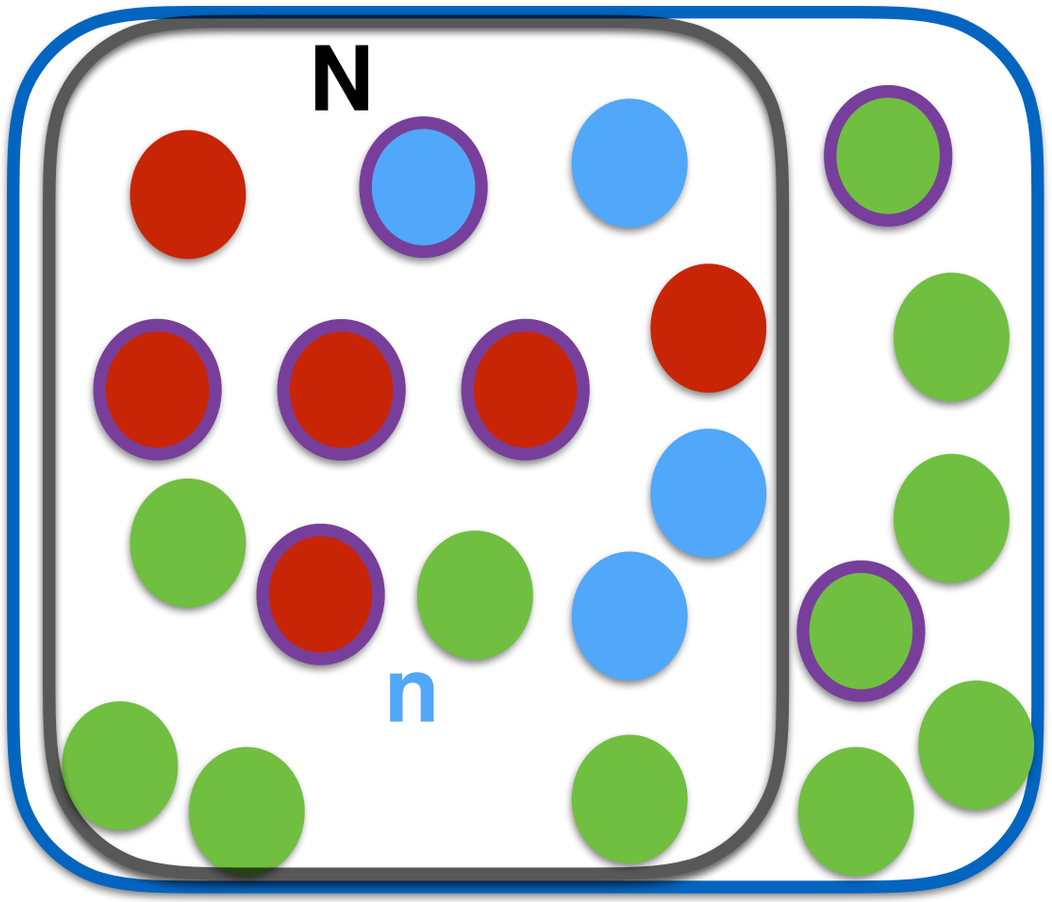
Basic Enrichment Analysis

GO Category or Pathway of interest



Experimentally tested

Global Gene Set



Interesting behavior

x out of n Vs **M out of N**

http://geneontology.org

Gene Ontology

Basic Enrichment Analysis

probability of getting at least x successes, i.e. x or more genes in the GO category (hypergeometric function sum)

$$p = \sum_{i=x}^{n} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

- N genes in big set
- M genes in category in big set
- n genes of interest
- looking of x or more hits

where $\binom{k}{l} = \frac{k!}{l!(k-l)!}$

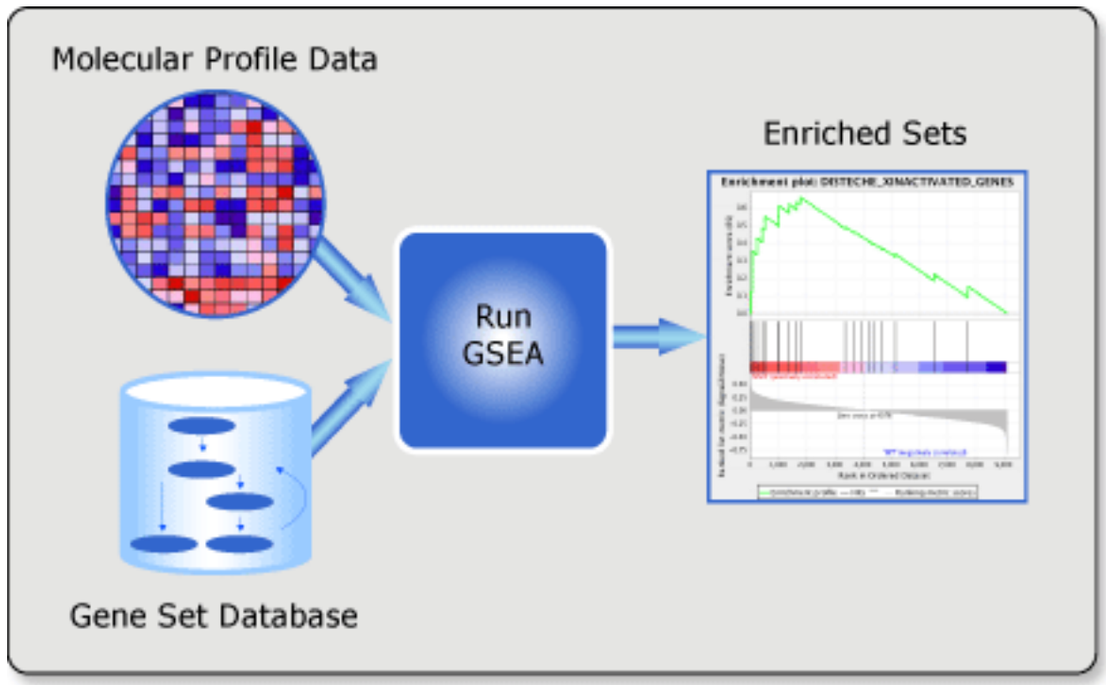
Remember to Correct For Multiple Hypothesis Tests

x out of n Vs **M out of N**

http://geneontology.org

Gene Ontology

- Panther (pantherdb.org)
- http://geneontology.org
- DAVID (https://david.ncifcrf.gov)
- GSEA (http://software.broadinstitute.org/gsea/index.jsp)
- many others…



http://geneontology.org