1. Introduction to Molecular & Systems Biology

EECS 600: Systems Biology & Bioinformatics, Fall 2008

Instructor: Mehmet Koyuturk

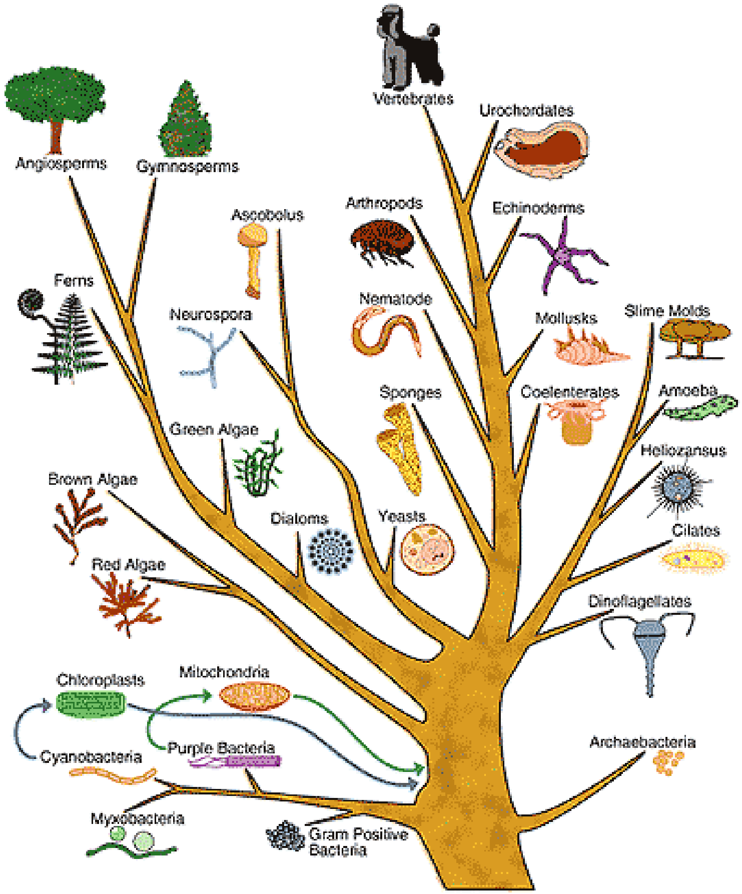
Case Western Reserve University est. 1826

Life

- There is no universal definition of life
  - The structural and functional unit of all living organisms is the cell
  - Living beings use energy to produce offsprings
  - Living beings feed on negative entropy
- Fundamental properties
  - Diversity
  - Unity
- In biology, almost every rule has an exception
  - Are viruses a form of life?

Evolution

- All organisms are part of a continuous line of ancestors and descendants
- Key principles
  - Self-replication: Inheritance of characters
  - Variation: Diversity and adaptation
  - Selection: Not all variation goes through
- Evolution is key to understanding the principles that underlie life



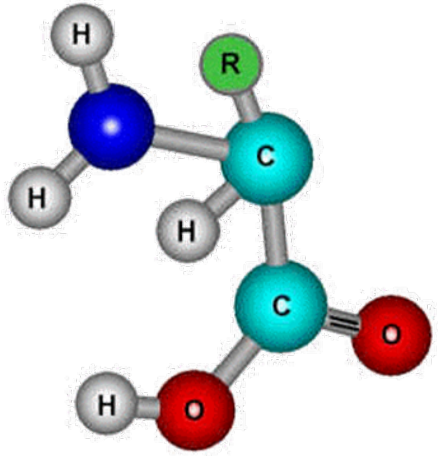
Molecular Biology

Structure & Function

- *Structure:* Physical composition and relationships of a molecule, cell, organism
- *Function:* The role of the component in the process of life
- The main function: Turn available matter & energy into offsprings
- Required structural components
  - Boundaries to separate organism from environment
    - Membranes, composed of lipids
  - Storage medium for inheritable characteristics
    - Chromosomes
  - All other materials necessary for survival and reproduction
    - Cytoplasm

Molecules

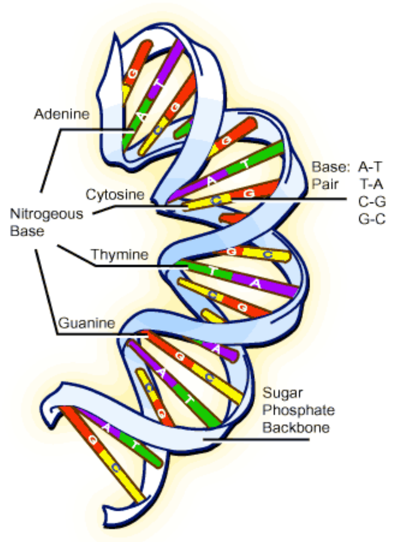
- Small molecules
  - Source of energy or material, structural components, signal transmission, building blocks of macromolecules
    - Water, sugars, fatty acids, amino acids, nucleotides

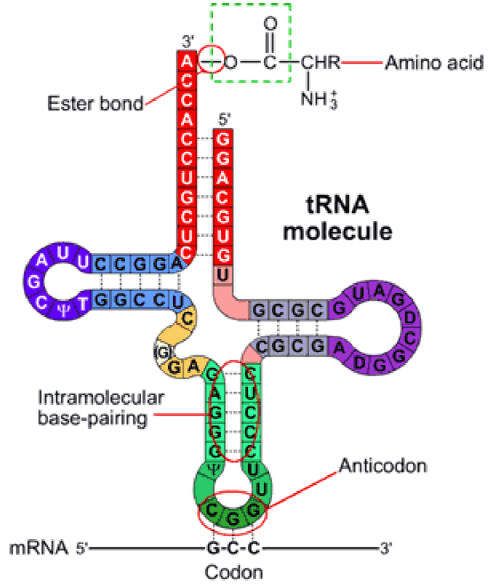


- Proteins
  - Main building blocks and functional molecules of the cell
    - Structure, catalysis of chemical reactions, signal transduction, communication with extracellular environment

Molecules

- DNA
  - Storage and reproduction of information
- RNA
  - Key role in transformation of genetic information to function





The Central Dogma

DNA
- Transcription

→

RNA
- Translation

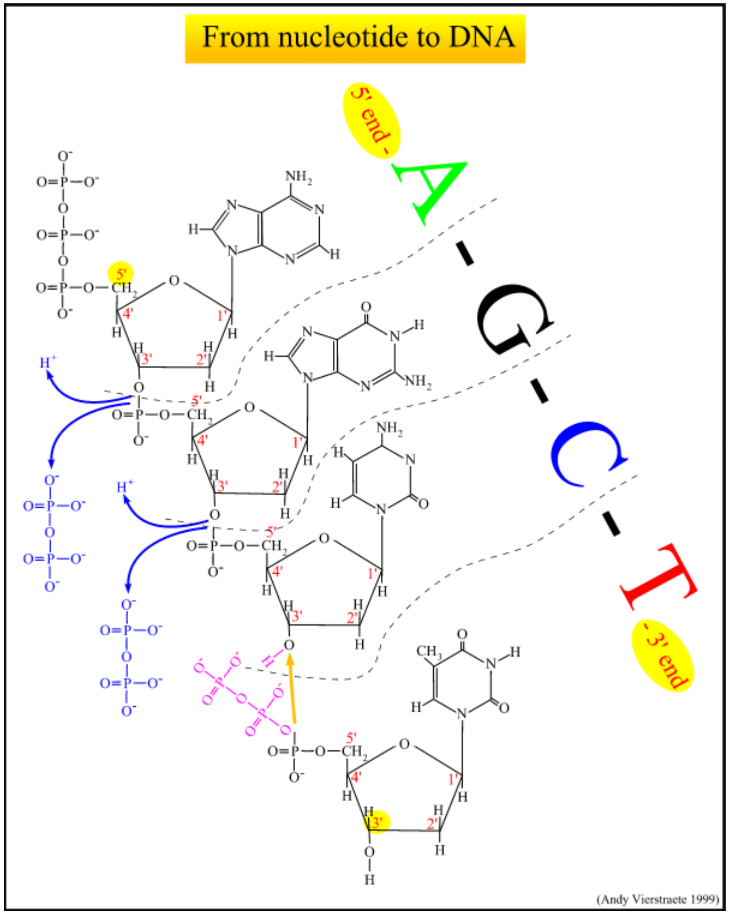
→

Proteins

- Proteins are in action, their structure determines their function
- DNA stores the information that determines a protein's structure
- RNA mediates transformation of genetic information into functional molecules
  - There are functional RNA molecules as well!

DNA

- Sequence of nucleotides
- Backbone is composed of sugars, linked to each other via phosphate bonds
- Each sugar is linked to a base
  - Adenine (A), Thymine(T), Guanine (G), Cytosine (C)
  - Base molecules compose the alphabet of genetic information

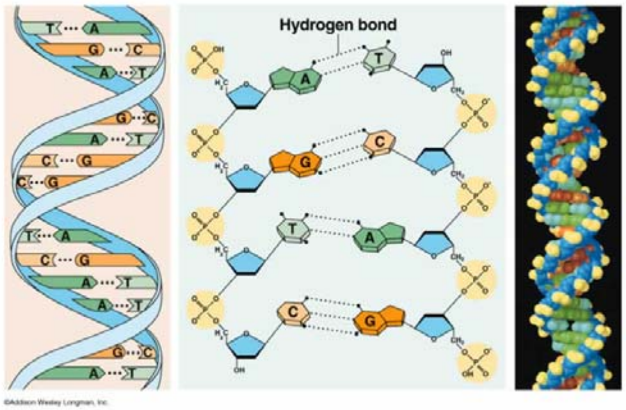


The Double Helix

- DNA is generally found in a double strand form
  - A and T, C and G form hydrogen bonds
  - Two strands with complementary sequences run in opposite directions

    5' A-T-C-T-G-A 3'

    3' T-A-G-A-C-T 5'
  - They are coiled around one another to form double helix structure



Storage of Genetic Information

- Chromosomes
  - Long double stranded DNA molecules
  - In eukaryotes, chromosomes reside in nucleus
  - Humans have 23 pairs of chromosomes
- Genome
  - All chromosomes (and mitochondrial DNA) form the genome of an organism
  - It is believed that almost all hereditary information is stored in the genome
  - All cells in an organism contain identical genomes

Genome Length Statistics

Organism		Genome Size (KB)	No. of Genes
Viruses	MS2	4	
	Lambda	50	~30
	Smallpox	267	~ 200
Prokaryotes	M. genitalium	580	470
	E. coli	4,700	4,000
Eukaryotes	S. cerevisiae (yeast)	12,068	5,885
	Arabidopsis	100,000	20 - 30,000
	Human	3,000,000	~ 100,000
	Maize	4,500,000	~ 30,000
	Lily	30,000,000	

RNA

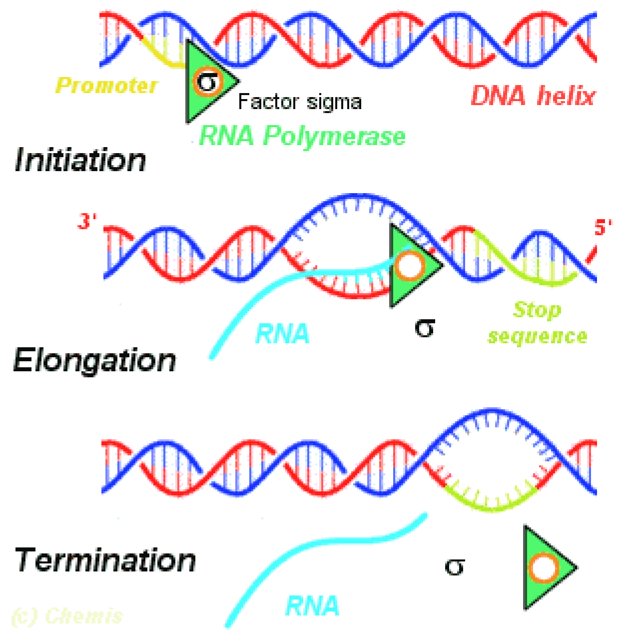
- RNA is made of ribonucleic acids instead of deoxyribonucleic acids (as in DNA)
  - RNA is single-stranded
  - In RNA sequences, Thymine (T) is replaced by Uracil (U)
- mRNA carries the message from genome to proteins
- tRNA acts in translation of biological macromolecules from the language of nucleic acids to aminoacids
- Several different types of RNA have several other functions
  - RNA is hypothesized to be the first organic molecule that underlies life

Proteins

- Proteins are chains of aminoacids connected by peptide bonds
  - Often called a polypeptide sequence
  - There are 20 different types of aminoacid molecules (each aminoacid in the chain is commonly referred to as a residue)
- Proteins carry out most of the tasks essential for life
  - Structural proteins: Basic building blocks
  - Enzymes: Catalyze chemical reactions that enable the mechanism transform forms of matter and energy to one another (metabolism)
  - Transcription factors: Genetic regulation, *i.e.*, control of which protein will be synthesized to what extent

Proteins: Synthesis, Structure, Function

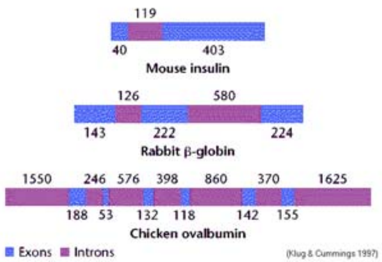
Transcription



- One strand of DNA is copied into complementary mRNA
- Carried out by protein complex RNA polymerase II

Splicing

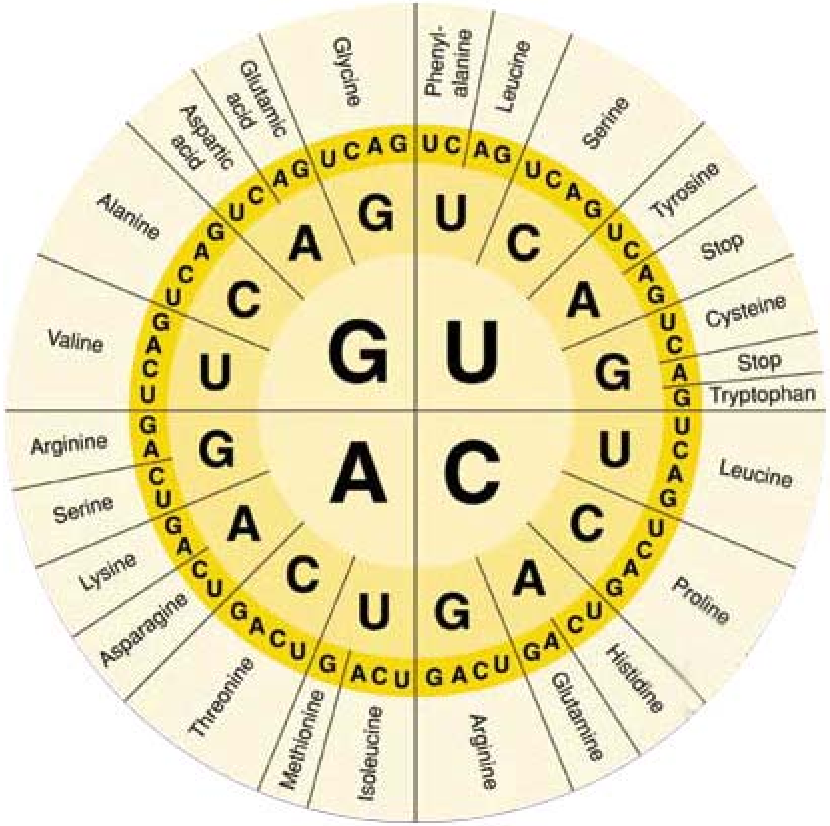
- A gene is a continuous stretch of genomic DNA from which one (or more) type(s) of protein(s) can be synthesized
- Genes contain coding regions (exons) separated by non-coding regions (intron)



- Introns are removed from pre-mRNA through a process called splicing, resulting in mRNA
- *Alternative splicing*: Different combinations of introns and exons may be used to synthesize different proteins from a single gene

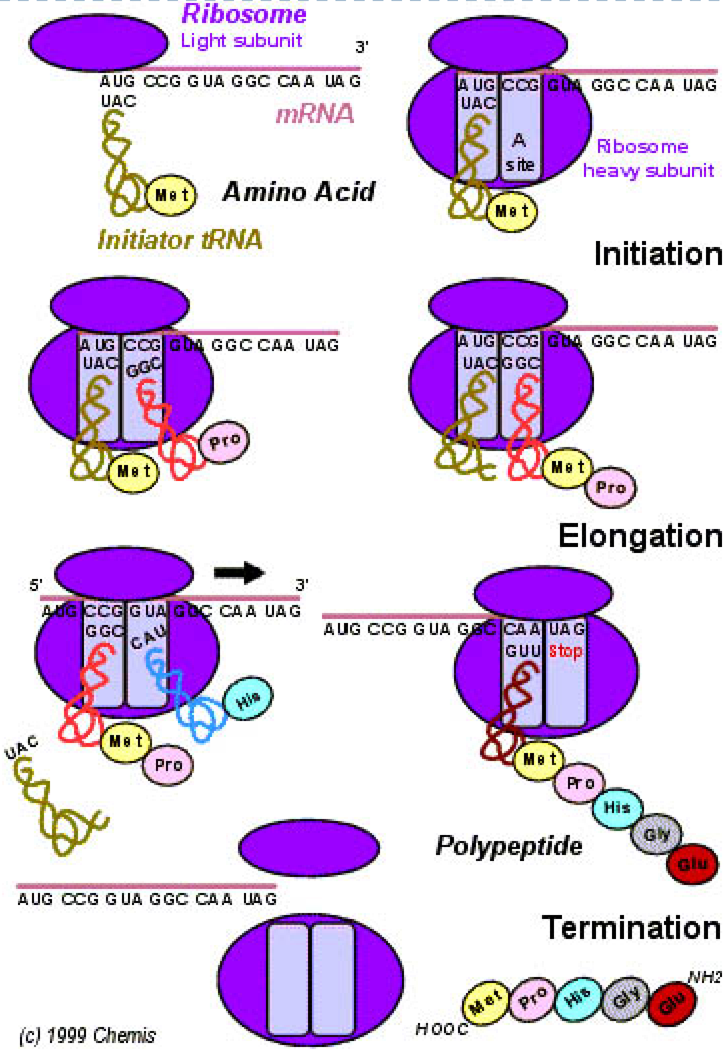
Genetic Code

- There are 4 different types of nucleotides, 20 different types of aminoacids
- A contiguous group of 3 nucleotides (codon) codes for a single aminoacid
  - 64 possible combinations, multiple codons code for a single aminoacid
  - There are codons reserved for signaling termination



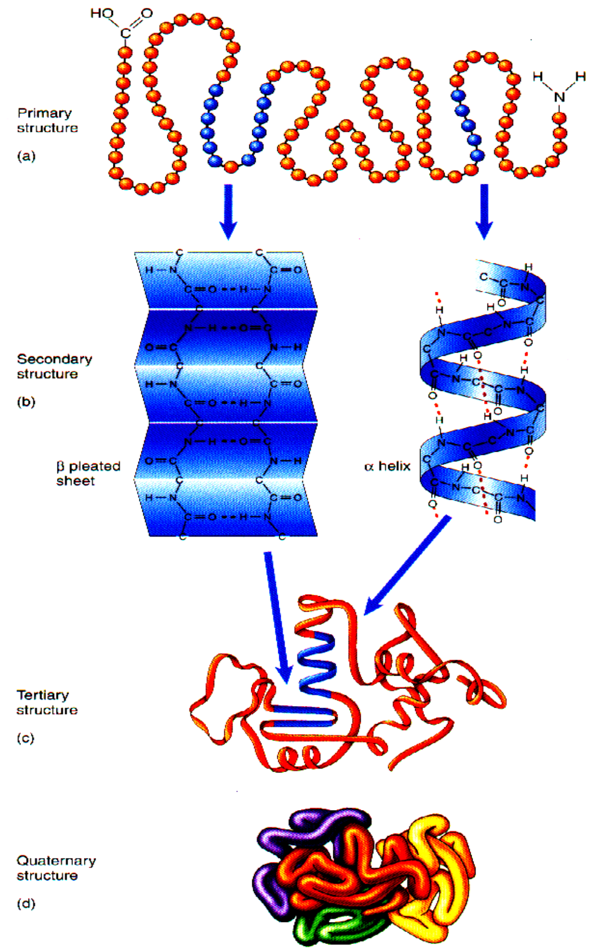
Translation

- The process of synthesizing a protein, using an mRNA molecule as template
- Carried out in ribosome
- tRNA
  - Cloverleaf structure, three bases at the hairpin loop form an *anticodon*
  - A single type of aminoacid may be attached to the 3' end of a single tRNA
- There is no tRNA with a stop anticodon



Protein Structure

- Primary structure
  - The aminoacid sequence and the chemical enviroment determine a protein's 3D structure
- Secondary structure
  - Alpha helices, beta sheets
- Tertiary structure
  - Folding: relatively stable 3D shape
  - Domain: functional substructure
- Quarternary structure
  - More than one aminoacid chain
- Structure is key in function



Protein Function

- Three aspects
  - Activity: What does the protein do? (e.g., an enzyme might break a particular kind of bond)
  - Specificity: The ability to act on particular targets
  - Regulation: Activity may be modulated by other molecules (on or off?)
- Each of these aspects is realized by a corresponding aspect of structure
- In this course, we will focus on analyzing data that provide clues on how proteins cooperate to perform complex functions

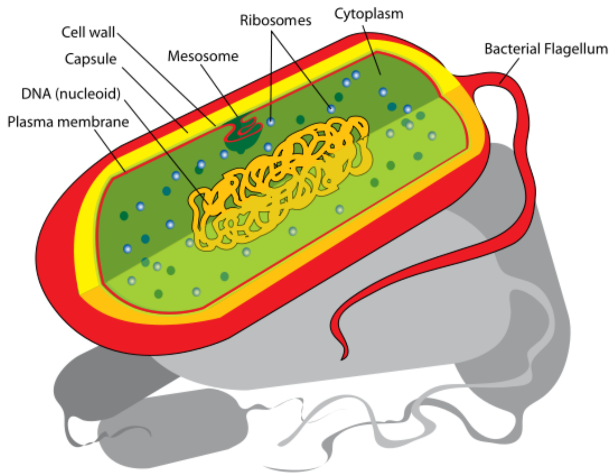
Domains of Life

Domains of Life

- Three cell types
  - Prokaryotes
  - Eukaryotes
  - Archaea
- Similarities
  - All have DNA as genetic material
  - All are membrane bound
  - All have ribosomes
  - All have similar basic metabolism
  - All are diverse in forms

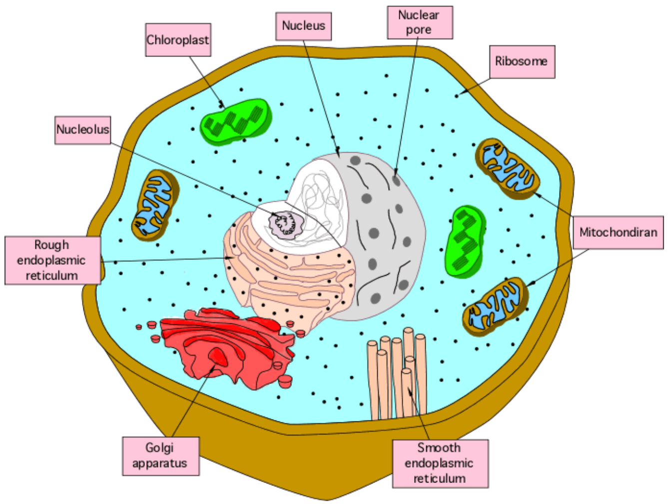
Prokaryotes

- Their genetic material is not membrane bound
- They do not have membrane bound cellular compartments
- They contain only a single loop of DNA (no chromosomes)
- All prokaryotes are unicellular (they do form colonies, though)
- They are ubiquitous
- All bacteria are prokaryotes
  - *E. coli, H. Pylori*



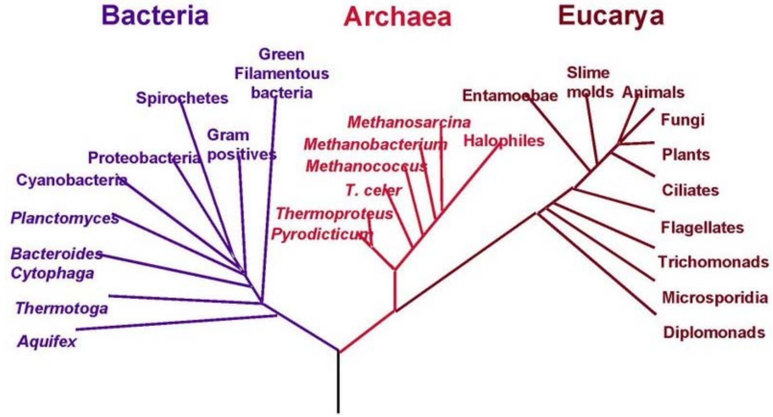
Eukaryotes

- Cells are organized into complex structures by internal membranes and a cytoskeleton
  - Nucleus is the most characteristic membrane bound structure
  - Genetic material is stored in chromosomes
- All multicellular organisms are eukaryotes
  - Can be unicellular as well
- Plants, animals, fungi, protists
  - Human (*H. sapiens*)
  - Mouse (*M. musculus*)
  - Weed (*A. thaliana*)
  - Fly (*D. melanogaster*)
  - Baker's yeast (*S. cerevisiae*)



Archaea

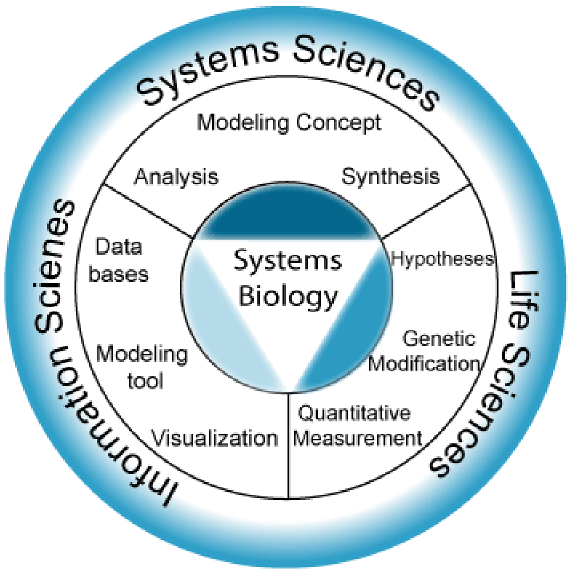
- Most recently discovered domain of life
- Generally extremophile
- Microorganisms like prokaryotes, therefore sometimes referred to as archaebacteria
  - Similar to prokaryotes in cell structure and metabolism
  - Genetic transcription and translation is more similar to that in eukaryotes



Systems Biology

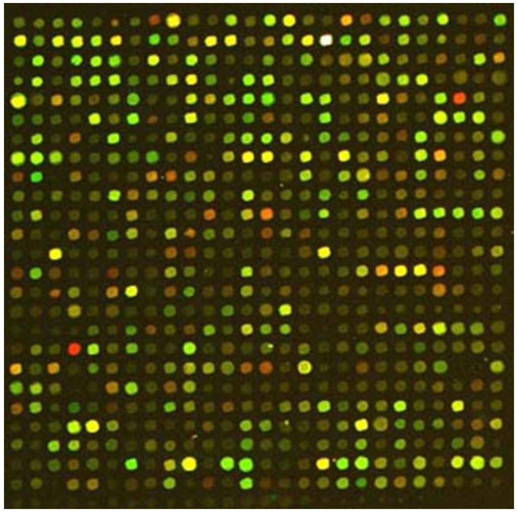
Why Systems Biology?

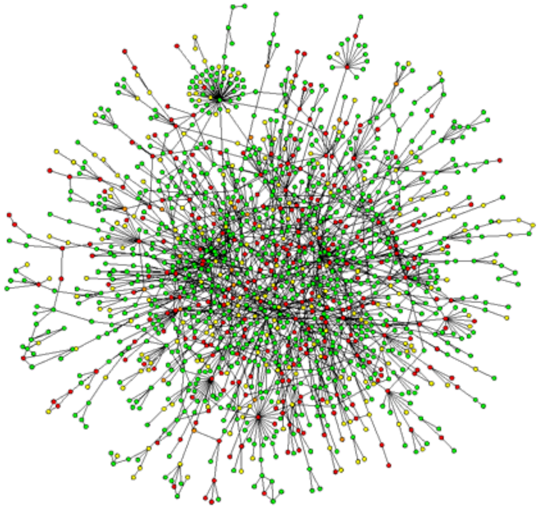
- "To understand biology at the system level, we must examine the **structure** and **dynamics** of **cellular and organismal function**, rather than the characteristics of **isolated** parts of a cell or organism." (Kitano, *Science*, 2002)
- Cell is not just an assembly of genes and proteins
- Systems biology **complements** molecular biology



Systems Perspective is Possible Today

- Progress in molecular biology
  - Genome sequencing
    - Information on underlying molecules
  - High-throughput measurements
    - Comprehensive data on system state





An Analogy

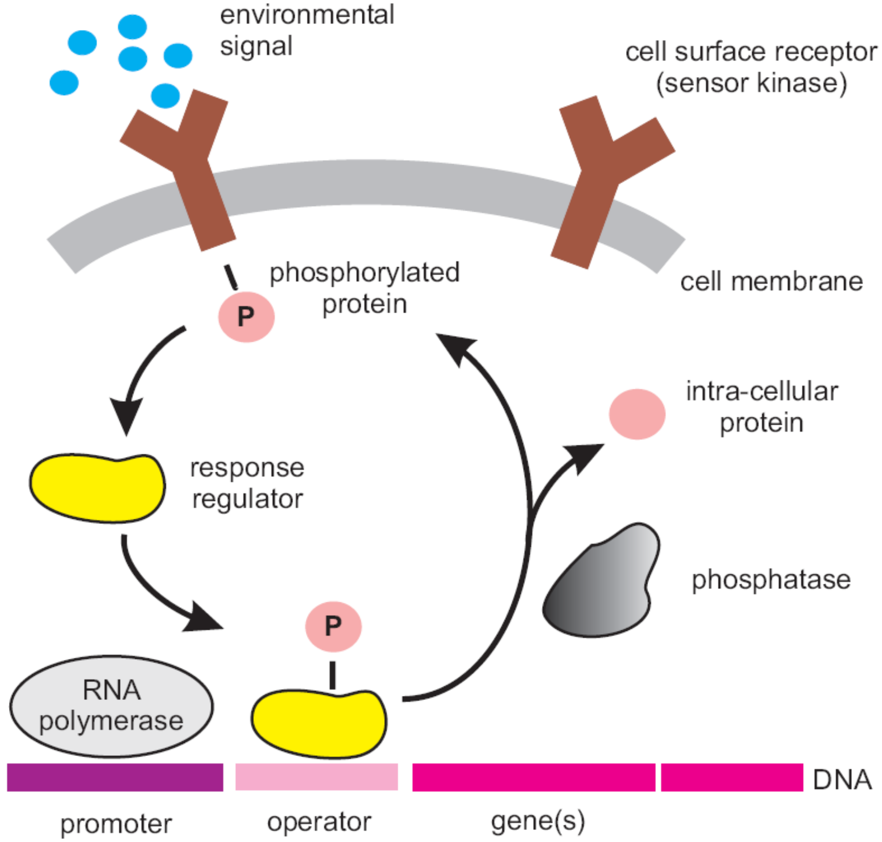
- Understanding how an airplane works
  - What do we learn if we list all parts of an airplane?
    - Identifying single genes or proteins
  - How are these parts assembled to form the structure of an airplane?
    - This tells us on what parts may have an effect what parts
    - Identifying regulatory effects of genes on one another, protein-protein interactions, etc.
  - How do individual components dynamically interact?
    - What is the voltage on each signal line?
    - How do voltages on different signal lines effect each other?
    - How do the circuits react when malfunction occurs?

What is a System?

System Concepts

- 1. System structures
  - Topology, wiring, architecture, organization
- 2. System dynamics
  - Behavior over time, under different conditions
- 3. System control
  - Mechanisms that systematically control the state of the cell
- 4. System design
  - Underlying design principles
- All interrelated!

An Example: Cellular Signaling



http://www.informatik.uni-rostock.de/~lin/GC/Slides/Wolkenhauer.pdf

System Structure

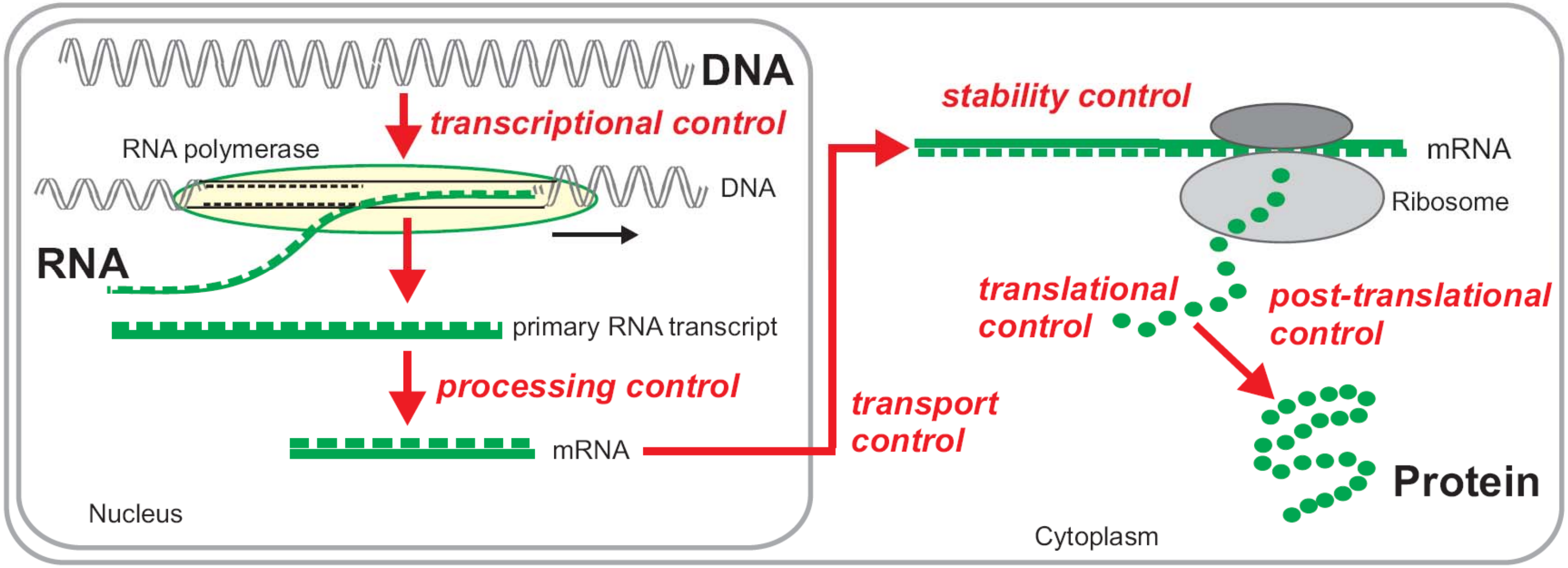
- Wiring, architecture, or organization of the system
  - Protein-protein interactions form a network
    - From direct physical relationships to large-scale orchestration between proteins
    - How are cellular signals are transmitted?
  - Metabolic network represents chains of reactions
  - Gene regulatory networks characterize the "control" of cellular state
- Has to go beyond intracellular wiring
  - How about organization of cells?
- Tools
  - Informatics, data analysis, knowledge discovery

System Dynamics

- The logic of system control in biological systems is *fuzzy*
  - Dimensions of time and space
- How does a system behave over time under various conditions?
  - How do concentrations of biochemical factors influence each other?
  - What is the effect of perturbation?
  - What are the essential mechanisms that underlie specific behaviors?
- Tools
  - Mathematical modeling
  - Simulation

System Control

- Mechanisms that systematically control the state of the cell
  - Robustness, how does the system respond to malfunction?



http://www.informatik.uni-rostock.de/~lin/GC/Slides/Wolkenhauer.pdf

System Design

- Engineering aspects of the system
  - Optimization, use of resources
- Are there general principles?
  - Convergent evolution
  - Evolutionary families of cellular circuitry?
  - “Periodic table” of functional regulatory circuits?
- In most cases, we may not know what we are looking for
  - Data mining & knowledge discovery
  - Pattern identification
  - Statistical evaluation: Which patterns are potentially relevant?

Organization & Dynamics

- Organization tells us about the architecture, but not how that architecture behaves
  - We have a road map, we want to characterize traffic patterns on the roads as well
  - The map is useful, but we need more information and more detailed modeling
- Organization underlies dynamics
  - If we understand network structure, we can start assigning functions on links (how do the gates behave?)
- Nevertheless, understanding of organization and dynamics is an overlapping process
  - Dynamic analysis may provide clues on identifying interactions

Properties of Complex Systems

Properties of Complex Systems

1. Emergence
2. Robustness
3. Modularity

Biological systems demonstrate these properties.

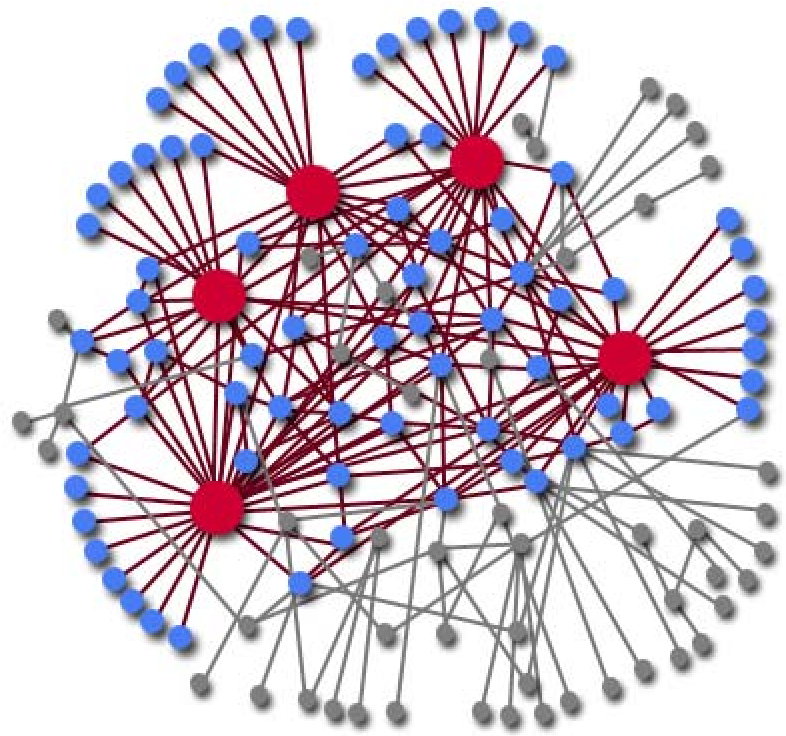
Emergence

- Emergent properties: Those that are not demonstrated by individual parts and cannot be predicted even with full understanding of the parts alone
  - Understanding hydrogen and oxygen is not sufficient to understand water
- Life is an emergent property
  - It is not inherent to DNA, RNA, proteins, carbohydrates, or lipids, but it is a consequence of their actions together
- Systems-level perspective is required to comprehensively understand emergent properties

Robustness

- Phenotypic stability under diverse perturbations
  - Environment, stochastic events, genetic variation
- Properties
  - Adaptation
    - Ability to cope with environmental changes
  - Parameter insensitivity
    - Not affected too much by slight perturbations
  - Graceful degradation
    - Slow degradation of a system's functions after damage (as compared to catastrophic failure)
  - Robustness might also cause fragility

Cost of Robustness



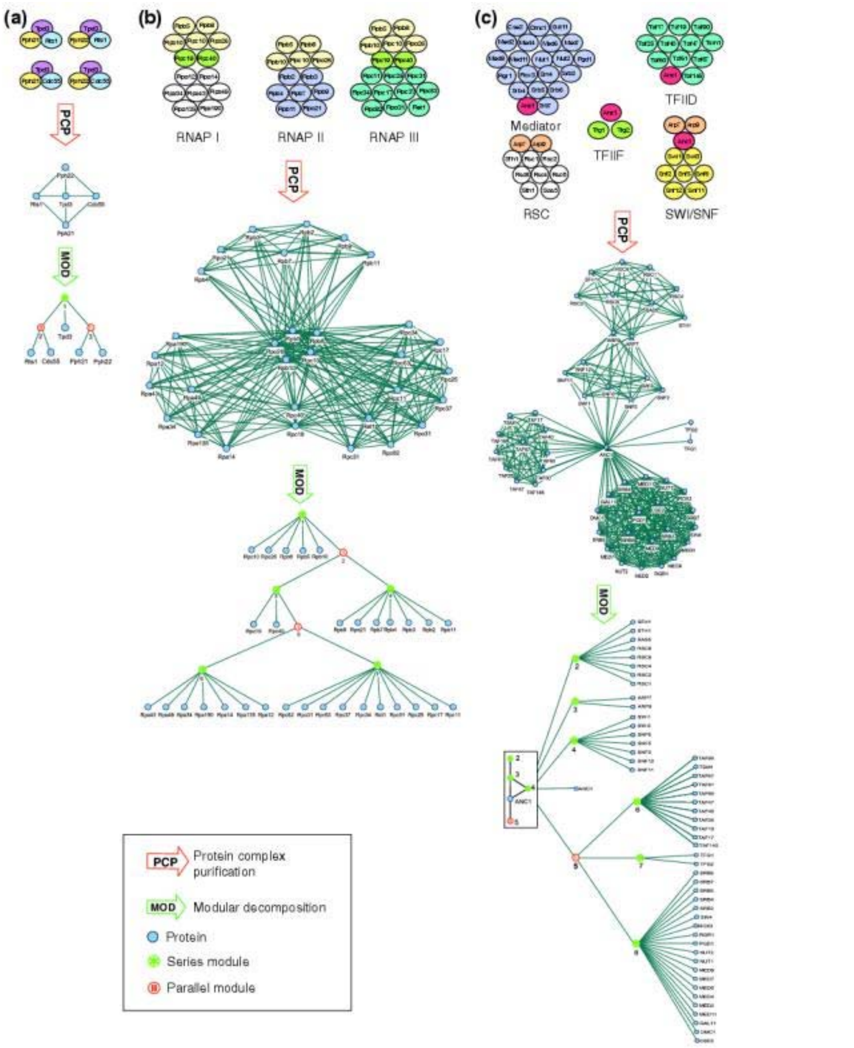
Scale-free networks: Robust against random attacks, vulnerable to targeted attacks

Robustness

- How can robustness be attained?
  - System control
    - Negative feedback: Insulates system from fluctuations imposed by the environment, dampens noise, rejects perturbations
    - Positive feedback: Enhances sensitivity
  - Redundancy
    - Multiple components with equivalent functions, alternate pathways
  - Structural stability
    - Intrinsic mechanisms that promote stability
  - Modularity
    - Sub-systems are physically or functionally isolated
    - Failure in one module does not spread to other parts

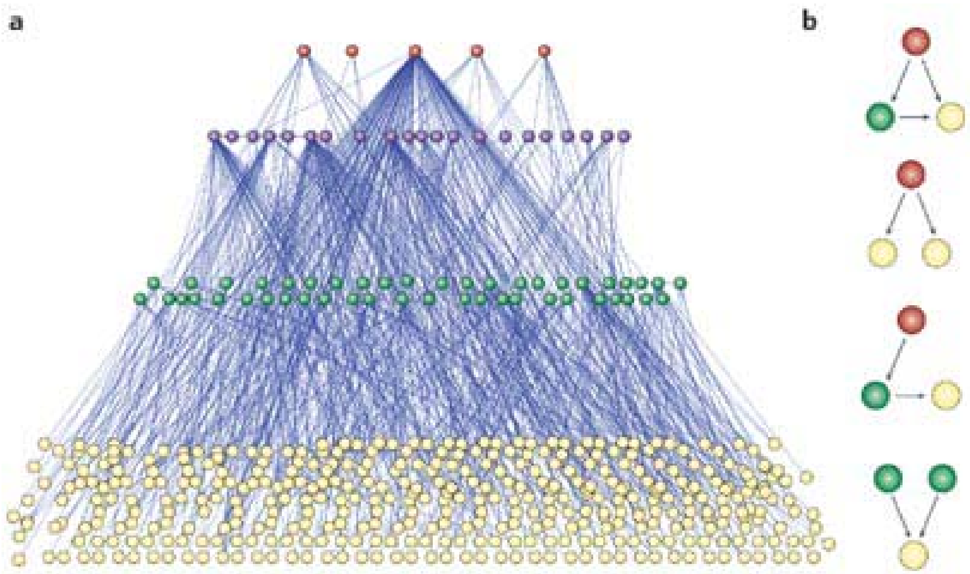
Modularity

- A module is a functional unit, a collection of parts that interact together to perform a distinct function
  - Inputs: signals that influence a module
  - Outputs: signals that are produced by a module



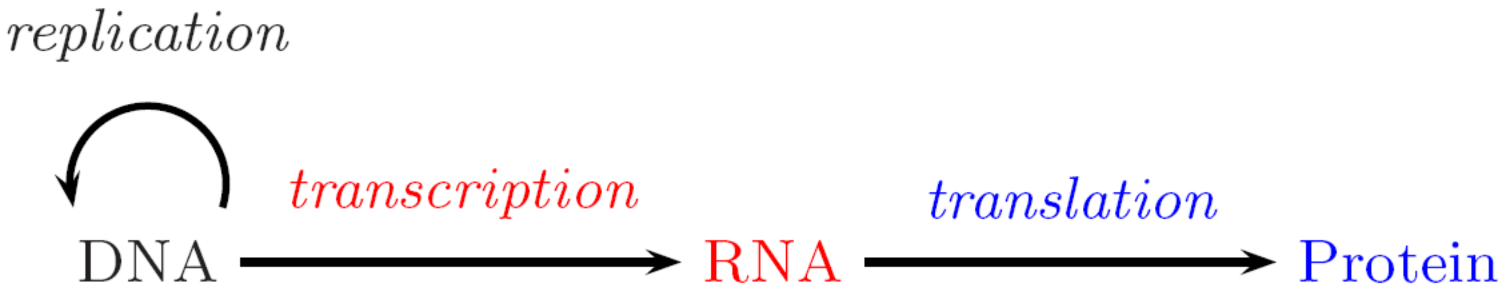
Modularity

- Contributes to robustness
- Contributes to development and evolution
  - Just multiply, rewire, revert a module
- Hierarchical modularity
  - Modules of modules of modules…



Omics of Systems Biology

Central Dogma Revisited



genome transcriptome proteome

http://www.informatik.uni-rostock.de/~lin/GC/Slides/Wolkenhauer.pdf

'Omes and 'Omics

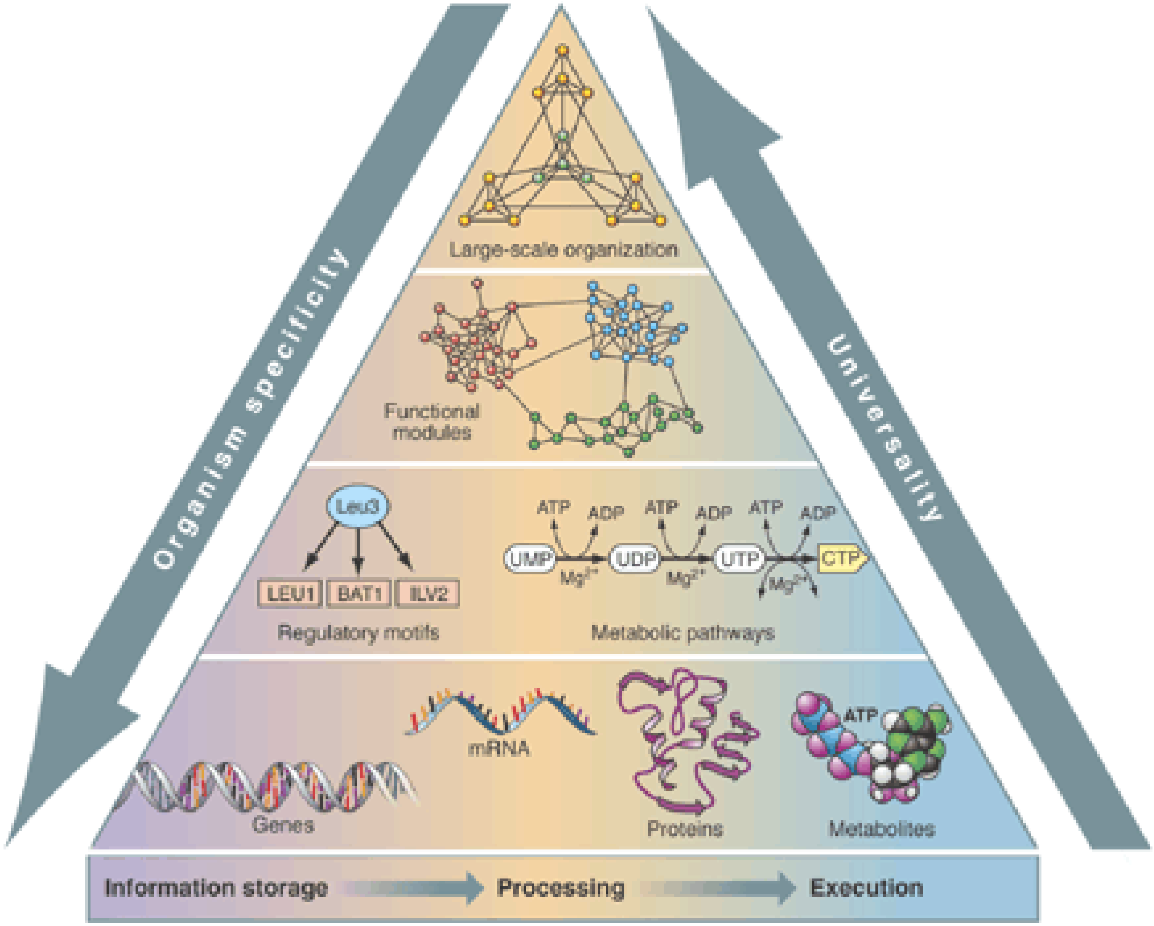
- …'ome: the complete set of …
  - Genome: genes
  - Transcriptome: mRNA (used to measure the state of a cell in terms of gene expression)
  - Proteome: proteins
  - Interactome: molecular interactions
  - Metabolome: chemicals involved in metabolic reactions
- …'omics': the study of…
- High-throughput methods
  - The same experiment is performed on many different molecules (genes, proteins, etc.) in a (partially) automated way
  - Make 'omics possible

Layers of Organization

- Genome
  - Long term information storage
- Transcriptome
  - Retrieval of information
- Proteome
  - Short term information storage
- Interactome
  - Execution
- Metabolome
  - State
- Analogies with computer hard/software?

Levels of Complexity

Life's Complexity Pyramid

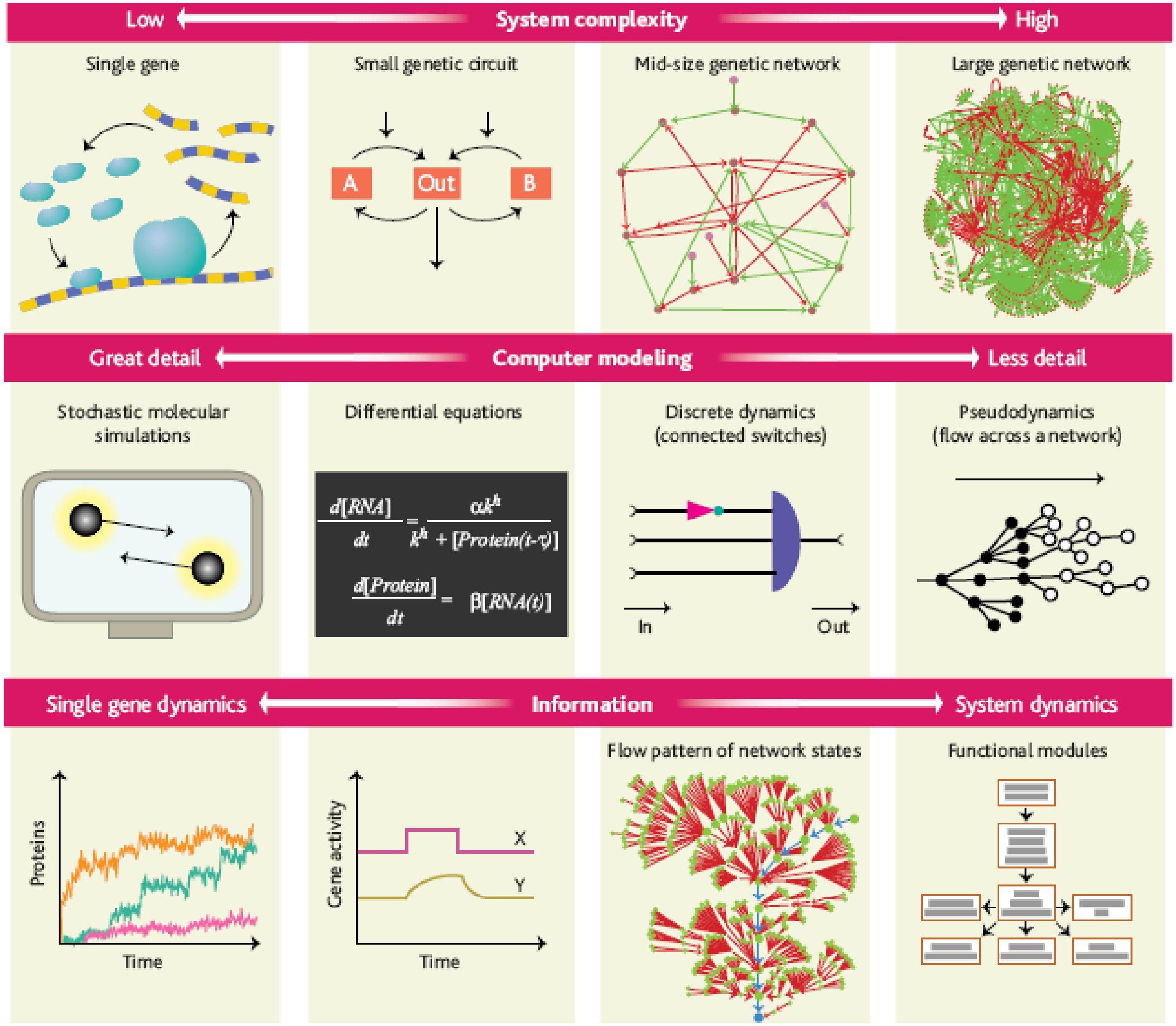


Oltvai & Barabasi, *Science*, 2002

Specificity vs. Universality

- Tendency toward universal as levels coarsen
  - Genes, metabolites, proteins are unique to organism
  - 43 organisms, for which metabolic information is available, share only about 4% of their metabolites
  - Key metabolic pathways are more frequently shared
- Higher degree of universality at module level?
  - Properties appear to be
    - Scale-free, hierarchical nature of wiring
  - Coherent regulatory motifs are common
  - Results on identified "modules" also demonstrate significant conservation
    - Still a lot to explore on modular conservation

Model Resolution



Bornholdt, *Science*, 2005

System Complexity

- Different models, different abstraction, different information, different computational needs
  - Boolean networks
    - General (thousands of genes)
    - Irrelevant to a particular system
    - Simple model
  - Flux networks
    - Specific (a few genes)
    - Relevant only to a particular system
    - Complex model

Level of Detail

- Trade off: Less is more
  - Less low level detail enables understanding at a larger scale
  - Computational limitations
  - Availability of data is an important consideration (e.g., gene expression provides correlation, what about causality?)
- What level of detail do we need?
  - The trajectory of segment polarity network in Drosophila was predicted solely on the basis of discrete binary modeled genes (Albert et al., *J. Theo. Biol.*, 2003)
  - A dynamic binary model of yeast cell cycle genetic network was constructed (Li *et al.*, *PNAS*, 2004)

Comprehensiveness of Data

1. Factor comprehensiveness
   - Number of components that can be inspected at a time
   - How many mRNA transcripts in an assay?
2. Time-line comprehensiveness
   - Time frame within which measurements are made
   - Longitude, resolution
   - Correlation vs causality
3. Item comprehensiveness
   - Simultaneous measurement of multiple items
   - mRNA & protein concentrations, phosporylation, localization

Studying Systems Biology

What Systems Biology Offers

- How genotype determines phenotype
  - Genes (and regulatory elements) have combinatorial effect on phenotype
  - Transcription factors combinatorially determine which genes are expressed
  - What determines the state of the cell?
  - What makes a difference during development?
  - Regulation, cooperation, redundancy
- Drug design
  - A ligand might influence multiple factors
  - A multiple drug system may guide a malfunctioning system to desired state with minimal *effects*

Challenges

- Data quality and standardization
  - Incompleteness
  - Not standardized or properly annotated
  - Quality is uncertain
- How do we use available data?
  - Hypotheses?
  - Iterative refinement
- Technology
  - Limited "comprehensiveness"
  - We cannot measure many things, so we have to make inference
    - Transient interactions

Case Western Reserve University est. 1826

Challenges

- Data Integration
  - How do different sources of data relate?
  - Interactions
    - Two-hybrid
    - Co-expression
    - Phylogenetic profiling
    - Linkage
    - What is an interaction?