Selected slides from BioNetwork slide by Dr. Nataša Pržulj

Dr. Nataša Pržulj
Department of Computing
Imperial College London
natasha@imperial.ac.uk
www.doc.ic.ac.uk/~natasha/

Introduction: biological networks

- The goal of ***systems biology***:
  - Systems-level understanding of biological systems
  - Analyze not only individual components, but their interactions as well and emergent behavior
  - In the rest of the course: Learn new biology from the topology/wiring/structure of such interaction networks

Introduction: biological networks

- Types of biological networks:
  - Intra-cellular networks
    1. Transcriptional regulation networks
    2. Protein structure networks
    3. Metabolic networks
    4. Protein-protein interaction (PPI) networks
    5. Cell signaling networks
  - Other biological networks
    - Neuronal synaptic connection networks
    - Brain functional networks
    - Ecological food webs
    - Phylogenetic networks
    - Correlation networks (e.g., gene co-expression)
    - Disease – “disease gene” association networks
    - Drug – “drug target” networks

Introduction: biological networks

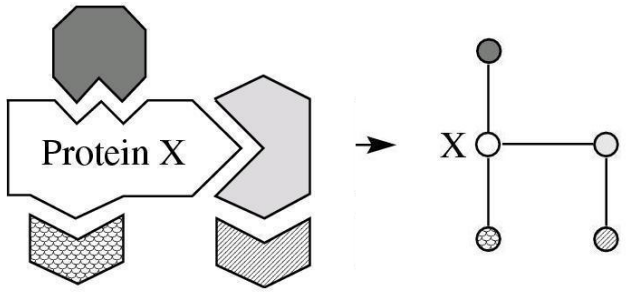
- **Biological nets**

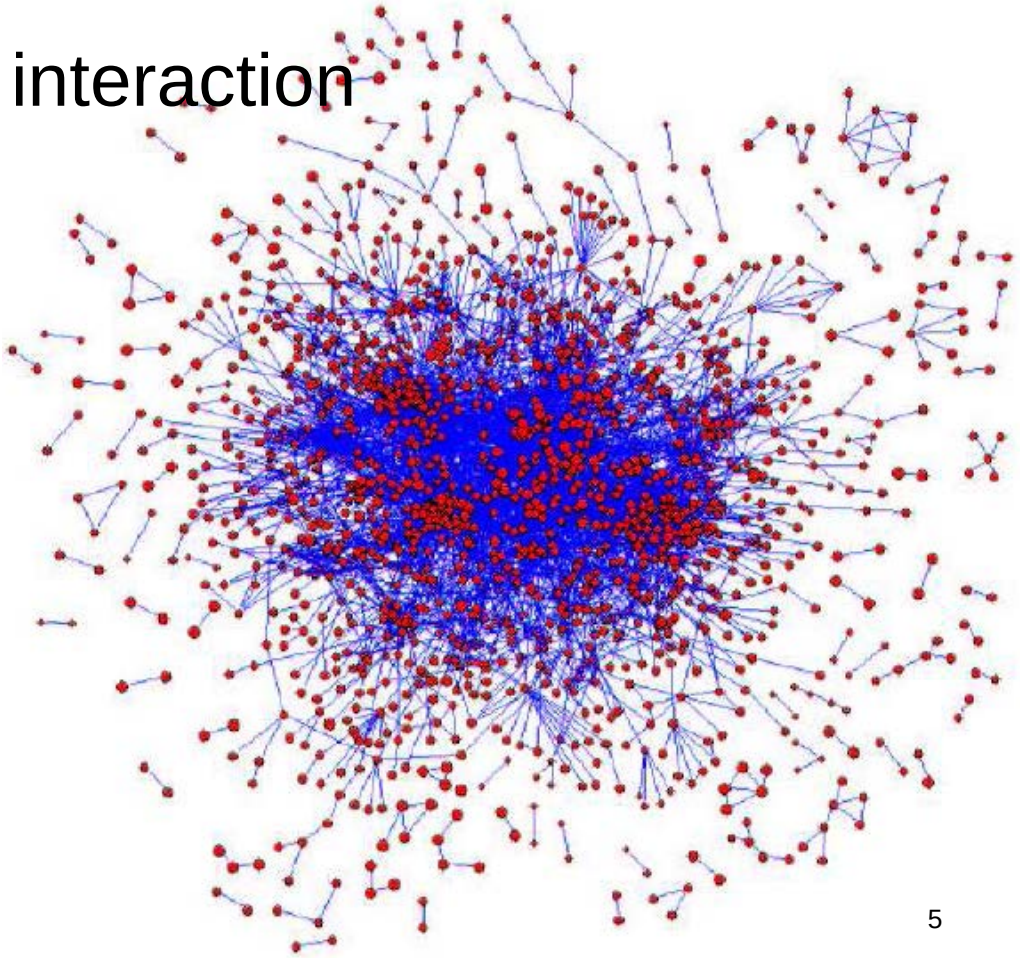
  E.g., Protein structure networks

Introduction: biological networks

- **Biological nets**

  E.g., Protein-protein interaction (PPI) networks

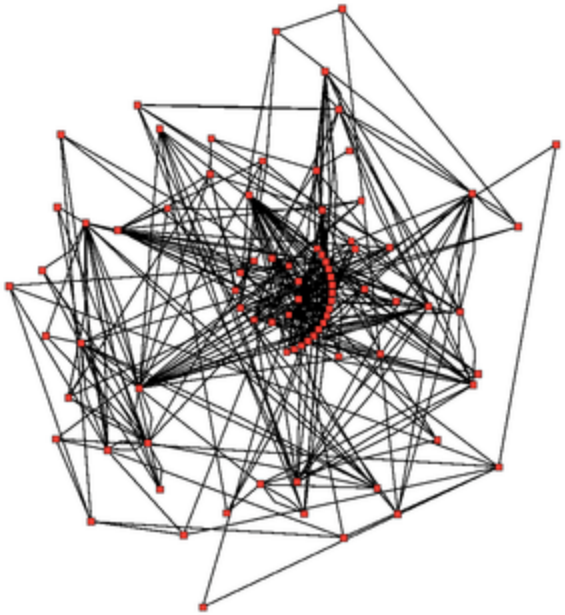




Introduction: biological networks

- **Biological nets**

  E.g., Metabolic networks

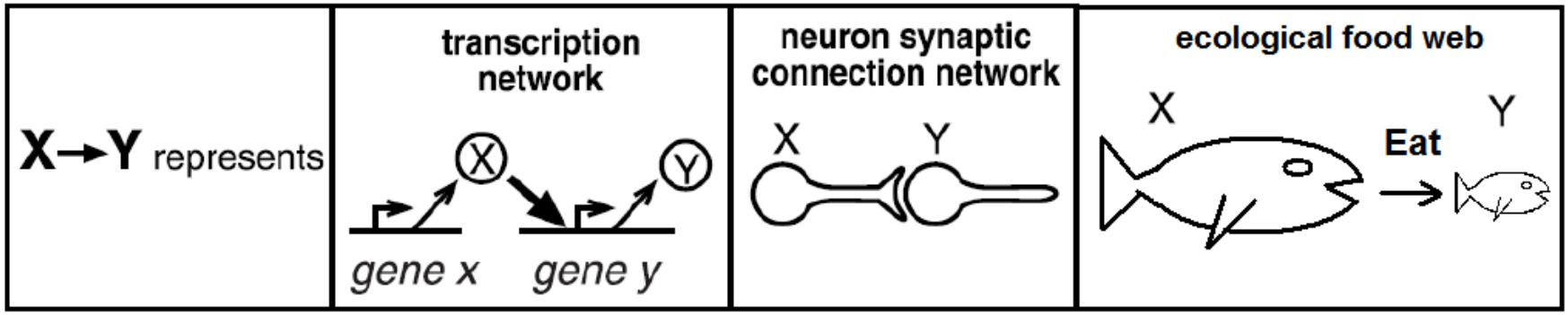


Metabolic network of *A. thaliana*

Introduction: biological networks

- **Biological nets**

  Other network types



Introduction: biological networks

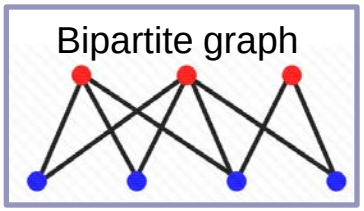
- Intra-cellular networks
  1. Transcriptional regulation networks
  2. Protein structure networks
  3. Metabolic networks
  4. Protein-protein interaction (PPI) networks
  5. Cell signaling networks
- All of these networks describe cellular functioning at different levels and often “overlap”
  - Cell relies on numerous highly interconnected interactions and chemical reactions between various types of molecules, e.g., proteins, DNA, RNA, metabolites, etc.
  - Various activities of cells are controlled by the action of molecules upon molecules
  - Proteins – central players
  - **Main application of methods in this course: PPI networks**

Metabolic networks

- Used for studying and modeling ***metabolism***
  - Biochemical reactions in cells that allow an organism to:
    - Respond to the environment
    - Grow
    - Reproduce
    - Maintain its structure
    - …
  - i.e., the main biochemical reactions needed to keep an organism in *homeostasis*
    - An internal regulation that maintains a stable, constant condition of a living system

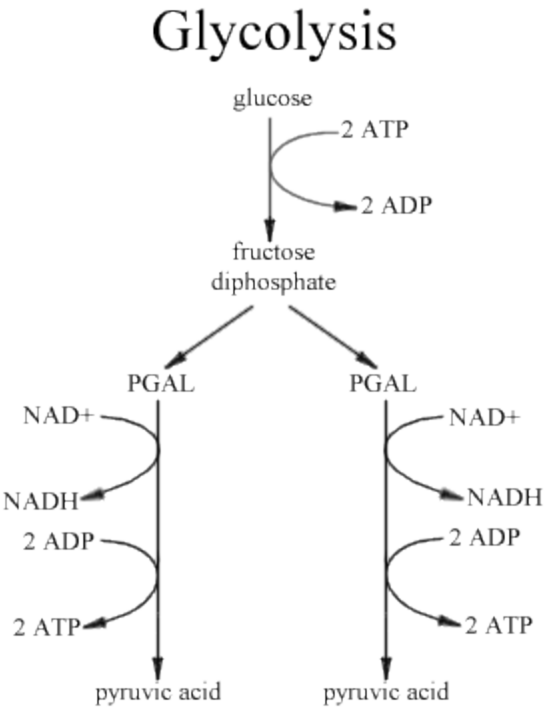
Metabolic networks

- ***Metabolites***
  - Small molecules such as glucose and amino acids
  - Also, macromolecules such as polysaccharides and glycans (carbohydrates)
- ***Metabolic pathways***
  - Series of successive biochemical reactions for a specific metabolic function, e.g., glycolysis, or penicillin synthesis, that convert one metabolite into another
  - ***Enzymes***: proteins that catalyze (accelerate) chem. reactions
- Thus, in a metabolic pathway:
  - Nodes correspond to metabolites and enzymes
    - In an alternate order → ***bipartite graphs***
  - Directed edges correspond to metabolic reactions
  - Simpler approaches: **nodes** are metabolites, directed edges are reactions that convert one metabolite into the other; or **nodes** are enzymes and metabolites as edges

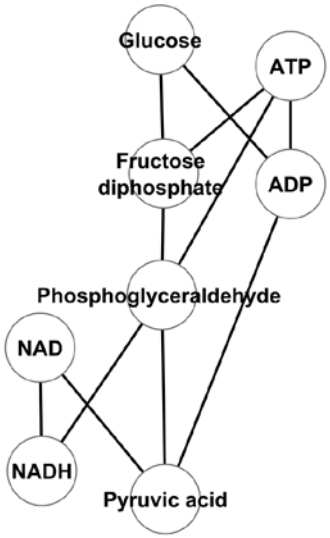


Metabolic networks

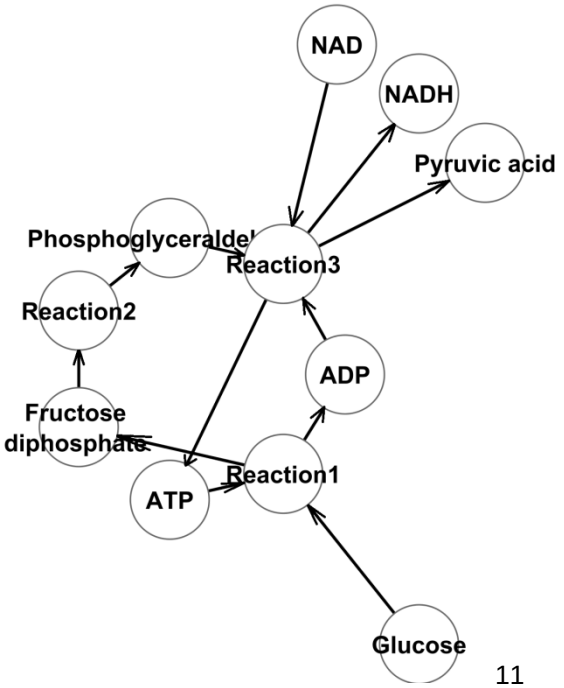
- Example: part of glycolysis pathway



Metabolite-centric representation:



Reactions + metabolites:



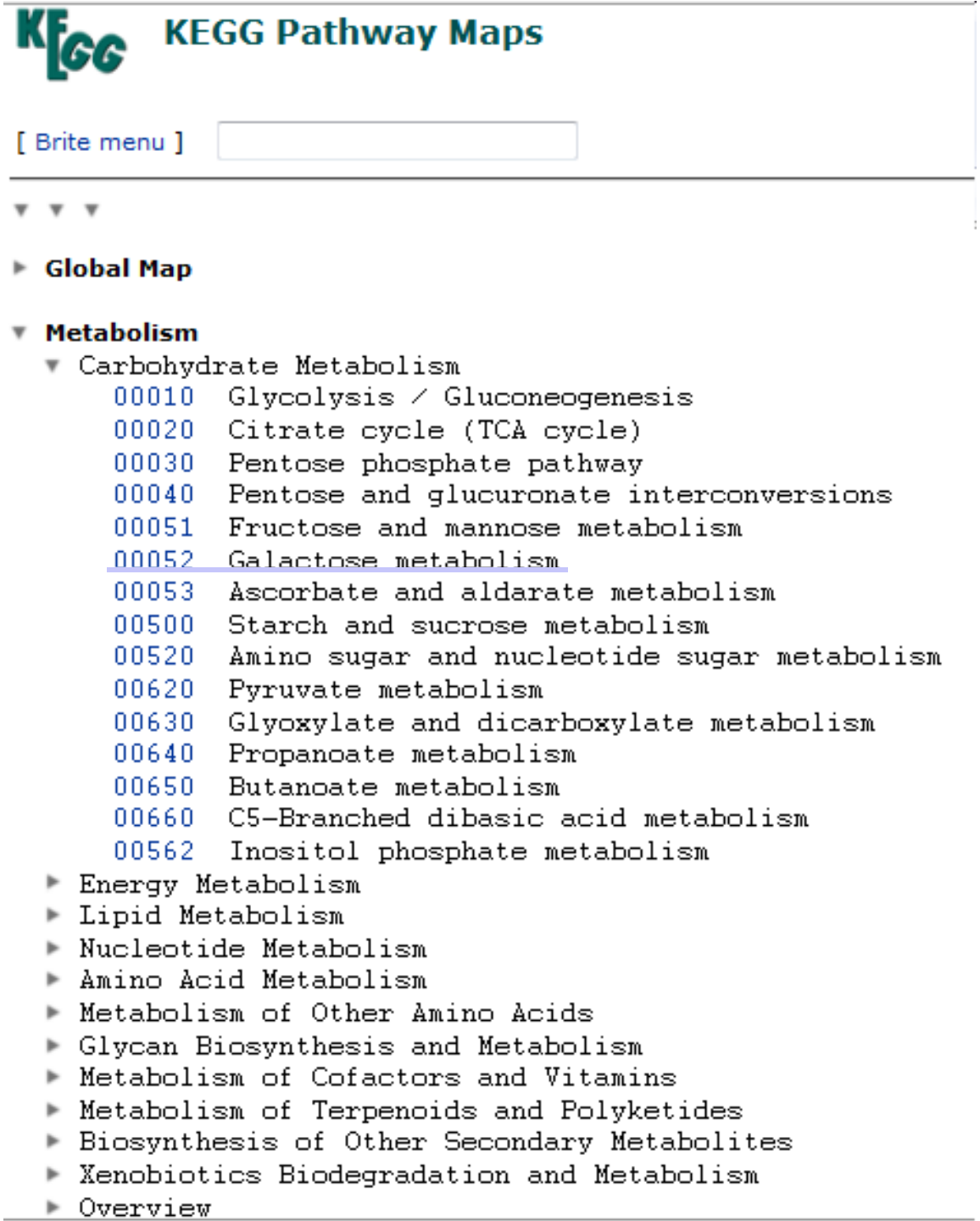
Metabolic networks

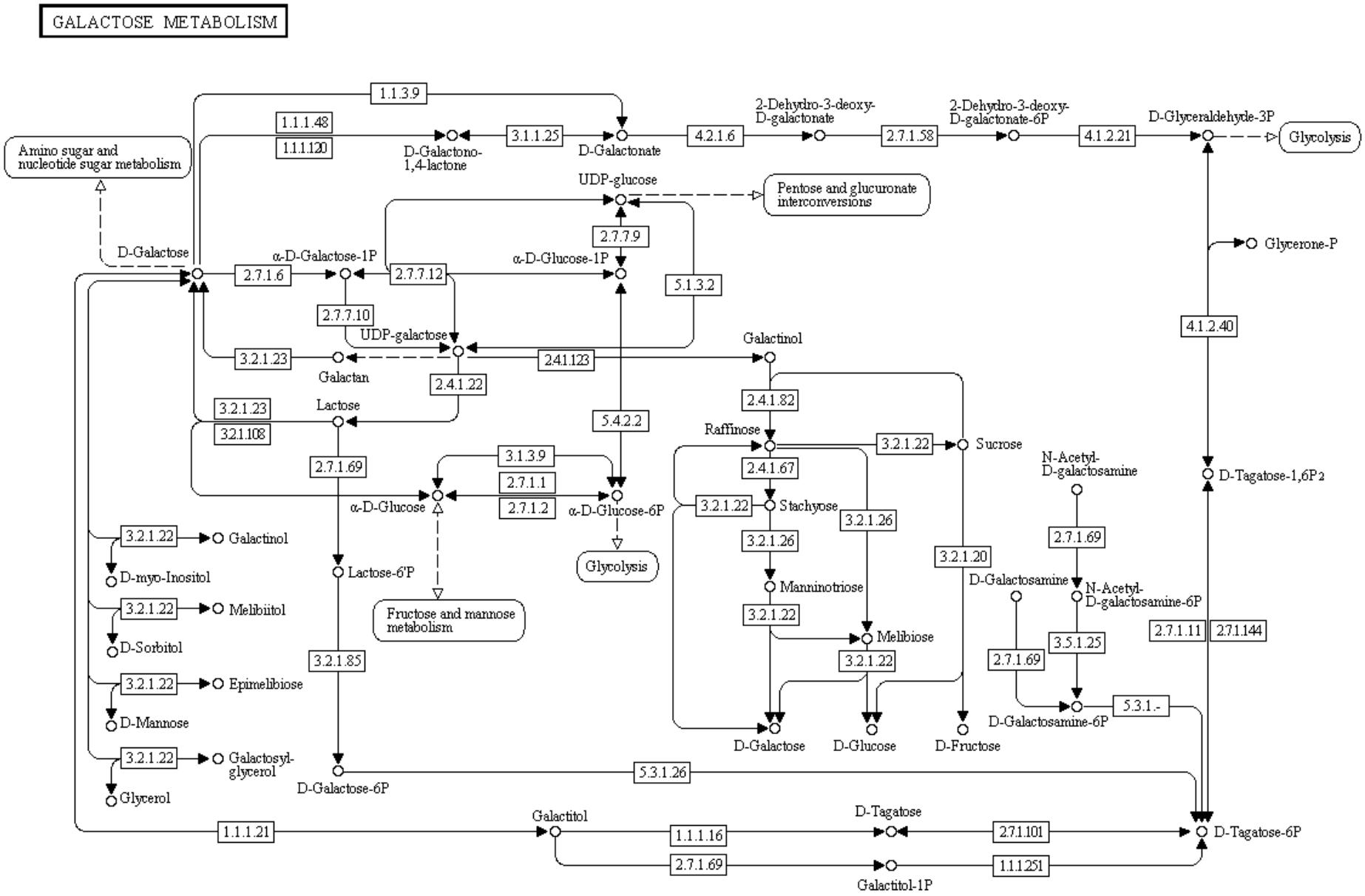
- All metabolic pathways of a cell form a ***metabolic network***
  - Complete view of cellular metabolism and material/mass flow through the cell
  - Cell relies on this network to digest substrates from the environment, generate energy, and synthesize components needed for its growth and survival
  - Insights from analyzing them used to, for example:
    - Cure human metabolic diseases through better understanding of the metabolic mechanisms
    - Control infections of pathogens by understanding the metabolic differences between human and pathogens

Metabolic networks

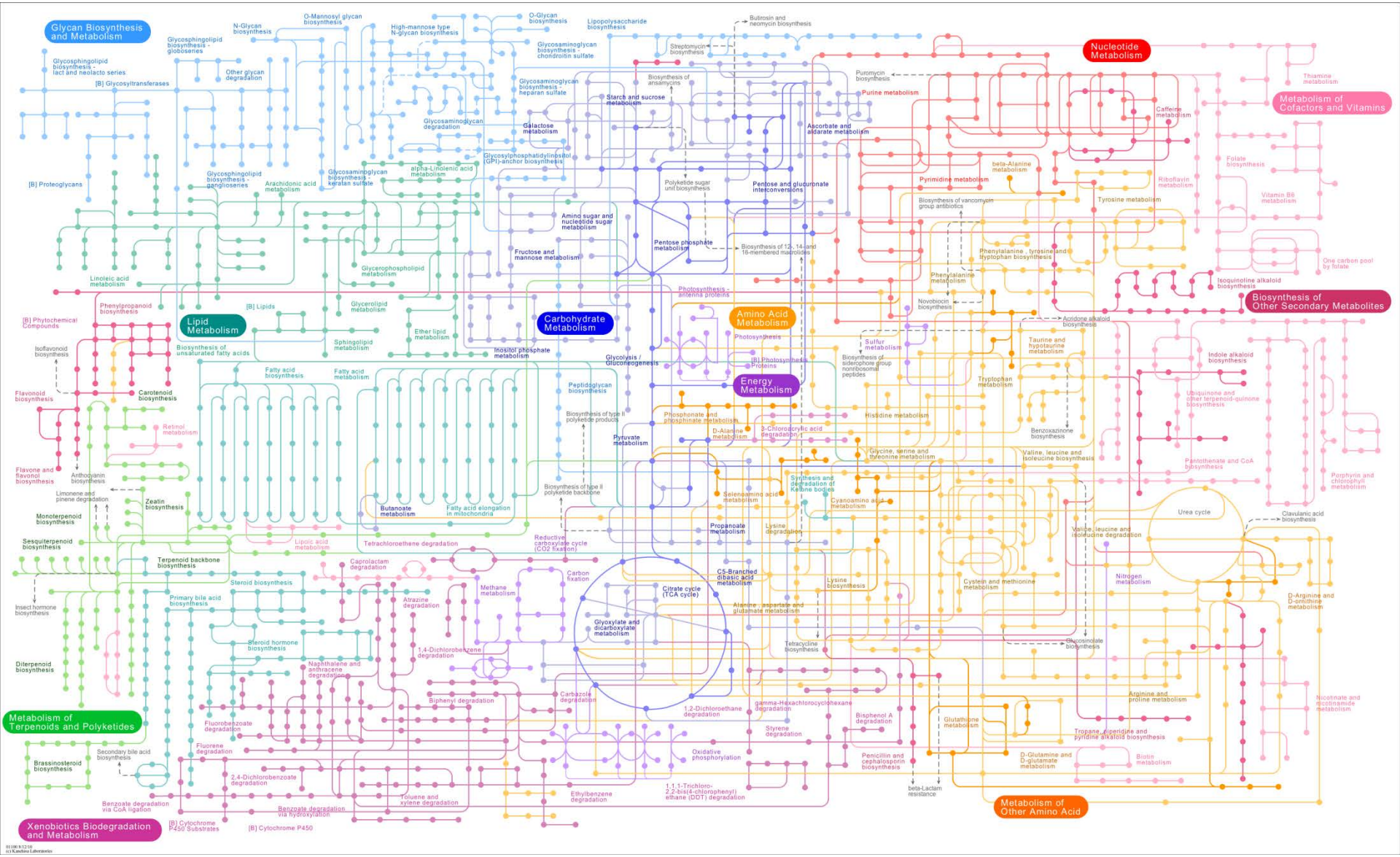
- Constructed:
  - Partially experimentally
  - Partially from genetic sequence (homology)
- Available for many organisms, from bacteria to human
- Available on-line:
  - KEGG (Kyoto Encyclopedia of Genes and Genomes)
    - Info on genes, proteins, reactions, pathways
    - Both for eukaryotes and prokaryotes
  - GeneDB–contains similar info
  - BioCyc, EcoCyc, MetaCyc
    - More specialized info on particular species
  - WIT, renamed to ERGO


File Edit View History Bookmarks Tools Help
http://www.genome.jp/kegg/
UCI ICS UCImail ICSmail FB EMBO NIF ND CSE CWP webfile
KEGG: Kyoto Encyclopedia of Genes a...
Go
Clear
» Japanese
KEGG Home
Introduction
Overview
Release notes
Current statistics
KEGG Identifiers
Pathway maps
Brite hierarchies
KEGG XML
KEGG API
KEGG FTP
KegTools
GenomeNet
DBGET/LinkDB
Feedback
KEGG: Kyoto Encyclopedia of Genes and Genomes
A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, the ecosystem, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.
Main entry point to the KEGG web service
KEGG2 KEGG Table of Contents Update notes Help
Data-oriented entry points
KEGG PATHWAY Pathway maps and pathway modules Pathway maps
KEGG BRITE Functional hierarchies and ontologies Brite hierarchies
KEGG DISEASE Human diseases Disease classification
KEGG DRUG Drugs ATC drug classification
KEGG ORTHOLOGY KO system and ortholog annotation KO system
KEGG GENES Genes and proteins
KEGG GENOME Genomes KEGG organisms
KEGG COMPOUND Chemical compounds Compound classification
KEGG GLYCAN Glycans
KEGG REACTION Reactions


KEGG Pathway Maps
[Brite menu]
Global Map
Metabolism
Carbohydrate Metabolism
00010 Glycolysis / Gluconeogenesis
00020 Citrate cycle (TCA cycle)
00030 Pentose phosphate pathway
00040 Pentose and glucuronate interconversions
00051 Fructose and mannose metabolism
00052 Galactose metabolism
00053 Ascorbate and aldarate metabolism
00500 Starch and sucrose metabolism
00520 Amino sugar and nucleotide sugar metabolism
00620 Pyruvate metabolism
00630 Glyoxylate and dicarboxylate metabolism
00640 Propanoate metabolism
00650 Butanoate metabolism
00660 C5-Branched dibasic acid metabolism
00562 Inositol phosphate metabolism
Energy Metabolism
Lipid Metabolism
Nucleotide Metabolism
Amino Acid Metabolism
Metabolism of Other Amino Acids
Glycan Biosynthesis and Metabolism
Metabolism of Cofactors and Vitamins
Metabolism of Terpenoids and Polyketides
Biosynthesis of Other Secondary Metabolites
Xenobiotics Biodegradation and Metabolism
Overview



00052 5/18/10
(c) Kanehisa Laboratories

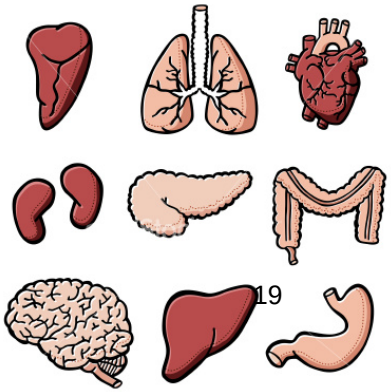

Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Metabolism of Other Amino Acid
Xenobiotics Biodegradation and Metabolism
N-Glycan biosynthesis
O-Mannosyl glycan biosynthesis
High-mannose type N-glycan biosynthesis
O-Glycan biosynthesis
Lipopolysaccharide biosynthesis
Glycosphingolipid biosynthesis - globoseries
Glycosaminoglycan biosynthesis - chondroitin sulfate
Glycosphingolipid biosynthesis - lacto and neolacto series
[B] Glycosyltransferases
Other glycan degradation
Glycosaminoglycan biosynthesis - heparan sulfate
Starch and sucrose metabolism
Glycosaminoglycan degradation
Galactose metabolism
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
[B] Proteoglycans
Glycosphingolipid biosynthesis - ganglioseries
Arachidonic acid metabolism
Glycosaminoglycan biosynthesis - keratan sulfate
alpha-Linolenic acid metabolism
Amino sugar and nucleotide sugar metabolism
Fructose and mannose metabolism
Pentose phosphate metabolism
Glycerophospholipid metabolism
Linoleic acid metabolism
Phenylpropanoid biosynthesis
[B] Phytochemical Compounds
[B] Lipids
Glycerolipid metabolism
Sphingolipid metabolism
Ether lipid metabolism
Isoflavonoid biosynthesis
Biosynthesis of unsaturated fatty acids
Inositol phosphate metabolism
Flavonoid biosynthesis
Carotenoid biosynthesis
Fatty acid biosynthesis
Fatty acid metabolism
Retinol metabolism
Flavone and flavonol biosynthesis
Anthocyanin biosynthesis
Limonene and pinene degradation
Zeatin biosynthesis
Monoterpenoid biosynthesis
Sesquiterpenoid biosynthesis
Terpenoid backbone biosynthesis
Steroid biosynthesis
Primary bile acid biosynthesis
Insect hormone biosynthesis
Steroid hormone biosynthesis
Diterpenoid biosynthesis
Brassinosteroid biosynthesis
Secondary bile acid biosynthesis
Peptidoglycan biosynthesis
Biosynthesis of type II polyketide products
Glycolysis / Gluconeogenesis
Pyruvate metabolism
Biosynthesis of type II polyketide backbone
Butanoate metabolism
Fatty acid elongation in mitochondria
Lipoic acid metabolism
Tetrachloroethene degradation
Reductive carboxylate cycle (CO2 fixation)
Caprolactam degradation
Carbon fixation
Methane metabolism
Atrazine degradation
Citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
1,4-Dichlorobenzene degradation
Naphthalene and anthracene degradation
Biphenyl degradation
Carbazole degradation
Fluorobenzoate degradation
Fluorene degradation
2,4-Dichlorobenzoate degradation
Benzoate degradation via CoA ligation
Benzoate degradation via hydroxylation
Toluene and xylene degradation
Ethylbenzene degradation
[B] Cytochrome P450 Substrates
[B] Cytochrome P450
Oxidative phosphorylation
1,1,1-Trichloro-2,2-bis(4-chlorophenyl) ethane (DDT) degradation
1,2-Dichloroethane degradation
gamma-Hexachlorocyclohexane degradation
Styrene degradation
Bisphenol A degradation
Penicillin and cephalosporin biosynthesis
beta-Lactam resistance
Glutathione metabolism
D-Glutamine and D-glutamate metabolism
Tropane, piperidine and pyridine alkaloid biosynthesis
Biotin metabolism
Arginine and proline metabolism
Nicotinate and nicotinamide metabolism
D-Arginine and D-ornithine metabolism
Nitrogen metabolism
Cysteine and methionine metabolism
Lysine biosynthesis
Alanine, aspartate and glutamate metabolism
Tetracycline biosynthesis
Glucosinolate biosynthesis
Urea cycle
Clavulanic acid biosynthesis
Valine, leucine and isoleucine degradation
Propanoate metabolism
Lysine degradation
C5-Branched dibasic acid metabolism
Selenoamino acid metabolism
Cyanoamino acid metabolism
Synthesis and degradation of ketone bodies
Valine, leucine and isoleucine biosynthesis
Glycine, serine and threonine metabolism
Phosphonate and phosphinate metabolism
D-Alanine metabolism
3-Chloroacrylic acid degradation
Histidine metabolism
Benzoxazinone biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Indole alkaloid biosynthesis
Tryptophan metabolism
Taurine and hypotaurine metabolism
Sulfur metabolism
Biosynthesis of siderophore group nonribosomal peptides
Acridone alkaloid biosynthesis
Isoquinoline alkaloid biosynthesis
Novobiocin biosynthesis
Phenylalanine metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
One carbon pool by folate
Photosynthesis - antenna proteins
Photosynthesis
[B] Photosynthesis Proteins
Biosynthesis of 12-, 14- and 16-membered macrolides
Polyketide sugar unit biosynthesis
Pentose and glucuronate interconversions
Biosynthesis of vancomycin group antibiotics
Pyrimidine metabolism
beta-Alanine metabolism
Tyrosine metabolism
Riboflavin metabolism
Vitamin B6 metabolism
Folate biosynthesis
Caffeine metabolism
Thiamine metabolism
Ascorbate and aldarate metabolism
Purine metabolism
Puromycin biosynthesis
Streptomycin biosynthesis
Biosynthesis of ansamycins
Butirosin and neomycin biosynthesis
Lipopolysaccharide biosynthesis

Metabolic networks

- Further readings
  - Junker and Schreiber, “Analysis of Biological Networks,” Wiley, 2008.
  - H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, “The large-scale organization of metabolic networks,” *Nature* 407, 2000.
  - R. Tanaka, “Scale-rich metabolic networks,” *Physical Review Letters* 94, 2005.

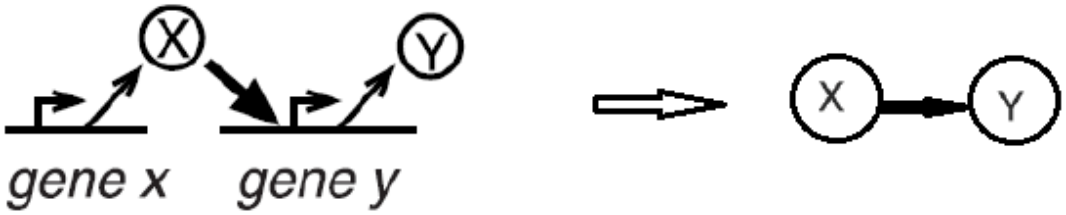
Transcriptional regulation networks

- Model regulation of *gene expression*
  - Recall: gene → mRNA → protein
- Gene regulation
  - Gives a cell control over its structure and function, e.g.:
    - *Cellular differentiation* – a process by which a cell turns into a more specialized cell type
    - *Morphogenesis* (a process by which an organism develops its shape)
    - ...



Transcriptional regulation networks

- Nodes correspond to genes
  - DNA sequences which are transcribed into mRNAs that translate into proteins
- Directed edges correspond to interactions through which the products of one gene affect those of another
  - Protein-protein, protein-DNA and protein-mRNA interactions

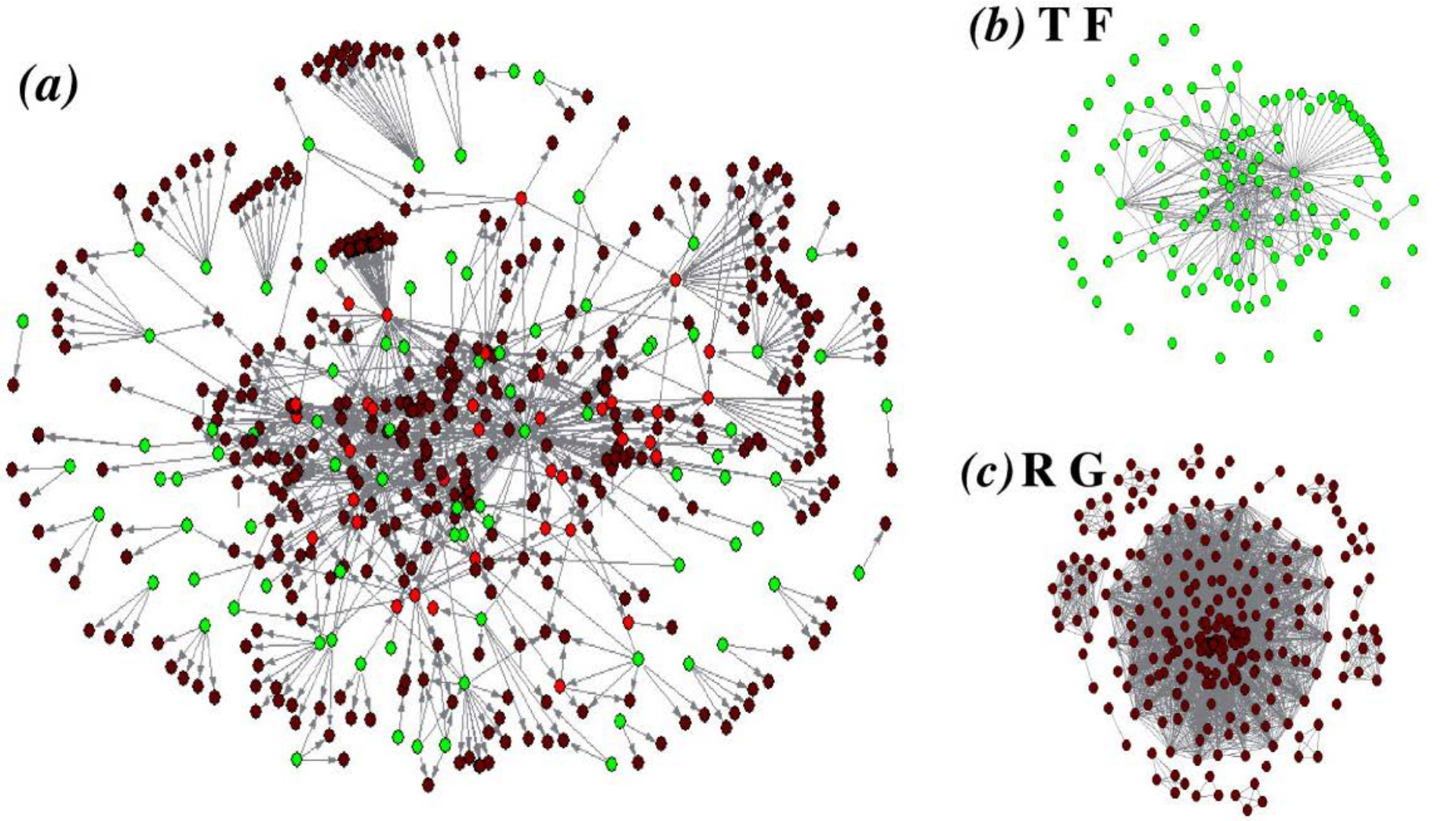


- *Transcription factor* X (protein product of gene X) binds regulatory DNA regions of gene Y to regulate the production rate (i.e., stimulate or repress transcription) of protein Y
  - Note: proteins are products of gene expression that play a key role in regulation of gene expression

Transcriptional regulation networks

- Problem
  - *Stimulation* and *repression* of gene transcription are both represented the same way in the network
- Available for *model organisms*
  - Non-human species manipulated and studied to get insights into workings of other organisms, e.g.:
    - Baker's yeast, *S. cerevisiae* (Milo et al., 2002)
    - *E. coli* (Shen-Orr et al., 2002)
    - Sea urchin (Davidson et al., 2002)
    - Fruitfly, *D. melanogaster*
  - Available from dBs: EcoCyc, GeneNet, KEGG, RegulonDB, Reactom, TRANSPATH, TRANSFAC

E. coli



Representation of the *E. coli* transcriptional regulatory network. a) Representation of the transcription-factor gene regulatory network of *E. coli*. Green circles represent transcription factors, brown circles denote regulated genes, and those with both functions are coloured in red. Projections of the network onto b) transcription factor and onto c) regulated gene nodes are also shown.

Guzmán-Vargas and Santillán *BMC Systems Biology* 2008 **2**:13 doi:10.1186/1752-0509-2-13

Transcriptional regulation networks

- Further readings:
  - Junker and Schreiber, “Analysis of Biological Networks,” Wiley, 2008.
  - List of databases:

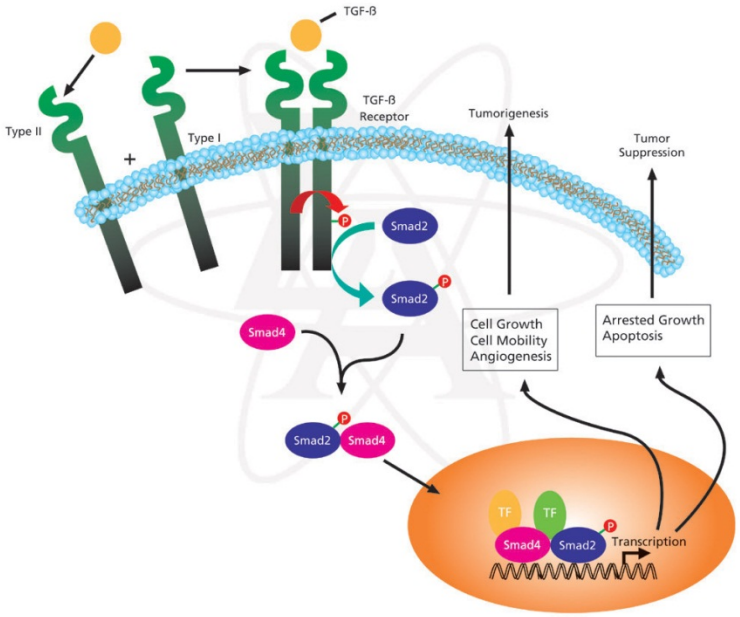
    University of Pittsburg, Health Science Library

    Online Bioinformatics Resources Collection

    http://www.hsls.pitt.edu/obrc/

Cell signaling networks

- ***Cell signaling***
  - Complex communication system that governs basic cellular activities, e.g., development, repair, immunity
- Errors in signaling cause diseases
  - E.g., cancer, autoimmune diseases, diabetes…



E.g.: **Transforming growth factor beta** (TGF-β) is a protein that controls proliferation, cellular differenciation, and other functions in most cells.

Cell signaling networks

- ***Signaling pathways***
  - Ordered sequences of signal transduction reactions in a cell, as shown in the previous figure
  - Cascade of reversible chemical modifications of proteins
    - E.g., phosphorylation catalyzed by *protein kineases:* enzymes that modify other proteins by adding phosphate groups to them (process called *phosphorylation*)
- Signaling pathways in the cell form the ***cell signaling network***
  - Nodes are proteins and edges are directed

Cell signaling networks

Famous examples (lots of literature on them):

- *Mitogen-activated protein kinase (MAPK)* pathway
  - Originally called “ERK” pathway
  - MAPK protein: an enzyme, a *protein kinase,* which can attach *phosphate groups* to a target protein, causing its spatial reorganization and affecting its function
    - Other enzymes can restore protein’s initial function
  - E.g.:
    - MYC
      - An *oncogene* transcription factor expressed in a wide range of human cancers (oncogene – when mutated or over-expressed, the gene helps turn a normal into a tumor cell)
      - MAPK can *phosphorylate* (attach phosphate group to) MYC and alter gene transcription and cell cycle progression
    - EGFR = “epidermal growth factor receptor”
      - Activates MAPK pathway
      - Mutations affecting its expression/activity can result in cancer

Cell signaling networks

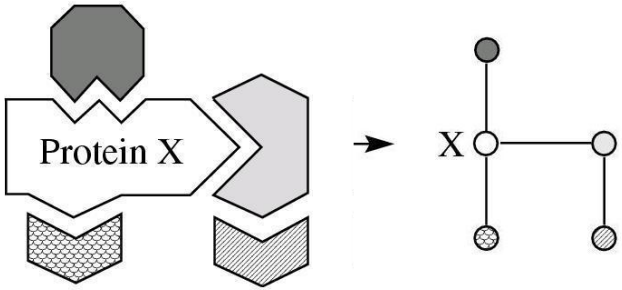
Famous examples (lots of literature on them) cont'd:

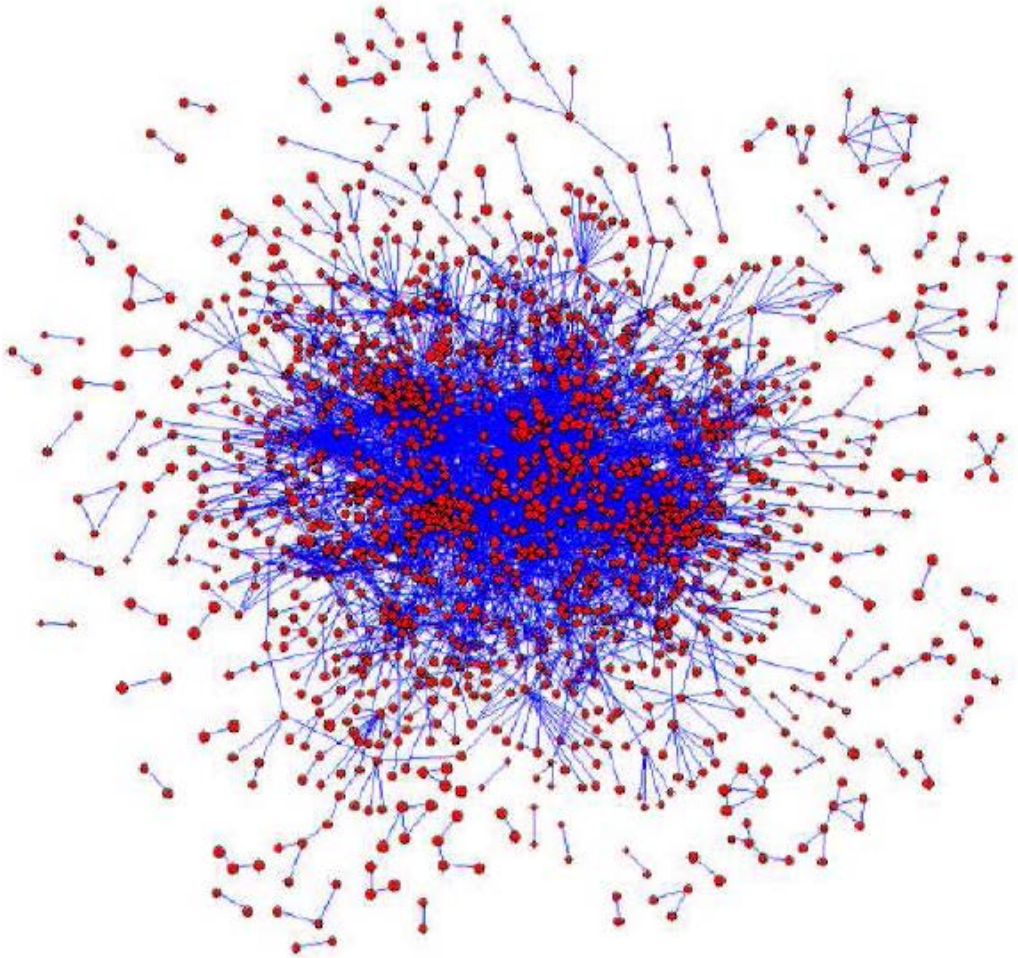
- *Hedgehog signaling pathway*
  - One of the key regulators of animal development
  - Conserved from fly to human
  - Establishes basis of fly body plan
  - Important during *embryogenesis* (the process by which the embryo develops) and *metamorphosis* (from larva to pupa to adult)
- *TGF-beta signaling pathway*
  - The "transforming growth factor" (TGF) signaling pathway
  - Involved in:
    - Cell growth
    - Cell differentiation
    - *Apoptosis* (programmed cell death)

Cell signaling networks

- Compared to metabolic networks:
  - Limited mass flow
  - Instead, sig. nets provide information transmission along a sequence of reactions – one enzyme modulates the activity of another one, which then modulates the activity of the third enzyme, etc., but *enzymes are not consumed* in the reactions they catalyze
- Compared to transcriptional reg. networks:
  - They overlap, but gene expression, i.e., transcription factors, can be seen as the “final targets” of signaling pathways
- Compared to PPI networks:
  - Signal transduction is indeed mediated between proteins, but PPIs are undirected without a defined input and output (as we will discuss soon)
  - Not all PPIs are involved in chemical reactions, or part of signal transduction
  - Also, many components of signaling are not proteins
- These networks have much in common
- At the same time, they reflect different aspects of cellular activity

Protein-protein interaction (PPI) networks





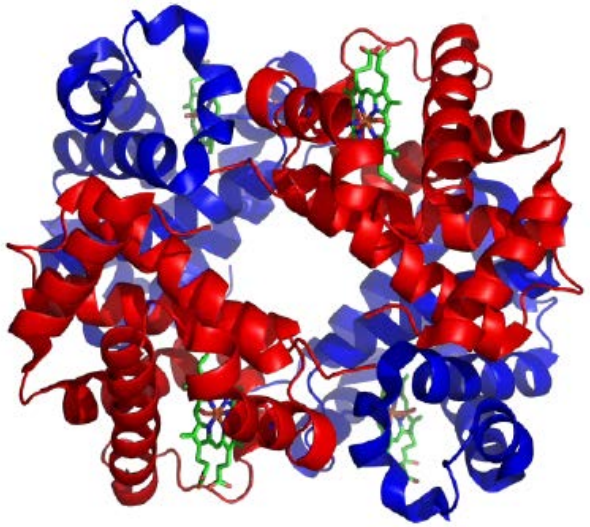
Protein-protein interaction (PPI) networks

- A *protein-protein interaction (PPI)* usually refers to a physical interaction, i.e., binding between proteins
- Can be other associations of proteins such as functional interactions – e.g., synthetic lethality: type of a “genetic interaction” (will introduce later)

Protein-protein interaction (PPI) networks

- PPIs are very important for structure and function of a cell:
  - Participate in signal transduction (*transient interactions*)
    - Play a role in many diseases (e.g., cancer)
  - Can be *stable interactions* forming a *protein complex*

    (a form of a quaternary protein structure, set of proteins which bind to do a particular function, e.g., ribosome, hemoglobin – illustrated below)



Protein-protein interaction (PPI) networks

- PPIs are very important for structure and function of a cell:
  - Can be *transient interactions*
    - Brief interactions that modify a protein that can further change PPIs e.g., protein kineases (add a phosphate group to a target protein)
    - A protein can carry another protein, e.g., *nuclear pore importins* (proteins that carry other proteins from cytoplasm to nucleus and vice versa)
    - Transient interactions form the *dynamic part of PPI networks*
  - Some estimates state that about 70% of interactions are stable and 30% are dynamic (transient)
- PPIs are essential to almost every process in a cell
- Thus, understanding PPIs is crucial for understanding life, disease, development of new drugs (most drugs affect PPIs)

Protein-protein interaction (PPI) networks

Methods to detect PPIs

- Biological and computational approaches
- None are perfect
  - High rates of *false positives*
    - Interactions present in the data sets that are not present in reality
  - High rates of *false negatives*
    - Missing true interactions

Protein-protein interaction (PPI) networks

Methods to detect PPIs

- PPIs initially studied individually by small-scale biochemical techniques (SS)
- However, large-scale (high-throughput) interaction detection methods (HT) are needed for high discovery rates of new protein interactions
- SS of better "quality," i.e., less noisy than HT
- However, HT are more standardized, while SS are performed differently each time
- SS are biased – the focus is on the subsets of proteins interesting to particular researchers
- HT – view of the entire proteome

Protein-protein interaction (PPI) networks

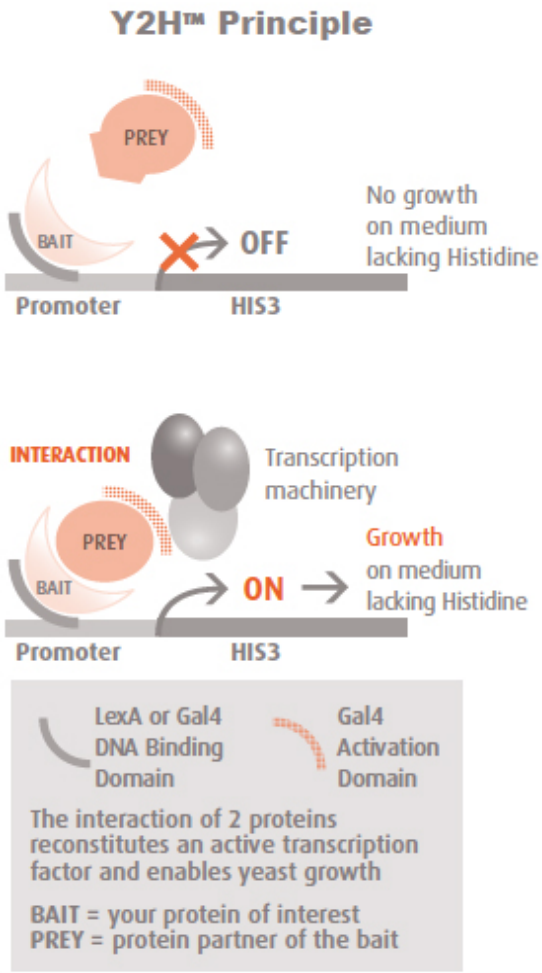
Methods to detect PPIs

- Physical binding
  - Yeast 2-hybrid (Y2H) screening
  - Mass spectrometry of purified complexes
- Functional associations
  - Correlated mRNA expression profiles
  - Genetic interactions
  - In silico (computational) methods
- In many cases, functional associations do take the form of physical binding

Protein-protein interaction (PPI) networks

Yeast two-hybrid assay

- *Binary* PPIs
- Pairs of proteins to be tested for interaction are expressed as *artificial* (genetically engineered) *fusion proteins* in yeast:
  - One protein is fused to a *reporter gene* (a gene attached to another gene of interest)
  - The other is fused to a *transcription factor*
  - Any interaction between them is detected by the transcriptional activation of the reporter gene



Protein-protein interaction (PPI) networks

Yeast two-hybrid assay

- One protein (in PPI) is “bait”, the other is “prey”
- Potential problem:
  - Interest in a particular pathway of, say 15 proteins
  - These 15 proteins are all “baits”
  - There is an order of magnitude more “preys”
  - This imposes a particular structure on the PPI network by experimental design without reflecting the underlying network topology
  - To avoid this, a matrix of *n x n* needs to be probed, where each bait is also a prey (Mark Vidal’s lab, Harvard)

Protein-protein interaction (PPI) networks

Yeast two-hybrid assay

- This method is scalable to the entire proteome
- Directly tests a protein pair for an interaction
- But high noise rate (50%, even up to 70%)
  - Because Y2H investigates interactions between:
    - artificial, fusion proteins
    - in the yeast
    - in the yeast's *nucleus*
  - **Each of these steps is noisy**
  - Proteins need to be in their native environment, not in nucleus
    - E.g., although proteins can physically bind, they never do so inside a cell, because of different localization, or because they are never simultaneously expressed

Protein-protein interaction (PPI) networks

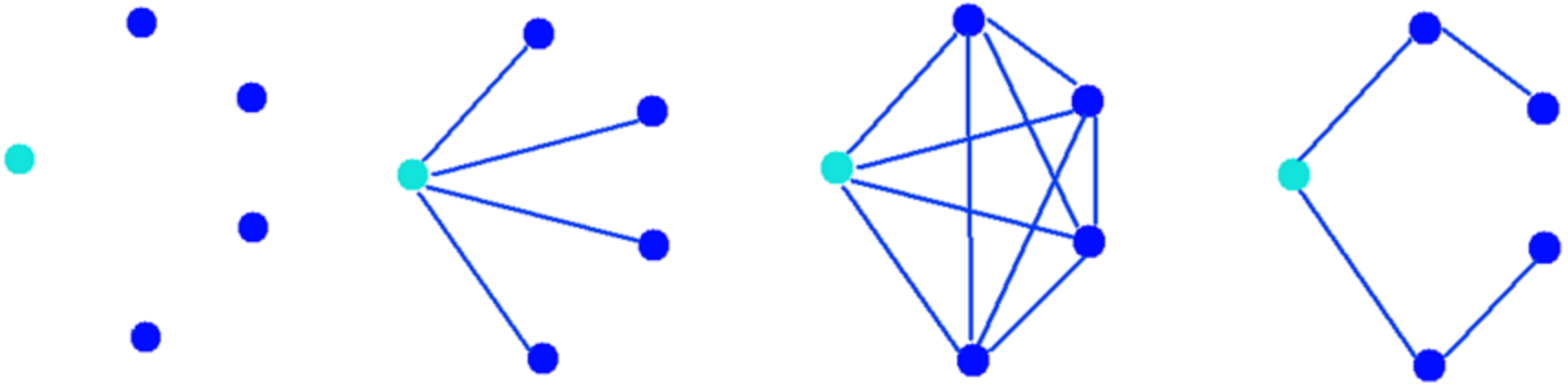
Mass spectrometry of purified complexes

- Individual proteins are tagged and used as hooks to biochemically purify whole protein complexes
- Complexes separated and components identified by mass spectrometry (MS)
  - MS measures mass-to-charge ratio of ions
- TAP (Tandem Affinity Purification)
- HMS-PCI (High-Throughput MS Protein Complex Identification)
- Not binary but co-complex data

Protein-protein interaction (PPI) networks

Mass spectrometry of purified complexes

- We know what proteins are in the complexes, but not how they are connected
  - Spoke model
  - Matrix model



Protein-protein interaction (PPI) networks

Mass spectrometry of purified complexes

- Pros:
  - Detects real complexes in their physiological settings
  - Consistency check is possible by tagging several members of a complex
  - Good for screening permanent/stable interactions
- Cons:
  - Might miss some complexes that are not present under given cellular conditions
  - Tagging may disturb complex formation
  - Loosely associated components can be washed off during purification

Protein-protein interaction (PPI) networks

Functional associations

- Correlated mRNA expression profiles (Dr. Rice's lectures)
  - Results in a gene expression correlation network
- Co-expression means that resulting proteins *could* interact
- Co-expression overlaid over PPI data, e.g. tool KeyPathwayMiner

Protein-protein interaction (PPI) networks

Functional associations

- Genetic interactions
  - Two non-essential genes that cause lethality when mutated at the same time form a *synthetic lethal* interaction
  - Such genes are often functionally associated and their encoded proteins may also interact physically
  - Charles Boone's group from University of Toronto published genetic interaction networks

Protein-protein interaction (PPI) networks

Functional associations

- Genetic interactions

Cell

Leading Edge
Perspective

Genetic Interactions in Cancer Progression and Treatment

Alan Ashworth,[1,2,*] Christopher J. Lord,[1,2,*] and Jorge S. Reis-Filho[1,2,*]
[1]The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, Fulham Road, London SW3 6JB, UK
[2]All authors contributed equally to this work
*Correspondence: alan.ashworth@icr.ac.uk (A.A.), chris.lord@icr.ac.uk (C.J.L.), jorge.reis-filho@icr.ac.uk (J.S.R.-F.)
DOI 10.1016/j.cell.2011.03.020

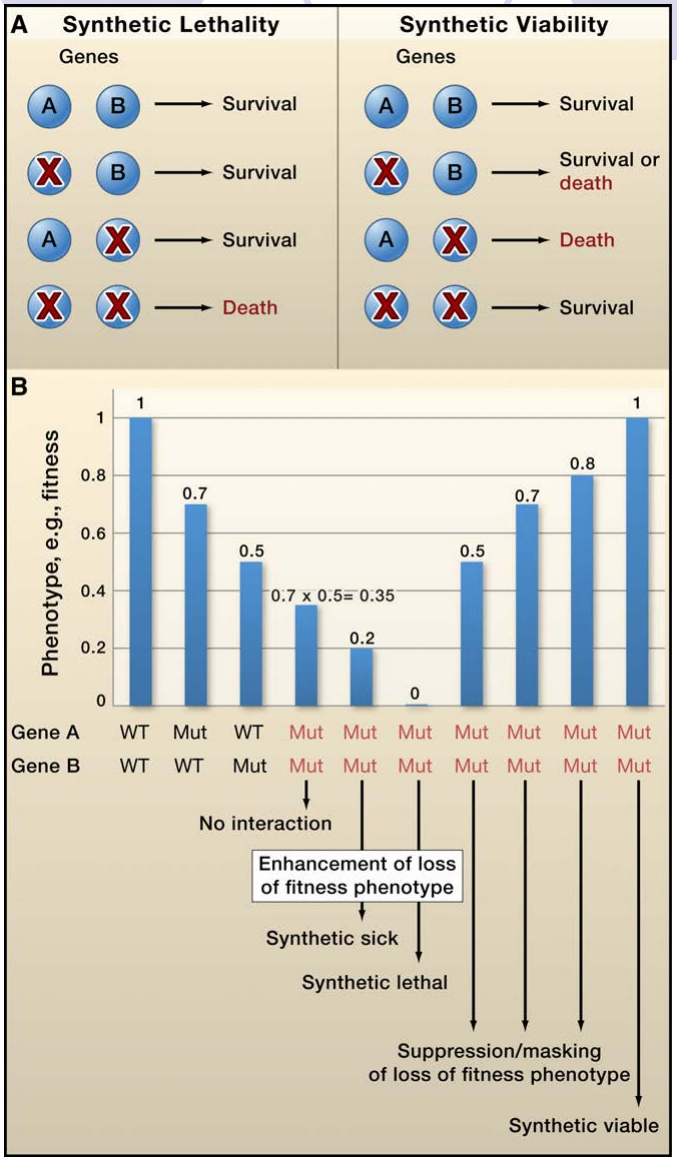


Figure 1. Gene Interactions in Cancer

(A) Extreme forms of genetic interaction are defined by synthetic lethality (in which a combination or synthesis of gene mutations causes cell death) and the reverse scenario, synthetic viability (in which a combination of gene effects rescues the lethal effects of a single gene change).

(B) Different modes of genetic interaction defined by quantitative effects on a phenotype, such as cell fitness. Here the value 1 represents the maximal fitness of cells, and the individual effects of changes in genes A (0.7) or B (0.5) are shown. When no interaction between genes A and B exists, the simple combination of effects (shown here as $0.7 \times 0.5 = 0.35$) is expected; any deviation from this value suggests an interaction between genes A and B.

Protein-protein interaction (PPI) networks

Functional associations

- In silico (computational) methods
  - Gene fusion (if two genes are present in one species and fused in another)
  - …

Protein-protein interaction (PPI) networks

Biases within PPI networks

- The following is lost:
  - Spatial information
  - Temporal information
  - Information about experimental conditions
  - Strength of interactions
  - Number of experiments confirming interactions
- PPI network: proteome + interactome
  - Proteome: a set of all **unique** proteins in an organism;
  - How does protein concentration affect the topology:
    - More instances of a protein in the cell → more interacting partners in the network?

Protein-protein interaction (PPI) networks

Quality and completeness of PPI data

- Data sets produced by different methods are often complementary
- Even data sets obtained by the same technique complement each other to some (large) extent
- Completeness of data sets:
  - Yeast: ~50% (~6K proteins, ~30K-60K interactions)
  - Human: ~10% (~25K proteins, ~260K interactions; ~300 million pairs to test)
  - Fly
  - Worm
  - Recently, herpes viruses (genome-wide coverage)

Protein-protein interaction (PPI) networks

PPI databases*

- Biological General Repository for Interaction Datasets (BioGRID)
- Human Protein Reference Database (HPRD)
- Saccharomyces Genome Database (SGD)
- Munich Information Center for Protein Sequences (MIPS)
- Database of Interacting Proteins (DIP)
- Molecular Interactions Database (MINT)
- Online Predicted Human Interaction Database (OPHID) → **I2D**
- VirusMINT

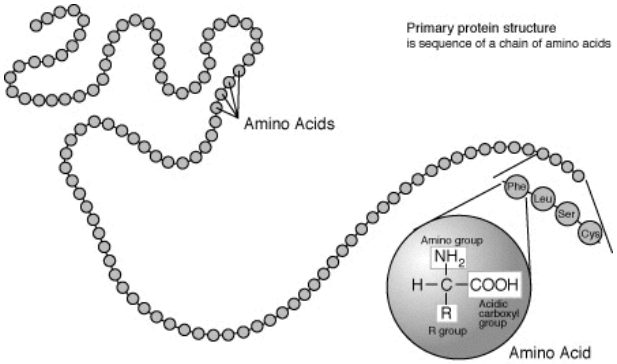
- The lack of standardization
  - Different databases use different naming conventions
  - Inconsistencies in mapping between them
  - This can seriously jeopardize network topological analyses

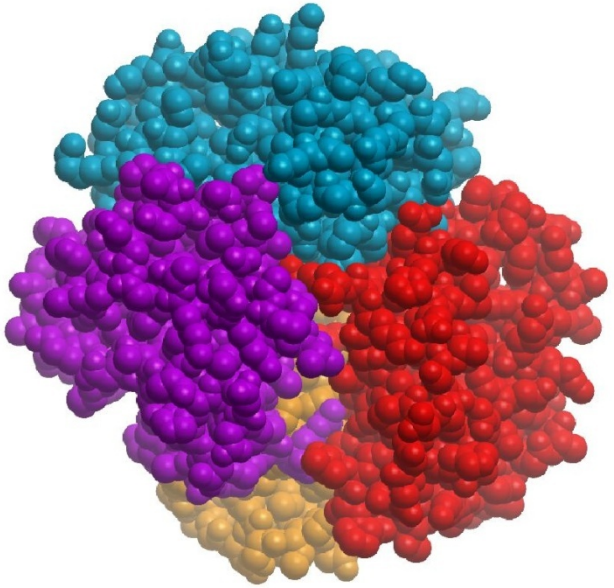
*Distinguish between binary and co-complex data.

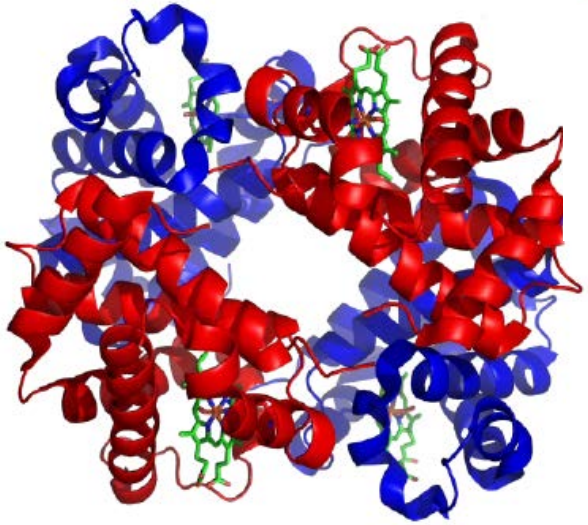
Protein-protein interaction (PPI) networks

- Additional readings:
  - Junker and Schreiber, “Analysis of Biological Networks,” Wiley, 2008
  - Chapter 4 of “Knowledge Discovery in Proteomics” by Wiggle and Jurisica
  - von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002, 417(6887):399–403
  - Mark Vidal lab’s work in *Nature Methods*, 2009

Protein structure networks



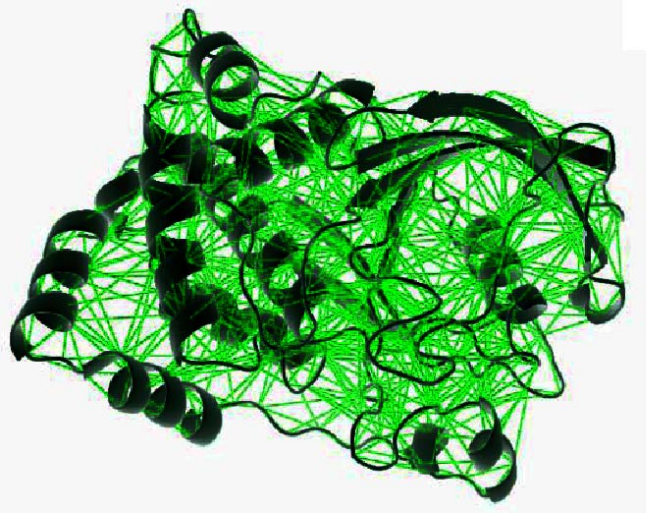




- PDB (Protein Data Bank): http://www.pdb.org/

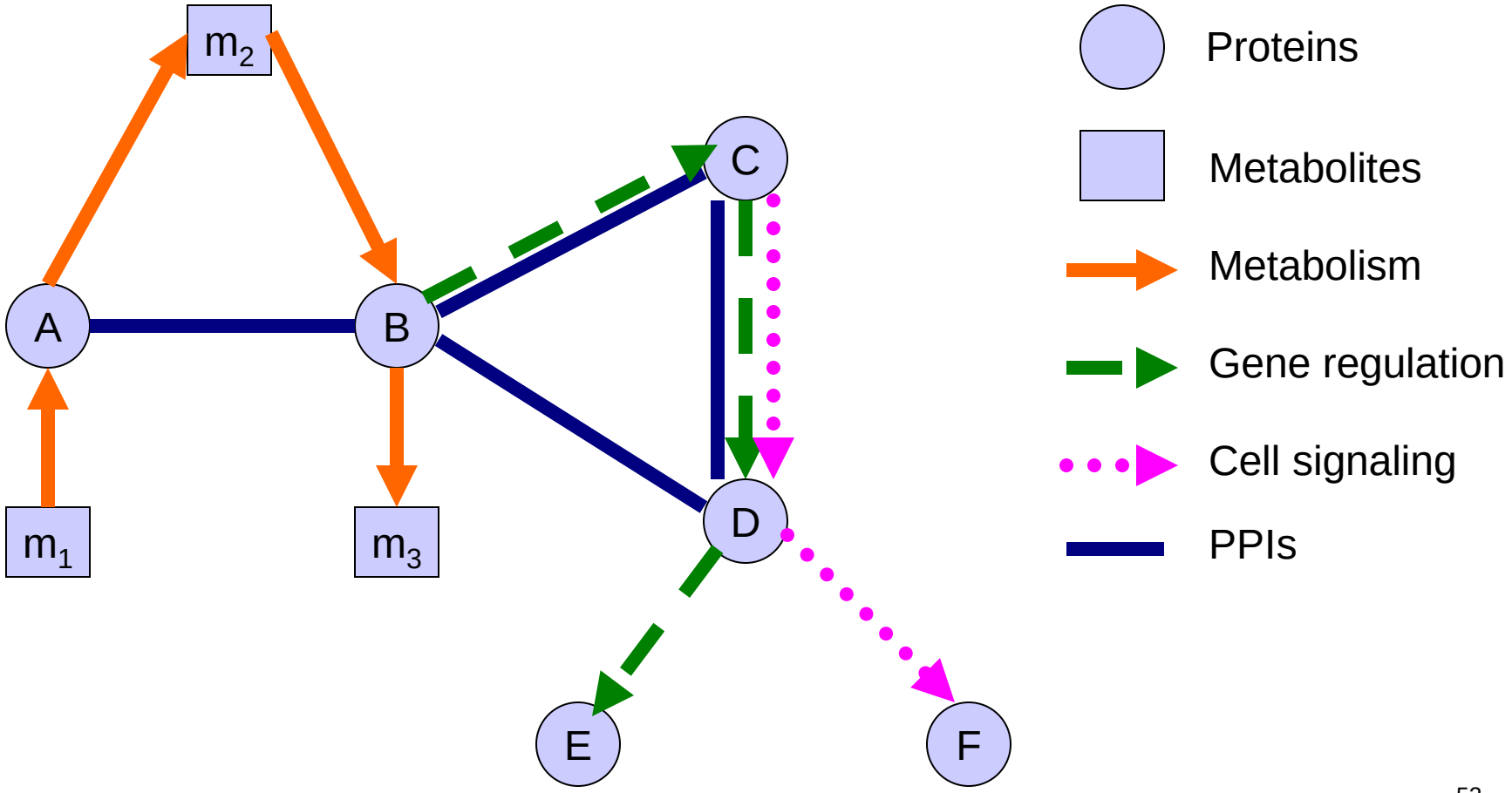
Protein structure networks

- "Residue interaction graphs" (RIGs) model protein structures (Dr. Malod-Dognin's lecture)
  - Nodes are amino acid residues
  - *Undirected, unweighted, e*dges exist between amino acids that are in close proximity in the protein's 3-dimensional structure
    - E.g., within 5 *Angstroms* (1 Å= 10^{-10} meters)



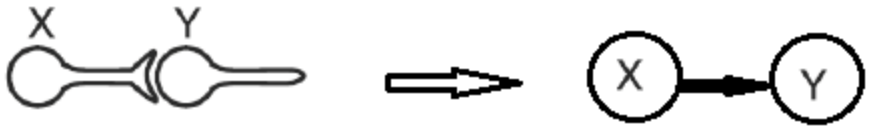
- Additional reading: Milenković et al., *PLoS ONE*, 2009

Different network types: summary



Other biological networks

- Neuronal synaptic connection networks



- Brain functional networks
  - Simultaneous (correlated) activities of brain regions during a task
- Ecological food webs

- Phylogenetic networks (trees)
  - Evolutionary relationships between species

Other biological networks

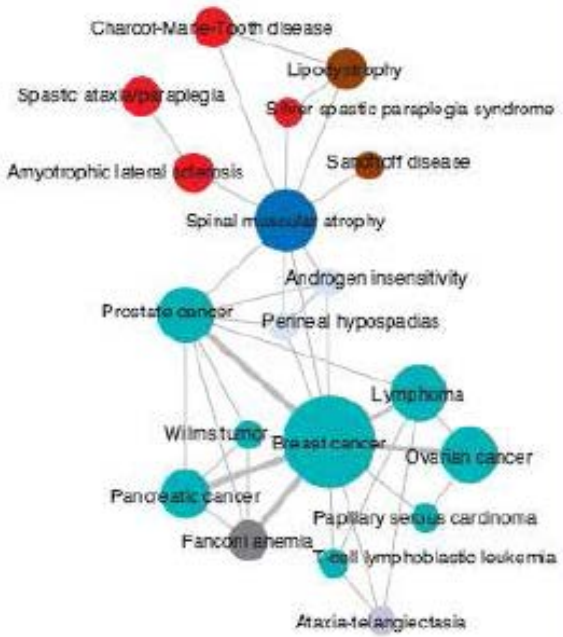
- Correlation networks (e.g., gene co-expression)
  - Different from transcriptional regulation networks
  - Not a direct result of experiments
  - Determined by:
    - Collecting large amounts of high-throughput data
    - Calculating the correlations between all elements
  - Biolayout Express 3-D: a tool for generating correlation networks

Other biological networks

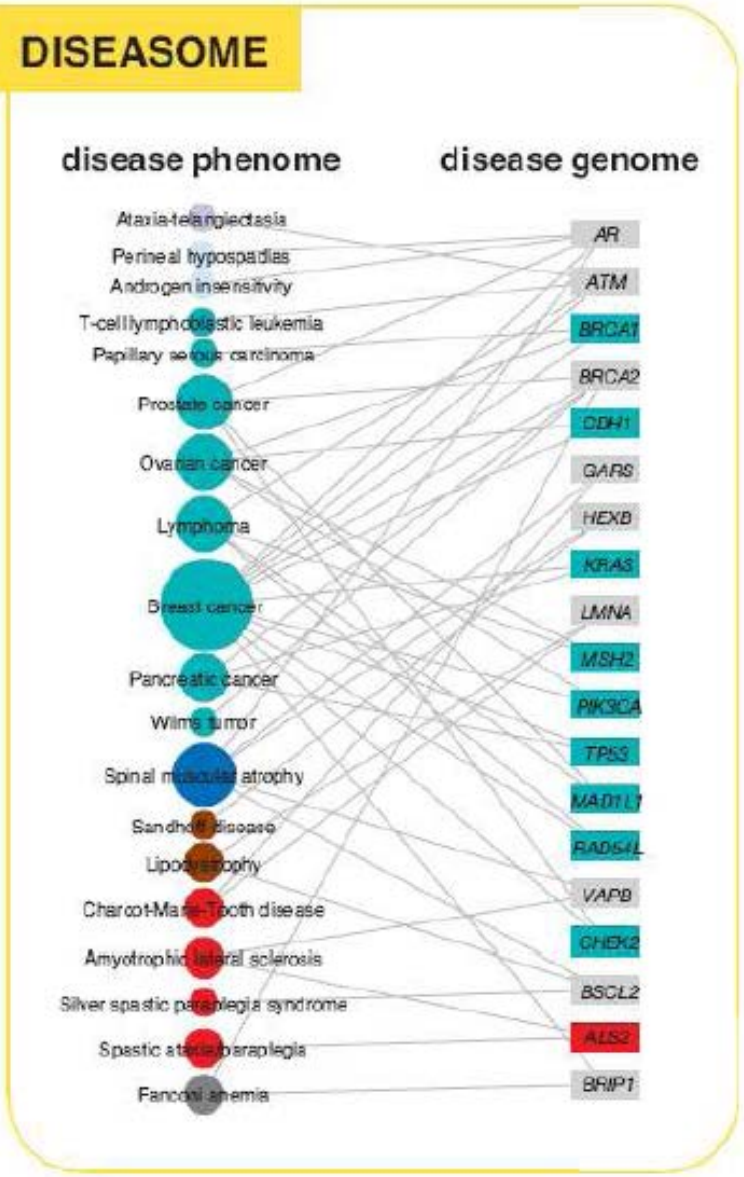
- Disease – “disease gene” association networks
  - Link diseases that are caused by the same gene
  - Link genes if they cause the same disease

- Drug – “drug target” association networks
  - Link drugs if they target the same gene (protein)
  - Link genes (protiens) if they are targeted by the same drug

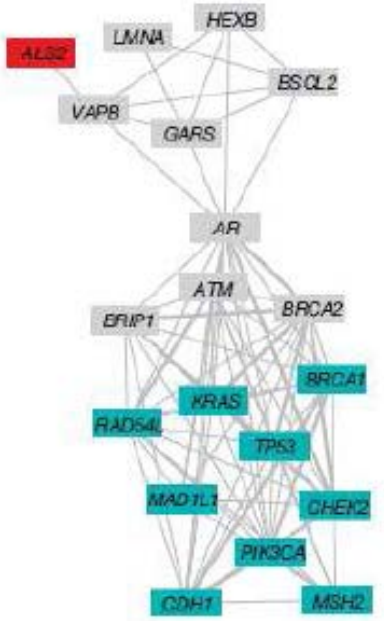
Human Disease Network (HDN)



DISEASOME



Disease Gene Network (DGN)



Goh K et al., The human disease network. *PNAS* 2007, 104(21):8685–8690.

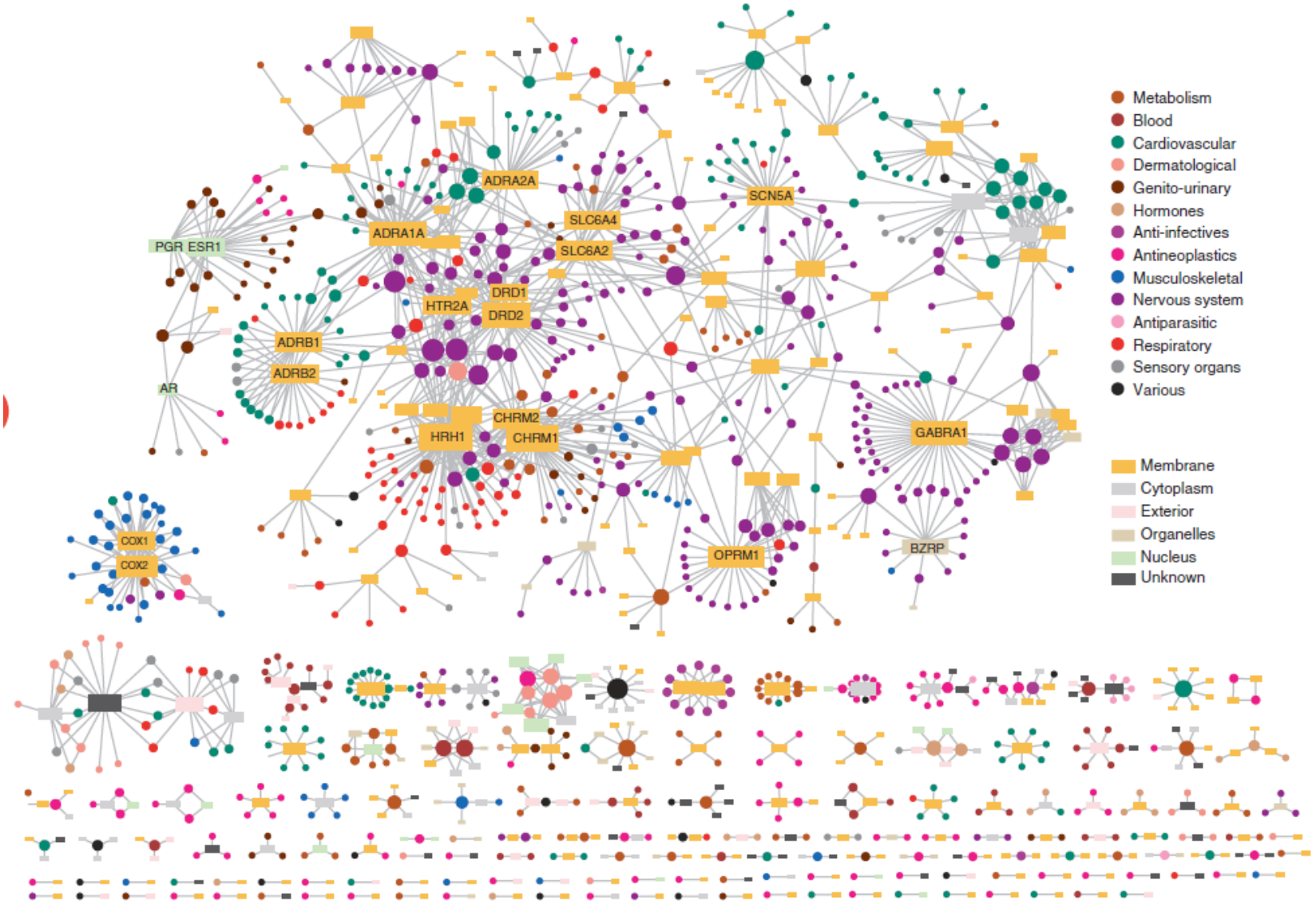


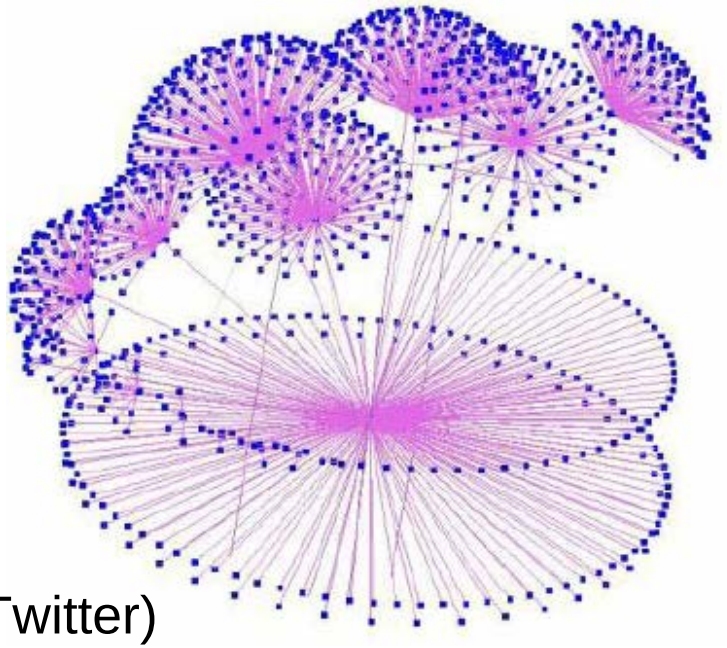
Figure 2 Drug–target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.

Other biological networks

Further readings

- Neuronal synaptic connection networks
  - White J et al., “The structure of the nervous system of the nematode C. elegans”, *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences,* 1986, 314:1340.
- Brain functional networks
  - Kuchaiev, O et al., “Structure of Brain Functional Networks”, *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009.
- Ecological food webs, phylogenetic networks, correlation networks
  - Junker and Schreiber, “Analysis of Biological Networks,” *Wiley*, 2008.
- Disease-disease gene association networks
  - Goh K et al., The human disease network. *PNAS* 2007, 104(21):8685–8690.
- Drug-drug target networks
  - Yidirim MA et al., Drug-target network. *Nature Biotechnology* 2007, 25(10).

Other real-world networks



- Techonlogical networks:
  - WWW
  - Internet
  - Electric circuits
  - Software call graphs
- Transportation networks:
  - Roads, airlines, railways
- Social networks:
  - Friendships/relationships (Facebook, Twitter)
  - Collaborations between scientists/movie stars
  - Spread of infections and diseases
  - Economic networks
  - Relationships between organizations (companies, NGOs, etc.)
  - City/country trading relationships
  - Migrations
  - Disaster response networks

Other real-world networks

- All use similar analysis and modeling tools, BUT
- We need to be application-specific
  - Some problems might be computationally hard in general, but easy for a particular application
  - E.g., finding isomorphism between *trees* (graphs with no cycles) can be done in linear time, but it is hard on graphs in general
- This is one of the reasons why it is important to find a *network model* (will be defined later) to which a real-world network belongs
  - Only with a good model, a network can be reproduced
  - And only then it can be understood

Topics

- Introduction to biology (cell, DNA, RNA, genes, proteins)
- Sequencing and genomics (sequencing technology, sequence alignment algorithms)
- Functional genomics and microarray analysis (array technology, statistics, clustering and classification)
- Introduction to biological networks
- Introduction to graph theory
- Network properties
  - Network/node centralities
  - Network motifs
- Network models
- Network/node clustering
- Network comparison/alignment
- Protein 3D structure / Network data integration
- Software tools for network analysis
- Interplay between topology and biology

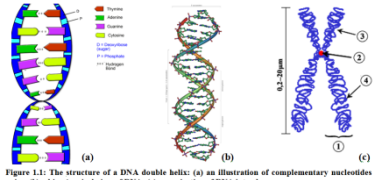


Figure 1.1: The structure of a DNA double helix: (a) an illustration of complementary nucleotides pairs; (b) a bio-chemical view of DNA; (c) organization of DNA into chromosome.