Introduction to System Biology

Dr. Carlo Cosentino
School of Computer and Biomedical Engineering
Department of Experimental and Clinical Medicine
Università degli Studi Magna Graecia
Catanzaro, Italy
carlo.cosentino@unicz.it
http://bioingegneria.unicz.it/~cosentino

Outline of the Lectures

- Introduction to Systems Biology
    - Basic Principles
    - Biology in a nutshell
    - Getting experimental data
- Modeling biochemical reactions
    - Deterministic Models
    - Stochastic Models
- Biological Networks
    - Types of biological networks
    - Dynamical models of biological networks
    - Inference of biological networks
- Analysis and Simulation of Biological Systems
    - A case-study: The cell cycle
    - Modeling the cell cycle
    - Analysis of biological models
    - Software tools

Introduction to System Biology

- Basic Principles
- Biology in a nutshell
- Experimental Techniques

What is Systems Biology?

- Systems biology is concerned with the study of biological functions and mechanisms, underpinning inter- and intra-cellular dynamical networks, by means of signal- and system-oriented approaches

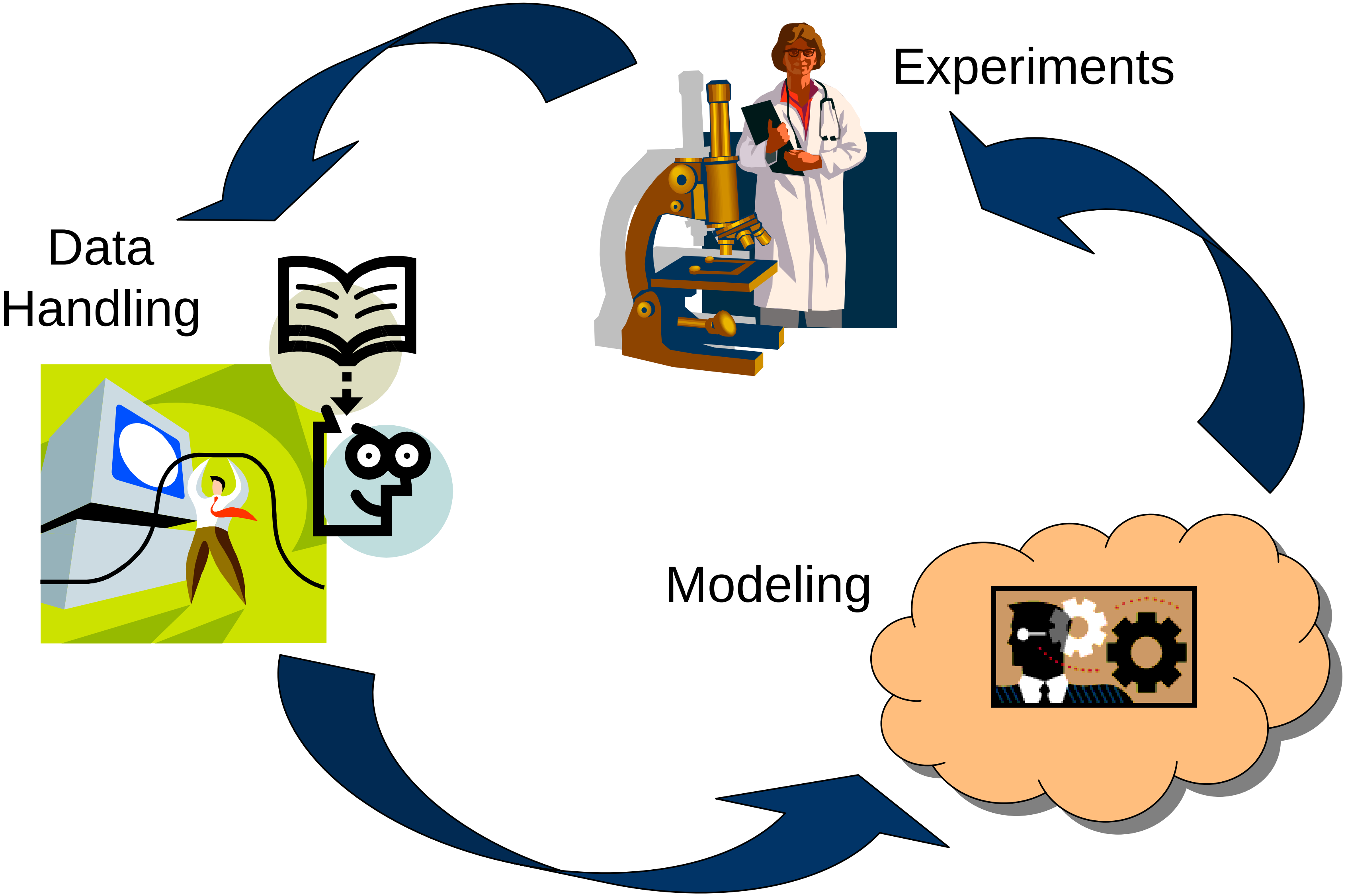
- “*Life is an emergent, rather than an immanent and inherent, property of matter. Although it arises from the material world, it cannot be reduced to it*” (E. Schrödinger)

- “*Science is built up of facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house*” (H. Poincaré)

Systems Biology Approach

- A systems biology approach means
  - Investigating the components of cellular networks and their interactions
  - Applying experimental high-throughput and whole-genome techniques
  - Integrating computational and theoretical methods with experimental efforts

An Iterative Approach



Experiments

Data Handling

Modeling

Main Related Disciplines

- Biology!
- Biotechnology
- Mathematics and Statistics
- Physics and Chemistry
- Information Science
- Engineering (Biomedical, Chemical, Computer, Systems & Control, Electronic)
- …

Foundations

- Improved biological knowledge with the prospect of utilization in biotechnology and health care
- New experimental techniques in genomics and proteomics
- Classical mathematical modeling of biological processes
- Computer power for simulation of complex systems
- Storage and retrieval capability in large databases and data mining techniques
- Internet as the medium for the widespread availability from multiple sources of knowledge

Goals

- Models to unveil mechanisms causing alterated phenotypes and devise novel therapies and drugs
- Predictive tools to design cells with desired properties cheaply and reliably
- Individualized and predictive medicine

Models are not Real, though Reality can be Modeled

- In almost any case, models are only rough representations of their biological counterparts
- Nevertheless, models enable to
  - Elucidate network properties
  - Check the reliability of basic assumptions
  - Uncover lack of knowledge and requirements for clarification
  - Create large repository of current knowledge, formalized in a non-ambiguous way and including quantitative data

What is a Model?

- It depends on whom you ask…
    - Genetist: the mouse family Ts65DN serves as a model for human trisomy 21
    - Chemist: a reaction network, described by dots (for metabolites) and arrows (for reactions)
    - Mathematician/Engineer: the same reaction network can be modeled by a system of nonlinear ODEs
- Abstract representation of objects or processes that explains features of these objects or processes

Mathematical Models

- Biological processes can be described in mathematical terms, however
    - A biological object can be investigated by means of different experimental methods
    - Each biological process can be described through different (mathematical) models
    - The choice of a mathematical model or an algorithm depends on the problem, the purpose, and the intention of the investigator
    - Modeling has to reflect essential properties of the system: Different models may highlight different aspects of the same instance

Model Development

- Formulation of the problem:
    - Identify the specific questions that shall be answered, along with background, problem and hypotheses
- Available Knowledge:
    - Check and collect quantitative and structural knowledge
        - Components of the system
        - Interaction map and kind of interactions
        - Experimental results with respect to phenotypic responses against different stimuli (gene knockout, RNAi, environmental conditions)

Model Development (cont'd)

- Selection of model structure:
  - Level of description (atomistic, molecular, cellular, physiological)
  - Deterministic or stochastic model
  - Discrete or continuous variables
  - Static, dynamical, spatio-temporal dynamical
- Robustness/Sensitivity Analysis:
  - Test the dependence of the system behavior on changes of the parameters
    - Numerical simulations
    - Bifurcation analysis

UMG
Dubium sapientiae initium

Model Development (cont'd)

- Experimental Tests
  - Hypotheses driven
  - Choice of parameters to be measured, different types of experiments, number of samples and repetitions, …
- Assessment of the agreement and divergences between experimental results and model behavior
- Iterative refinement of the hypotheses (and of the model)

Data Integration

- Observation of biological phenomena is restricted to the granularity and precision of the available experimental techniques
- A strong impulse to the development of a systematic approach in the last years has been given by the new high-throughput biotechnologies
  - Sequencing of human and other genomes (genomics)
  - Monitoring genome expression (transcriptomics)
  - Discovering protein-protein and -DNA interactions (proteomics)
- Different types of information need to be integrated

Issues in Data Integration

- Data representation and storage:
  - (too) Many databases (GO, KEGG, PDB, Reactome…)
  - XML-like annotation languages (SBML, CellML)
- Information retrieval
  - Tools for retrieving information from multiple remote DBs
- Data correlation
  - Find the correlation between phenotypes and genomic/proteomic profiles
  - Statistics, data mining, pattern analysis, clustering, PCA, …

Interdisciplinary Information Exchange

- The interdisciplinary nature of systems biology requires the exchange of information among scientists from different fields
- Mathematical formulas have to be made understandable for biologists
- People acquainted with the rigid rules of mathematics and computers have to understand the diversity of biological objects and the uncertainty in the outcome of experiments

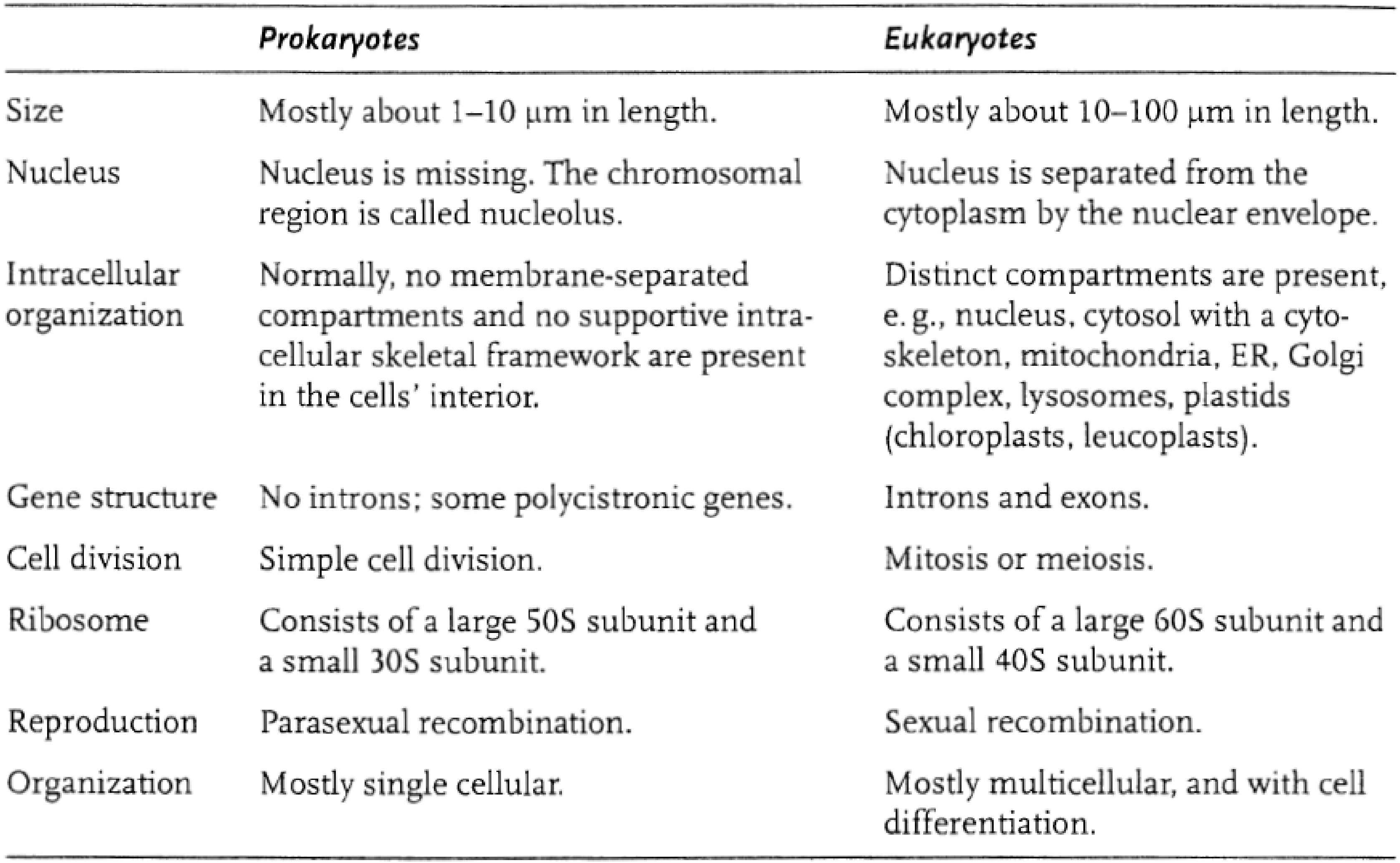
Outline

- Basic principles
- Biology in a nutshell
- Experimental techniques

Branches of Biology

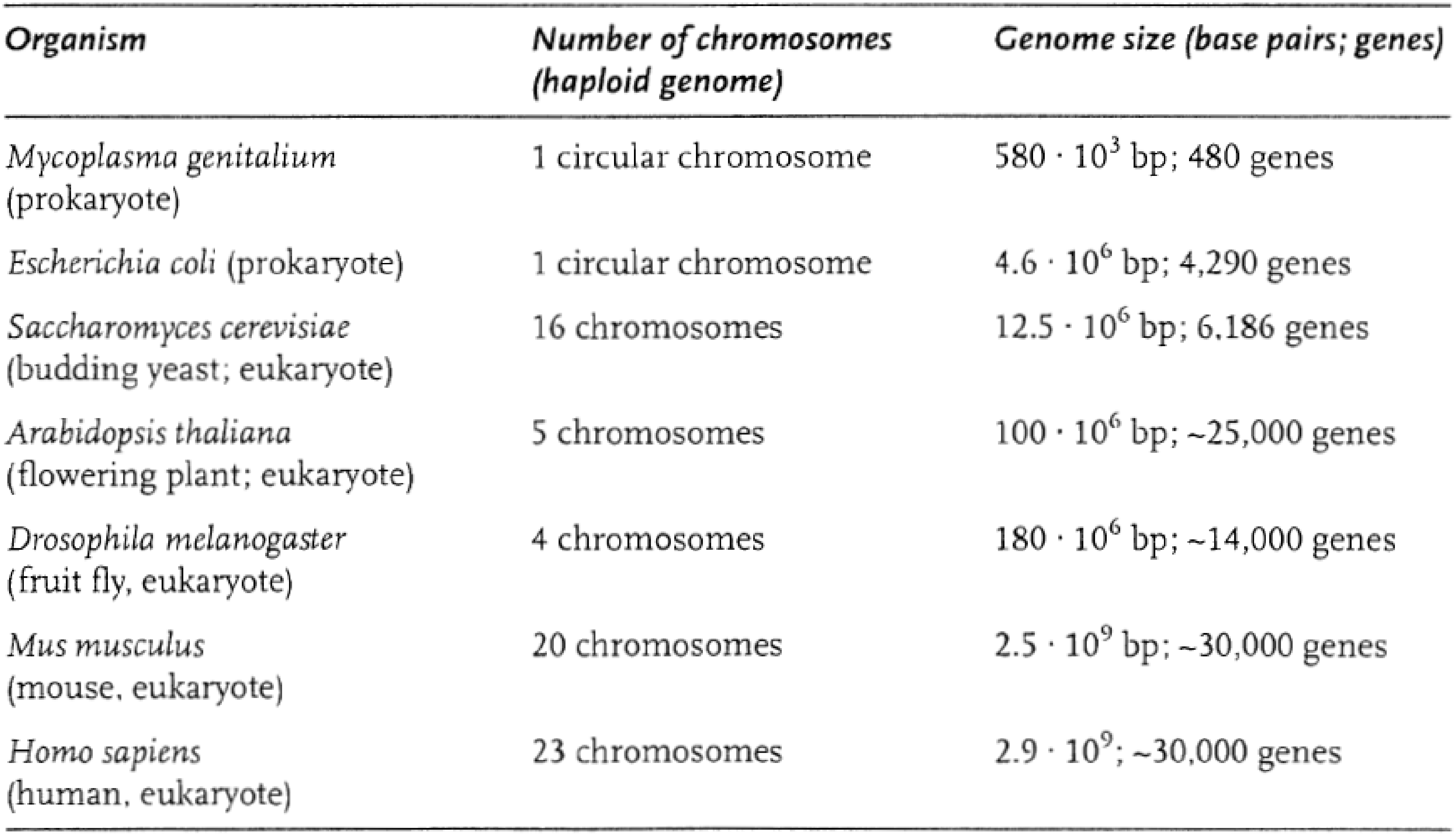
- Biology is the science that deals with living organisms and their interrelationships between each other and their environment in light of the evolutionary origin
  - Biochemistry
  - Molecular biology
  - Genetics
  - Physiology
  - Morphology
  - Cytology
  - Ecology
  - …

Prokaryotes and Eukaryotes



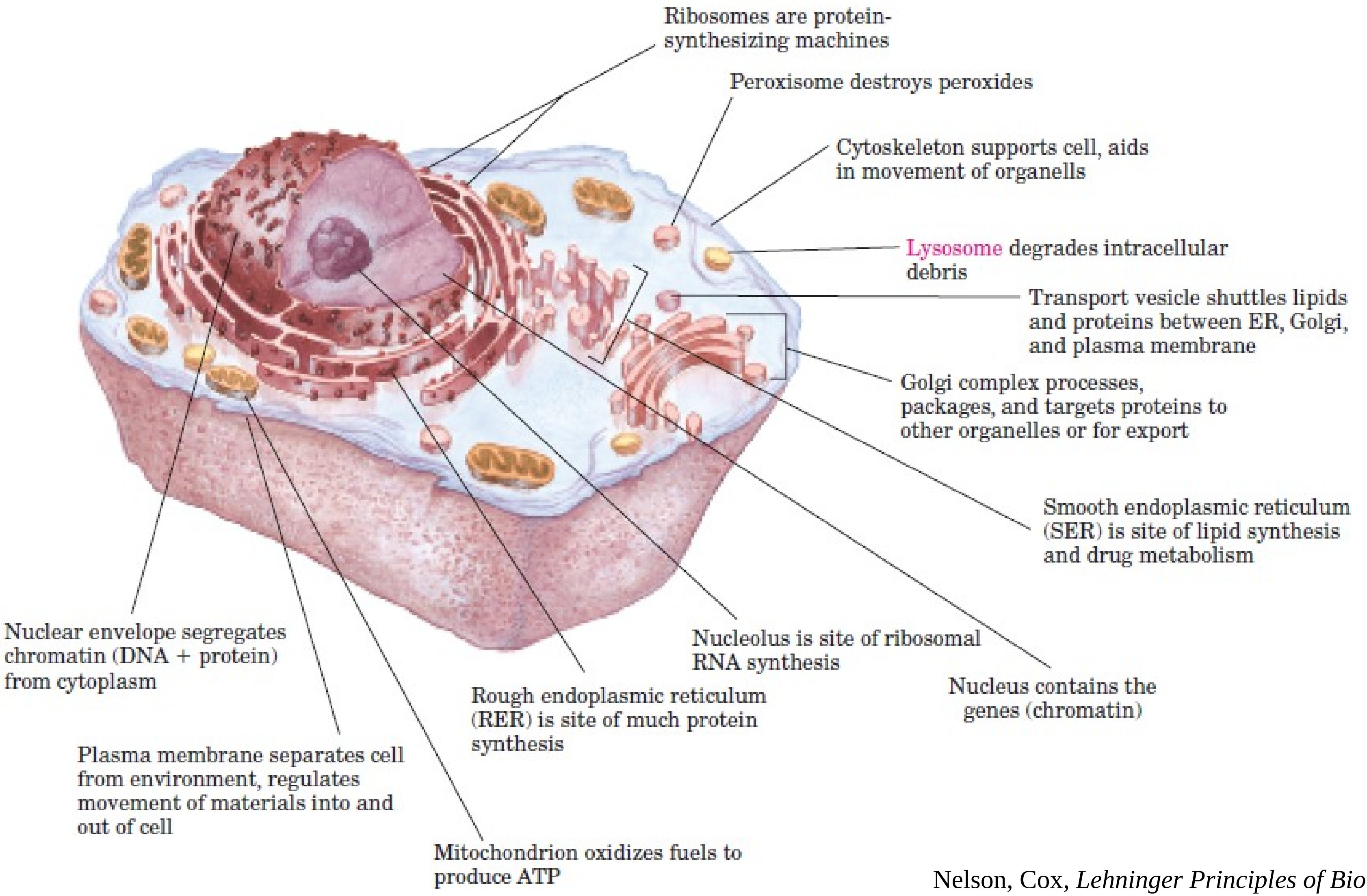
	Prokaryotes	*Eukaryotes*
Size	Mostly about 1–10 μm in length.	Mostly about 10–100 μm in length.
Nucleus	Nucleus is missing. The chromosomal region is called nucleolus.	Nucleus is separated from the cytoplasm by the nuclear envelope.
Intracellular organization	Normally, no membrane-separated compartments and no supportive intracellular skeletal framework are present in the cells' interior.	Distinct compartments are present, e. g., nucleus, cytosol with a cytoskeleton, mitochondria, ER, Golgi complex, lysosomes, plastids (chloroplasts, leucoplasts).
Gene structure	No introns; some polycistronic genes.	Introns and exons.
Cell division	Simple cell division.	Mitosis or meiosis.
Ribosome	Consists of a large 50S subunit and a small 30S subunit.	Consists of a large 60S subunit and a small 40S subunit.
Reproduction	Parasexual recombination.	Sexual recombination.
Organization	Mostly single cellular.	Mostly multicellular, and with cell differentiation.

Model Organisms



Organism	*Number of chromosomes (haploid genome)*	*Genome size (base pairs; genes)*
Mycoplasma genitalium (prokaryote)	1 circular chromosome	$580 \cdot 10^3$ bp; 480 genes
Escherichia coli (prokaryote)	1 circular chromosome	$4.6 \cdot 10^6$ bp; 4,290 genes
Saccharomyces cerevisiae (budding yeast; eukaryote)	16 chromosomes	$12.5 \cdot 10^6$ bp; 6,186 genes
Arabidopsis thaliana (flowering plant; eukaryote)	5 chromosomes	$100 \cdot 10^6$ bp; ~25,000 genes
Drosophila melanogaster (fruit fly, eukaryote)	4 chromosomes	$180 \cdot 10^6$ bp; ~14,000 genes
Mus musculus (mouse, eukaryote)	20 chromosomes	$2.5 \cdot 10^9$ bp; ~30,000 genes
Homo sapiens (human, eukaryote)	23 chromosomes	$2.9 \cdot 10^9$; ~30,000 genes

Structure of the (Animal) Cell



Ribosomes are protein-synthesizing machines

Peroxisome destroys peroxides

Cytoskeleton supports cell, aids in movement of organells

Lysosome degrades intracellular debris

Transport vesicle shuttles lipids and proteins between ER, Golgi, and plasma membrane

Golgi complex processes, packages, and targets proteins to other organelles or for export

Smooth endoplasmic reticulum (SER) is site of lipid synthesis and drug metabolism

Nuclear envelope segregates chromatin (DNA + protein) from cytoplasm

Nucleolus is site of ribosomal RNA synthesis

Nucleus contains the genes (chromatin)

Rough endoplasmic reticulum (RER) is site of much protein synthesis

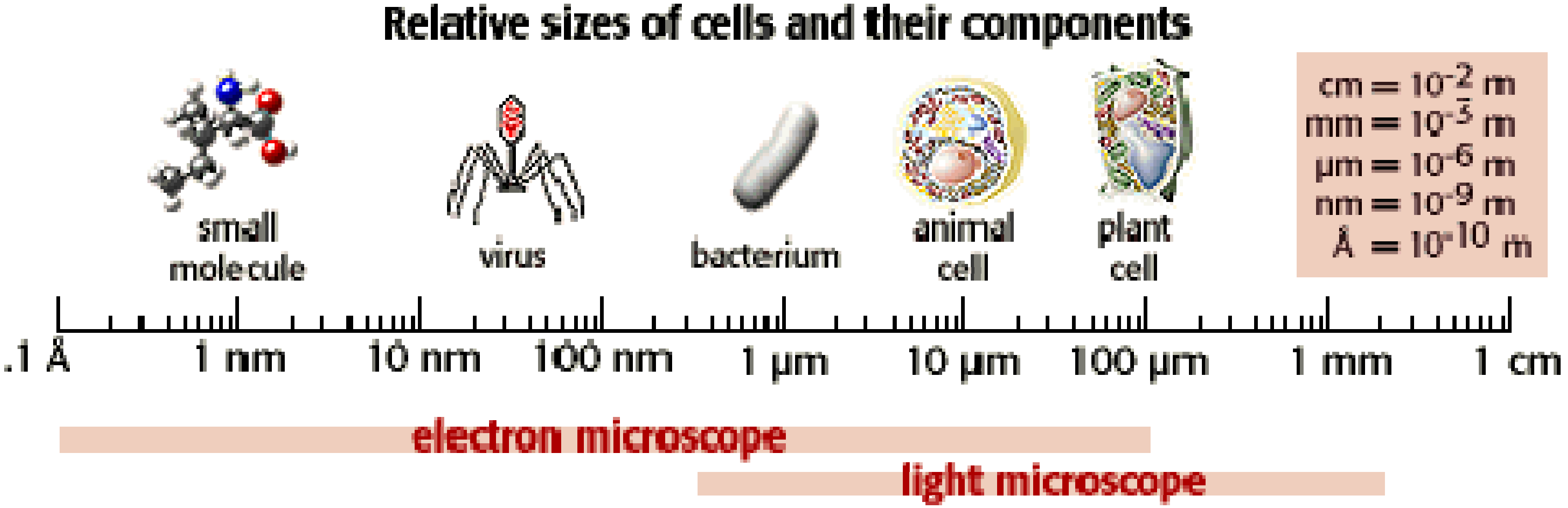
Plasma membrane separates cell from environment, regulates movement of materials into and out of cell

Mitochondrion oxidizes fuels to produce ATP

Nelson, Cox, *Lehninger Principles of Biochemistry*

Composition and Sizes of Biological Objects

- Biological compounds are prevalently composed of four basic components: C, N, O, H
- Many compounds are made out of a set of few repeated basic blocks (nucleic acids, polypeptides, polymers)

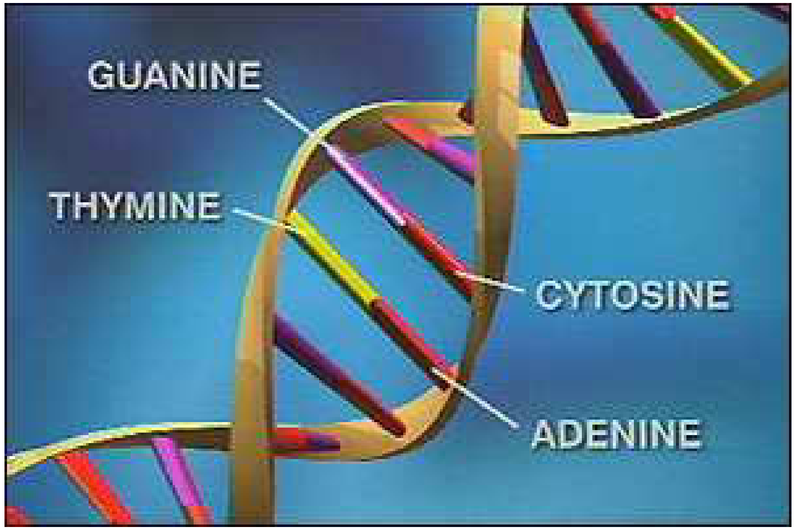


Biomolecules Classifications

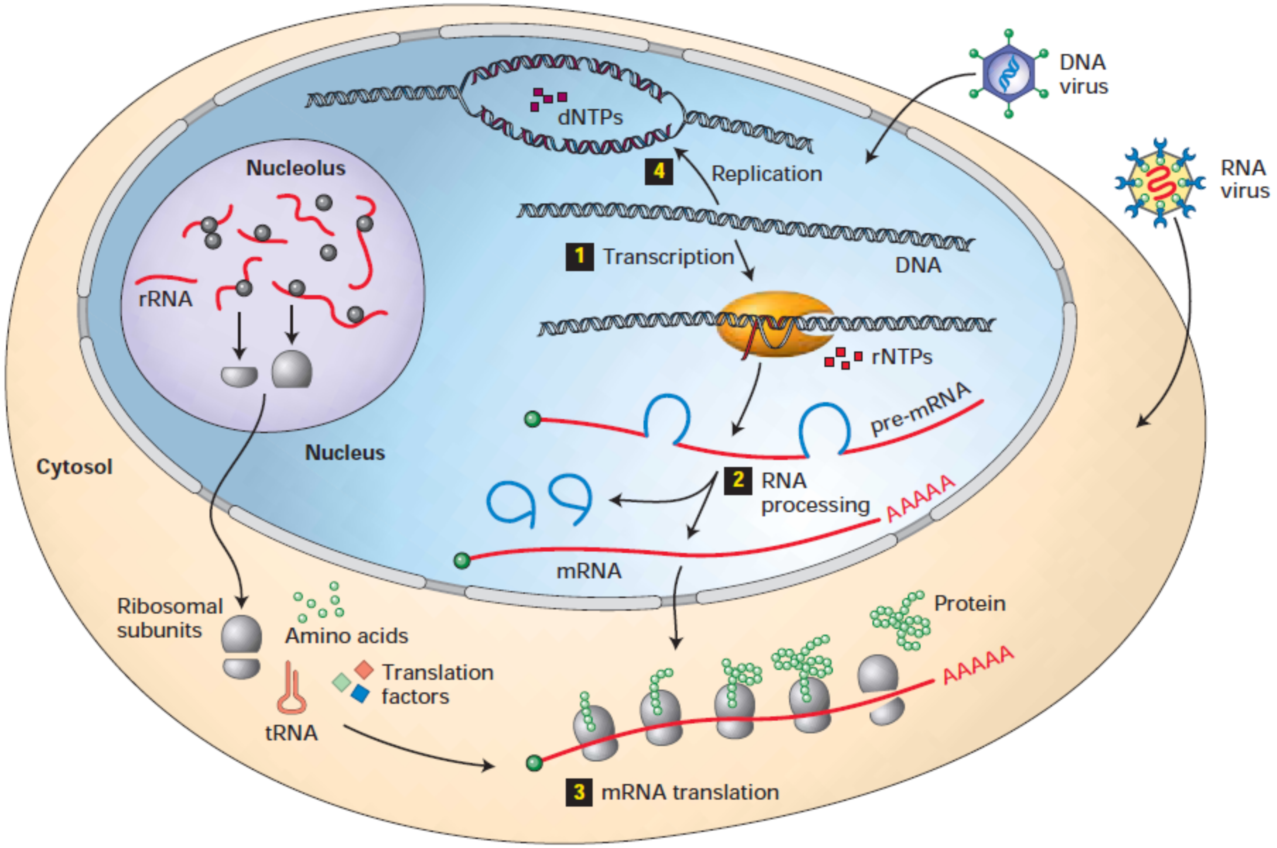
- Biomolecules can be divided into four families
  - Carbohydrates (energy storage)
  - Lipids (structure, signaling)
  - Proteins (catalysis, regulation, signaling, structure and support)
  - Nucleic acids (Information coding, regulation)
- Each family has multiple functions (e.g. RNAs are coding polymers, but can also have regulatory functions, whereas some lipids act as hormones, i.e. signals)
- Most molecules are made up of sub-complexes belonging to different families

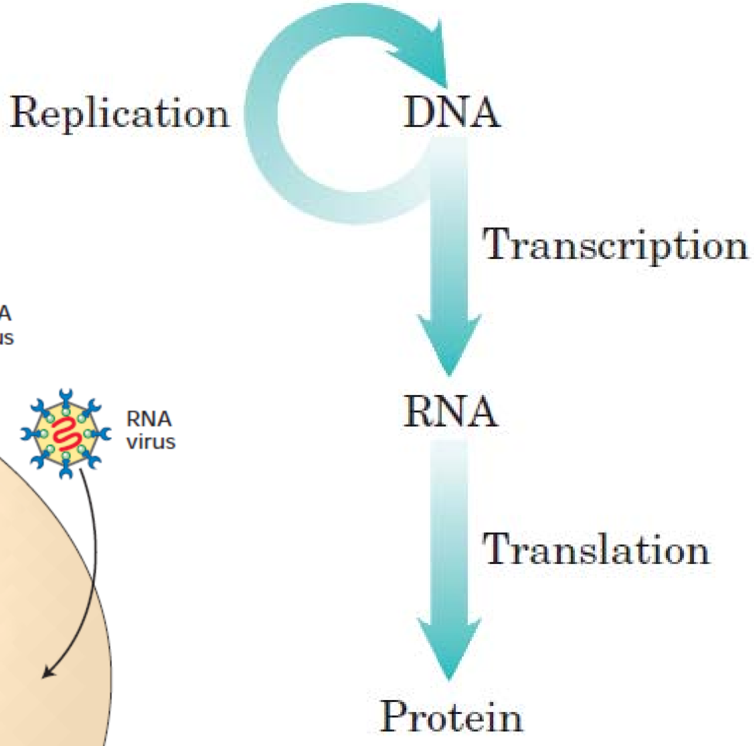
Nucleic Acids

- Nucleic acids are polymers built up of covalently bound mononucleotides, they comprise DNA and RNA
- DNA contains coding sequences (genes) for proteins and RNAs and non-coding sequences
- Nucleotides are also important in energy transfer in a multitude of reactions (e.g. ATP – ADP)
- Some nucleotides also serve as signaling packets (e.g. cAMP)



Central Dogma of Molecular Biology

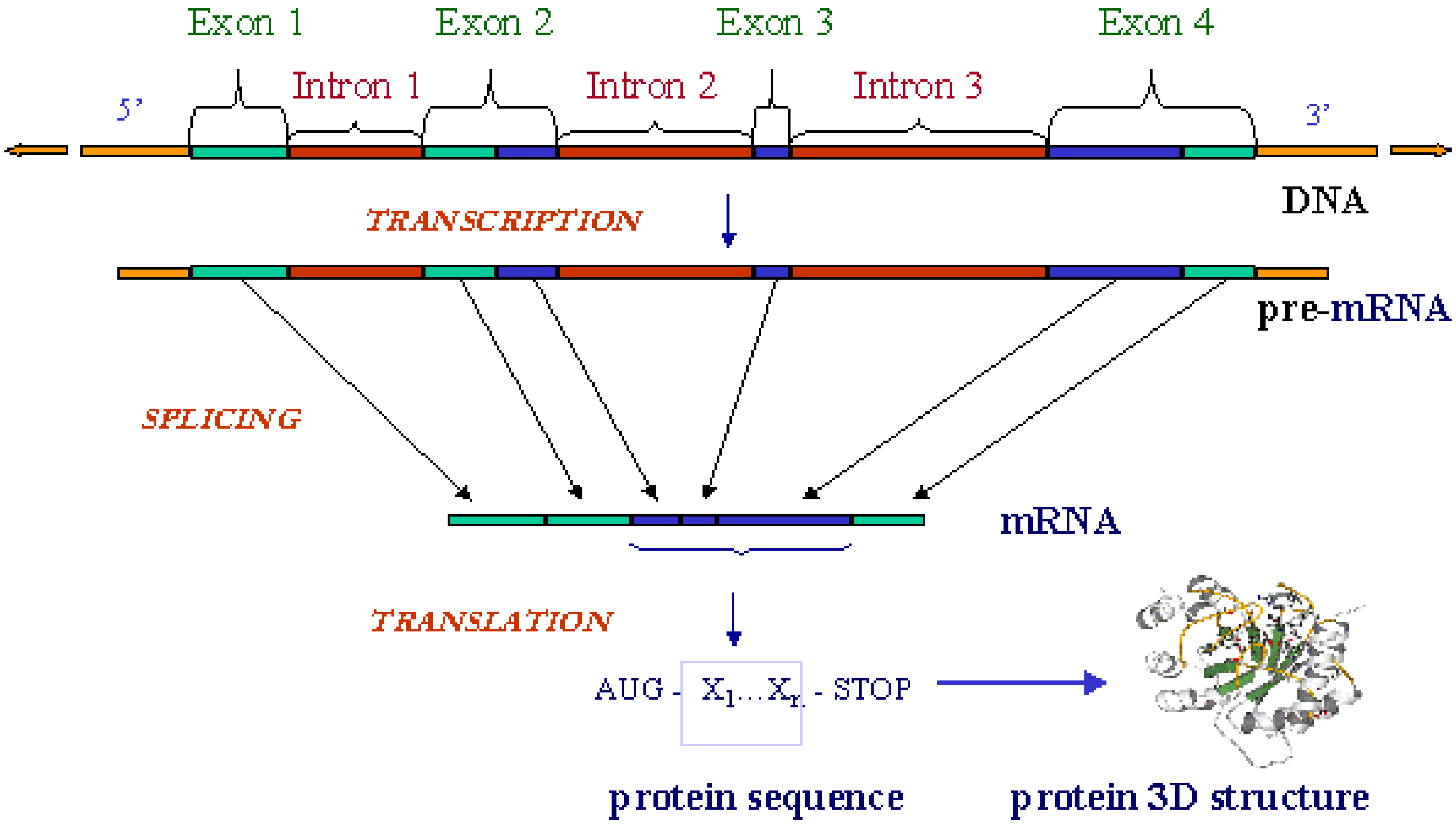



Replication
DNA
Transcription
RNA
Translation
Protein

Lodish et al, *Molecular Cell Biology*

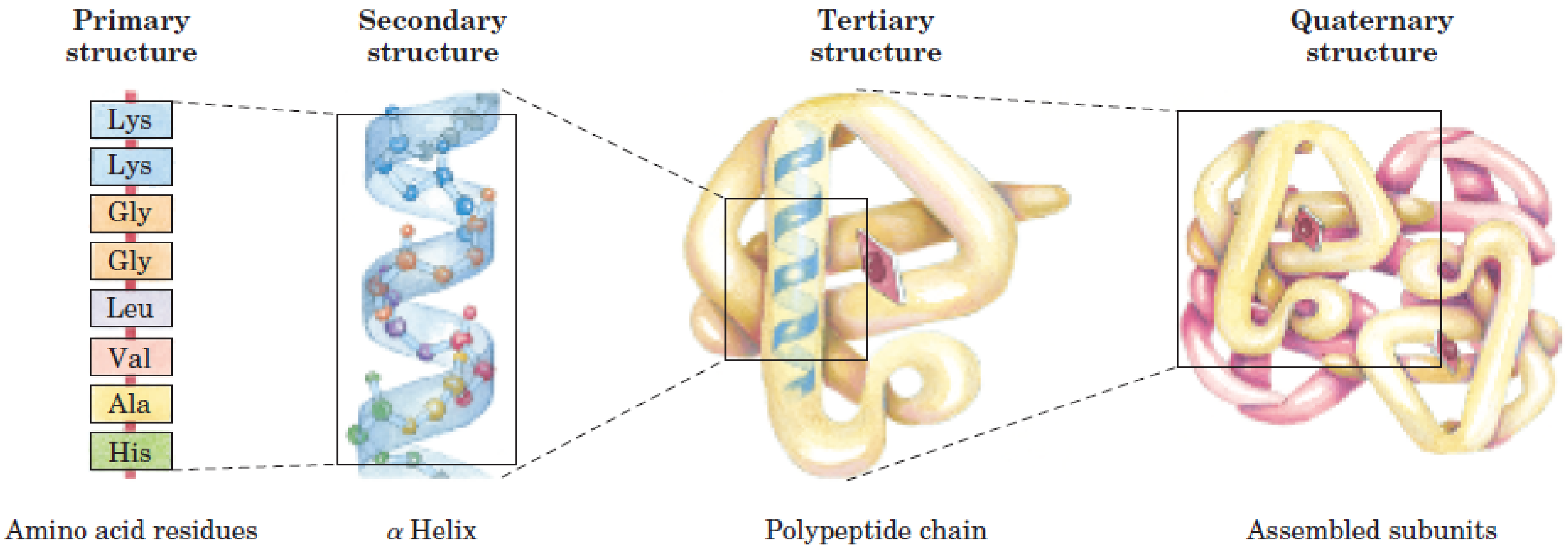
Transcription and Translation

- Through alternative splicing, the same coding region can produce more than one protein



Structure of Proteins

- Proteins consist of one or more polypeptides
- Each peptide is composed of covalently linked amino acids, chosen from a set of 20 different ones



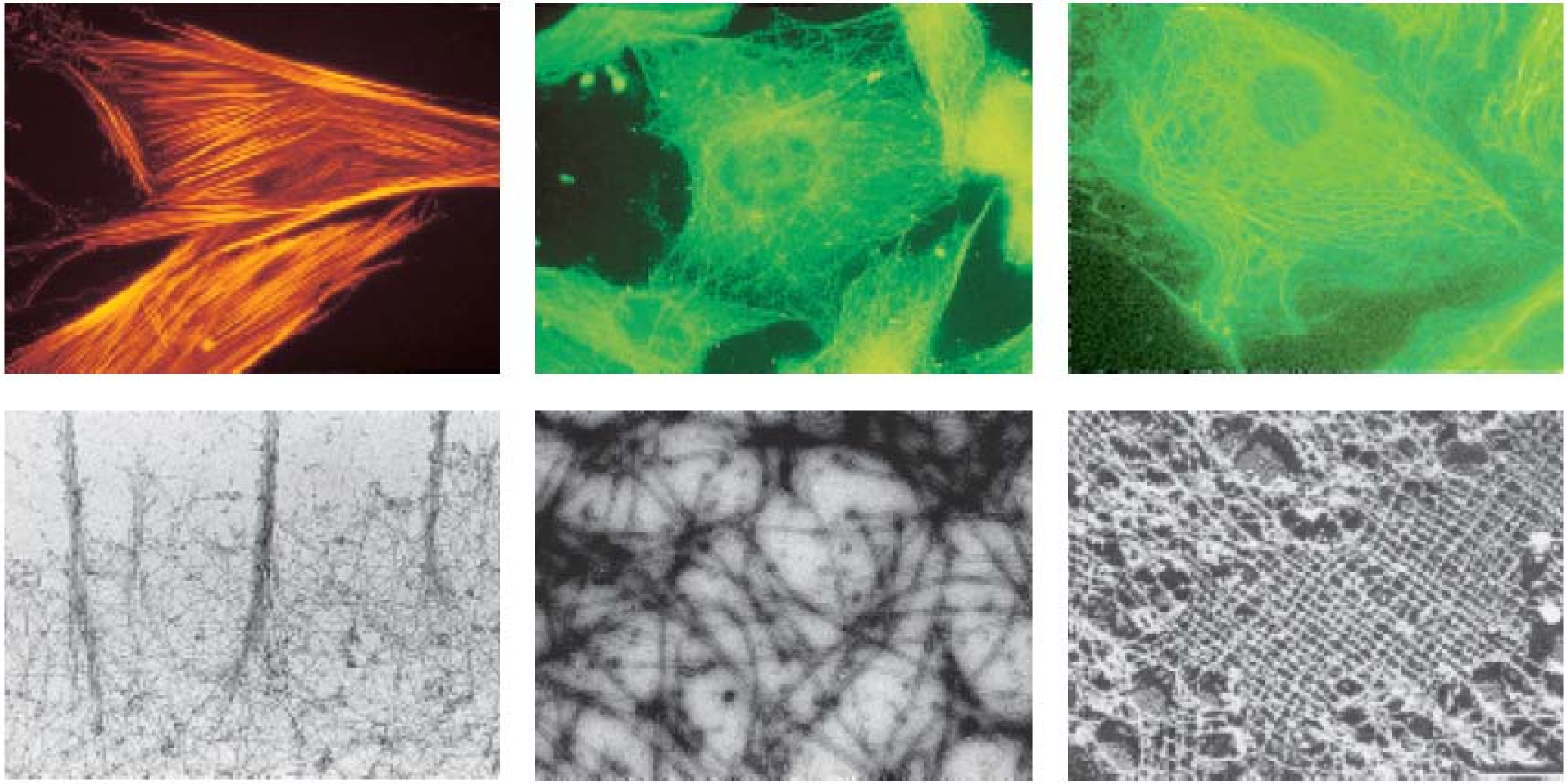
Nelson, Cox, *Lehninger Principles of Biochemistry*

Structure of Proteins (cont'd)

- The functioning of a protein is strictly dependent on its three-dimensional structure
- Chemical modifications (e.g. phosphorylation) can induce relevant conformational changes, exposing or hiding the active sites
- On the other hand, environmental changes (e.g. pH, temperature) or endogenous alterations (e.g. genetic mutations) usually modify the protein structure, hindering its correct functioning

The Cytoskeleton

- Translocation of compounds is an important mechanism, though it cannot be described by standard diffusion processes



Actin stress fibers
(a)

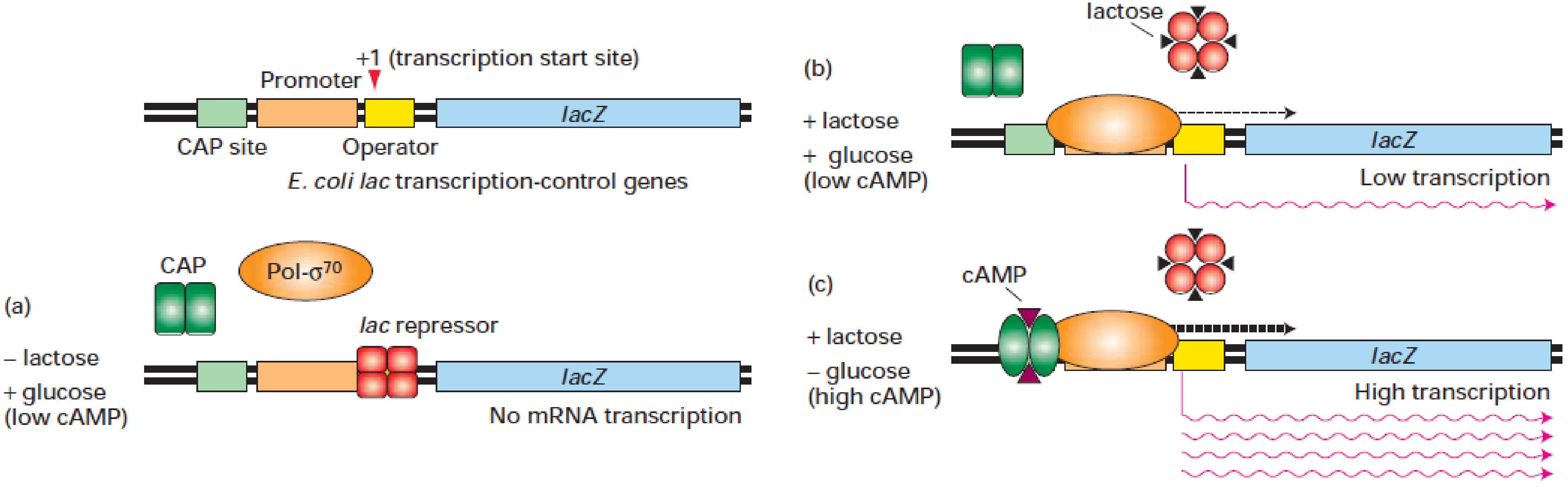
Microtubules
(b)

Intermediate filaments
(c)

Nelson, Cox, *Lehninger Principles of Biochemistry*

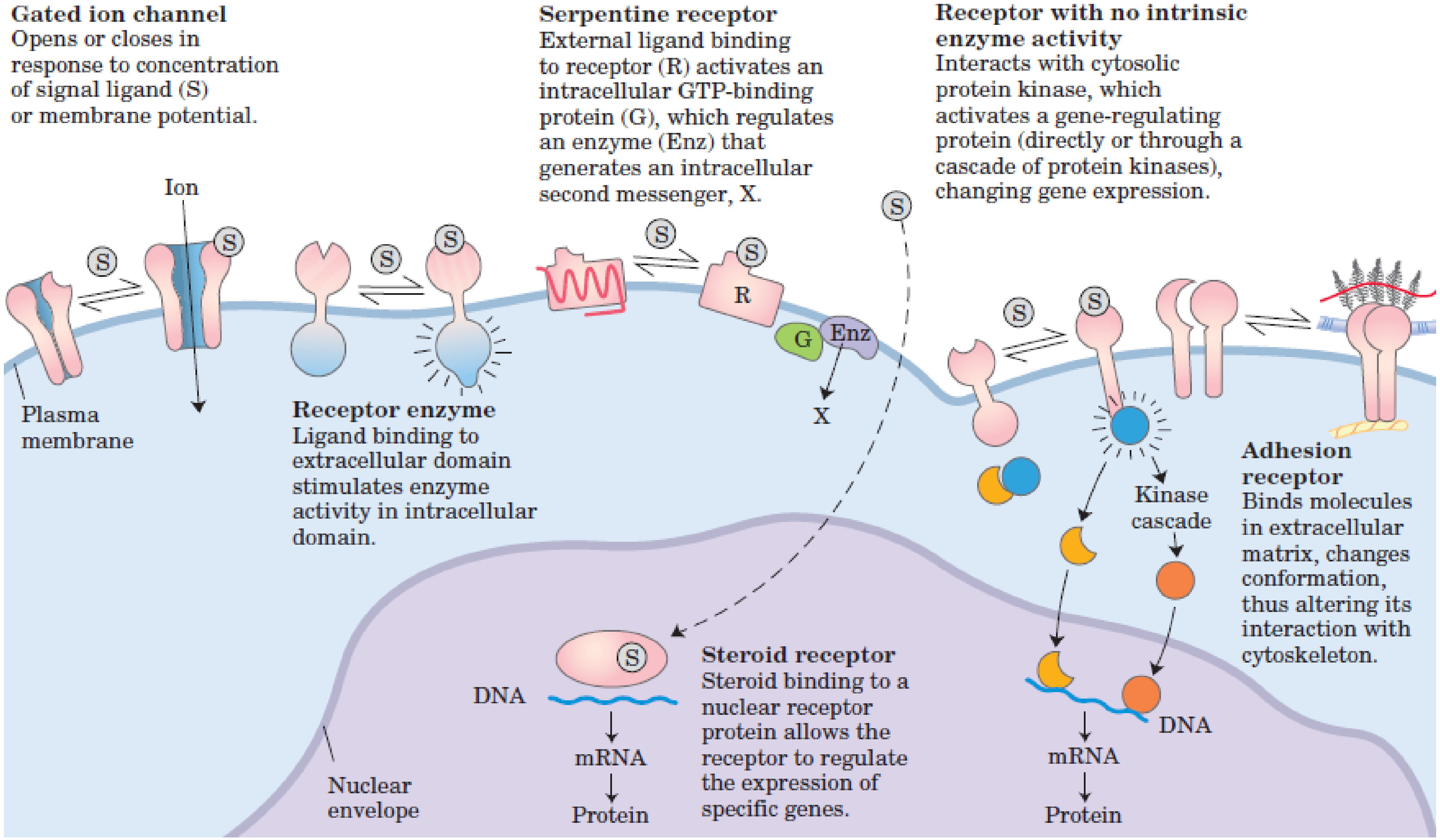
Combinatorial Regulation

- Genes expression is regulated via adjacent transcription-control regions in a combinatorial way
- Only a subset of the whole genome is expressed at a particular time or in a specific cell type



Lodish et al, *Molecular Cell Biology*

Biosignaling: Membrane Receptors



Gated ion channel
Opens or closes in response to concentration of signal ligand (S) or membrane potential.

Serpentine receptor
External ligand binding to receptor (R) activates an intracellular GTP-binding protein (G), which regulates an enzyme (Enz) that generates an intracellular second messenger, X.

Receptor with no intrinsic enzyme activity
Interacts with cytosolic protein kinase, which activates a gene-regulating protein (directly or through a cascade of protein kinases), changing gene expression.

Receptor enzyme
Ligand binding to extracellular domain stimulates enzyme activity in intracellular domain.

Adhesion receptor
Binds molecules in extracellular matrix, changes conformation, thus altering its interaction with cytoskeleton.

Steroid receptor
Steroid binding to a nuclear receptor protein allows the receptor to regulate the expression of specific genes.

Nelson, Cox, *Lehninger Principles of Biochemistry*

UMG
Dubium sapientiae initium

Outline

- Basic Principles
- Biology in a nutshell
- Experimental Techniques

Getting Experimental Data

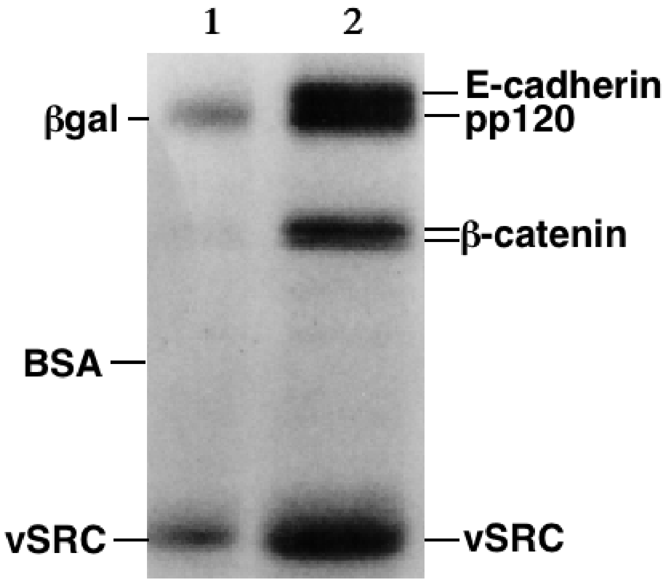
- Systems scientists are typically not concerned about the difficult and often indirect processes by which biological relationships and interactions are identified
- It is worth having a glimpse at the basic techniques used in experimental biology
- Two main issues arise when using experimental techniques:
  - How to get quantitative data out of experiments designed to give qualitative answers
  - How to get time-sampled (synchronized) measurements of many biological objects (possibly cheaply and in a suitable time)

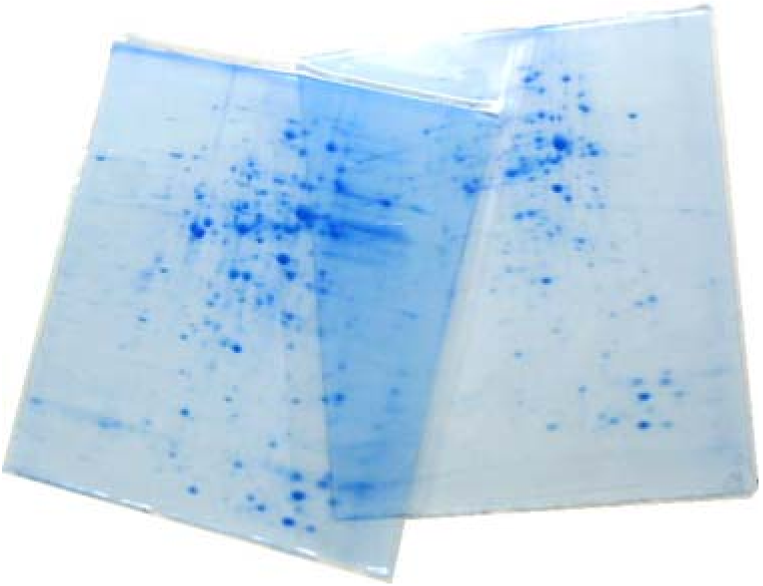
Gel Electrophoresis

- Both nucleic acids and proteins are electrically charge, thus they can be separated by applying electrical fields
- While the negative charge of DNA is proportional to the length of the fragment under consideration, for proteins the charge distribution is much more involved and case-specific
- By using sodium dodecyl sulfate (SDS) it is possible to confer to a protein a negative charge that is proportional to the number of its amino acids

Gel Electrophoresis

- SDS-PAGE does not provide absolute quantitative data, rather it can be used for relative measurements
- The samples have to be purified before measurement (at least in one-dimensional GE)



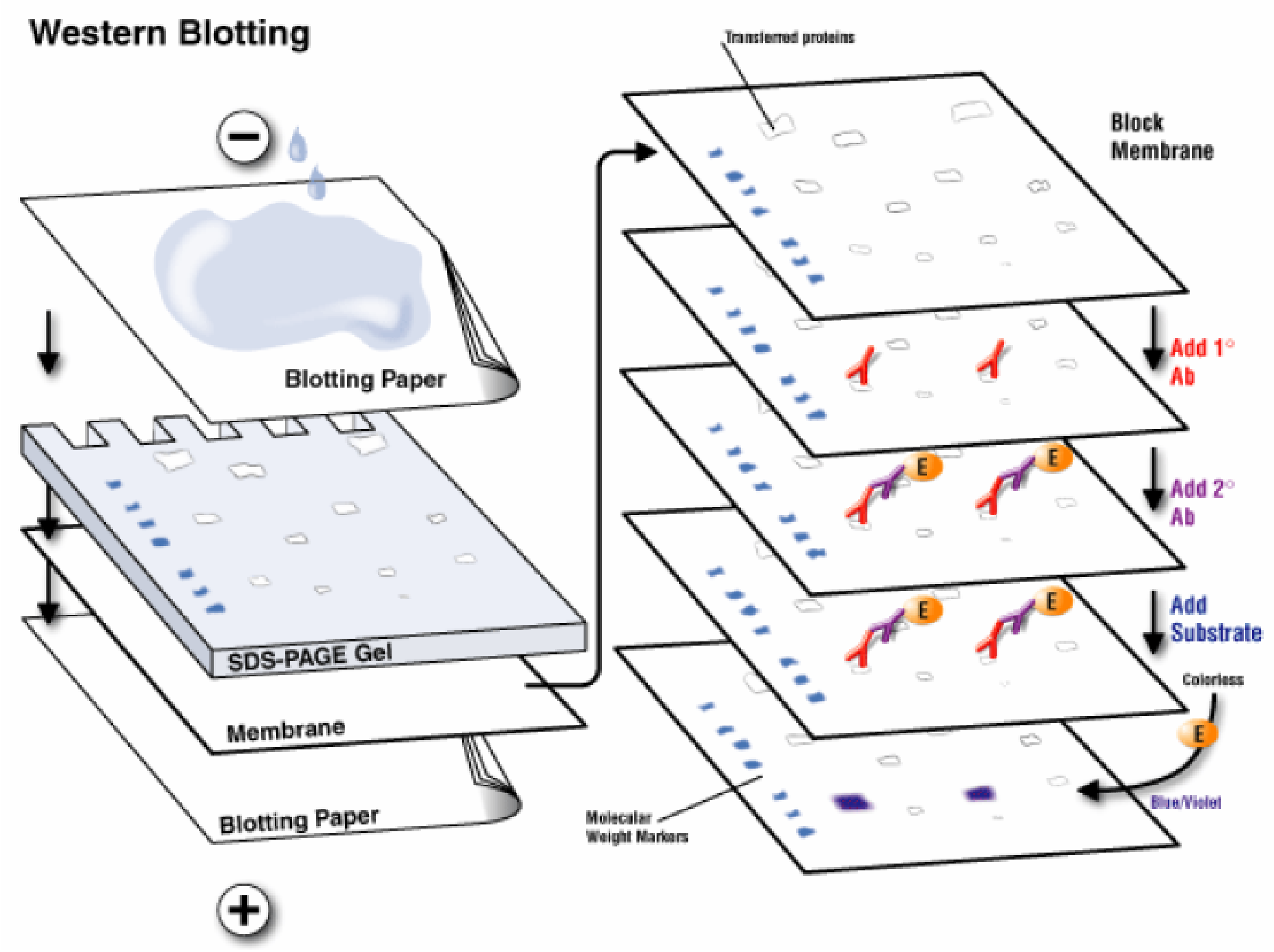


Hybridization and Blotting

- Hybridization techniques are based on the specific recognition of a probe and target molecule
- Used to selectively visualize only the molecules of interest in a complex mixture
- It can be applied both to DNA (Southern Blotting) and to mRNA samples (Northern Blotting)
- In the case of protein it is not possible to exploit hybridization, hence it is necessary to use a specific antibody for the protein of interest

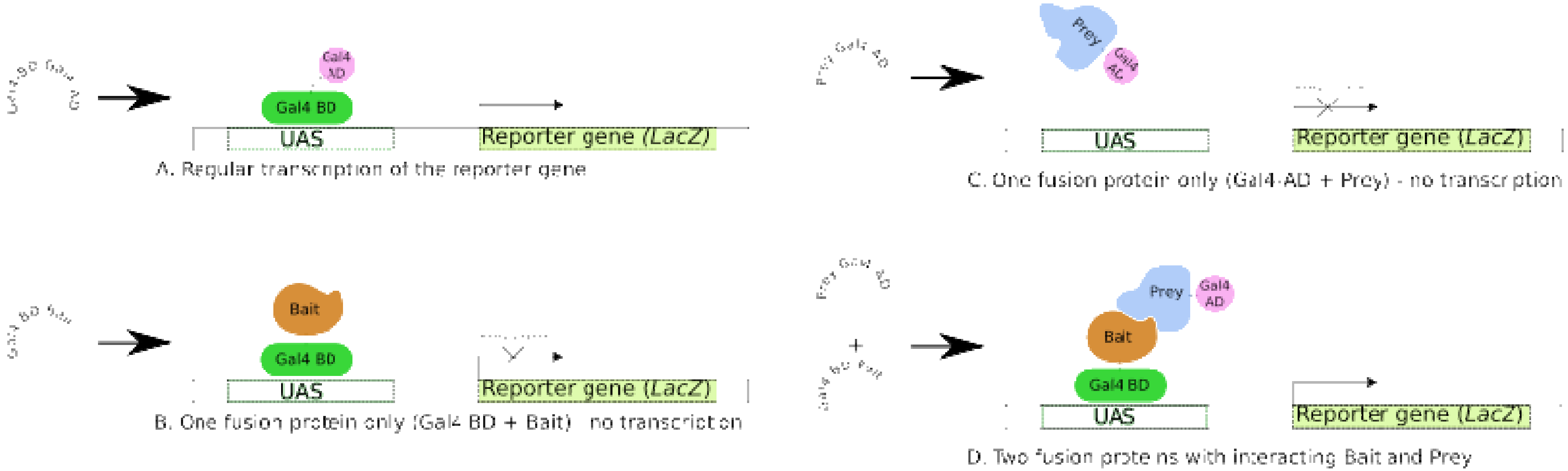
Western Blotting (Immunoblotting)

- Measurements of intracellular protein concentration



Yeast Two-Hybrid (Y2H)

- Two-hybrid screening is used to determine protein – protein and protein – DNA interactions
- The protein to be tested are hybridized with two fragments of a transcription factor
- The TF interacts with the upstream activation site (UAS), activating the reported gene



A. Regular transcription of the reporter gene

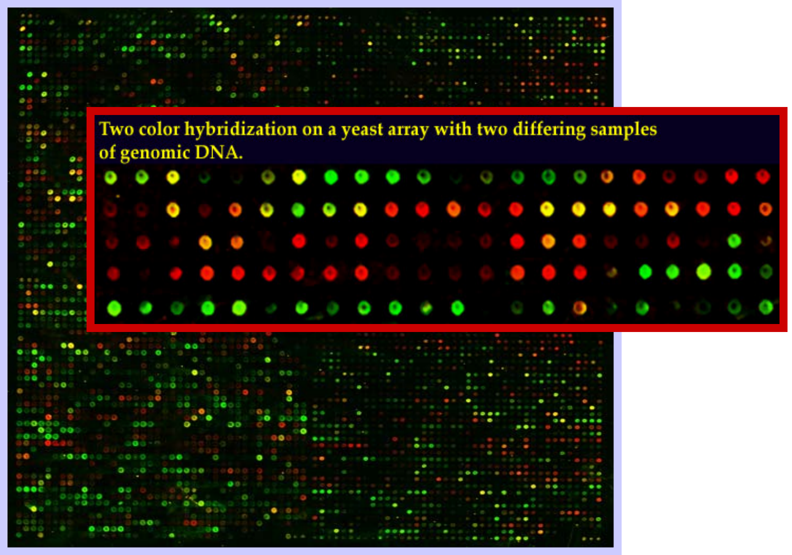
C. One fusion protein only (Gal4-AD + Prey) - no transcription

B. One fusion protein only (Gal4 BD + Bait) - no transcription

D. Two fusion proteins with interacting Bait and Prey

DNA Microarrays

- DNA chips, also called DNA microarrays, are a recently developed method for the high-throughput analysis of gene expression
- Instead of looking at the expression of a single gene, microarrays allow one to monitor the expression of several thousand genes in a single experiment, resulting in a global picture of cellular activity
- Hence they represent a key tool for implementing a systems biology approach



Two color hybridization on a yeast array with two differing samples of genomic DNA.

Single Feature

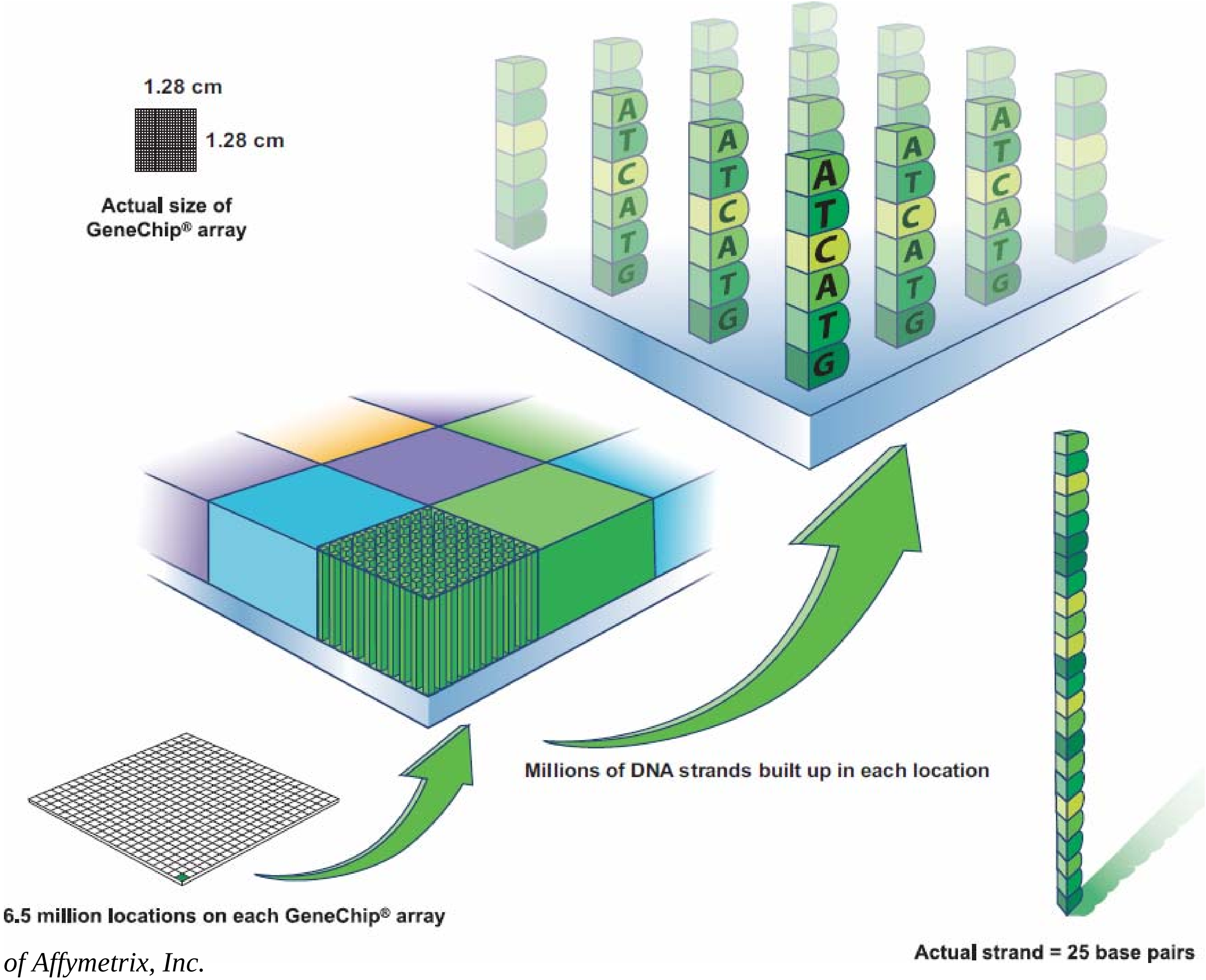


Image courtesy of Affymetrix, Inc.

Probes Hybridization

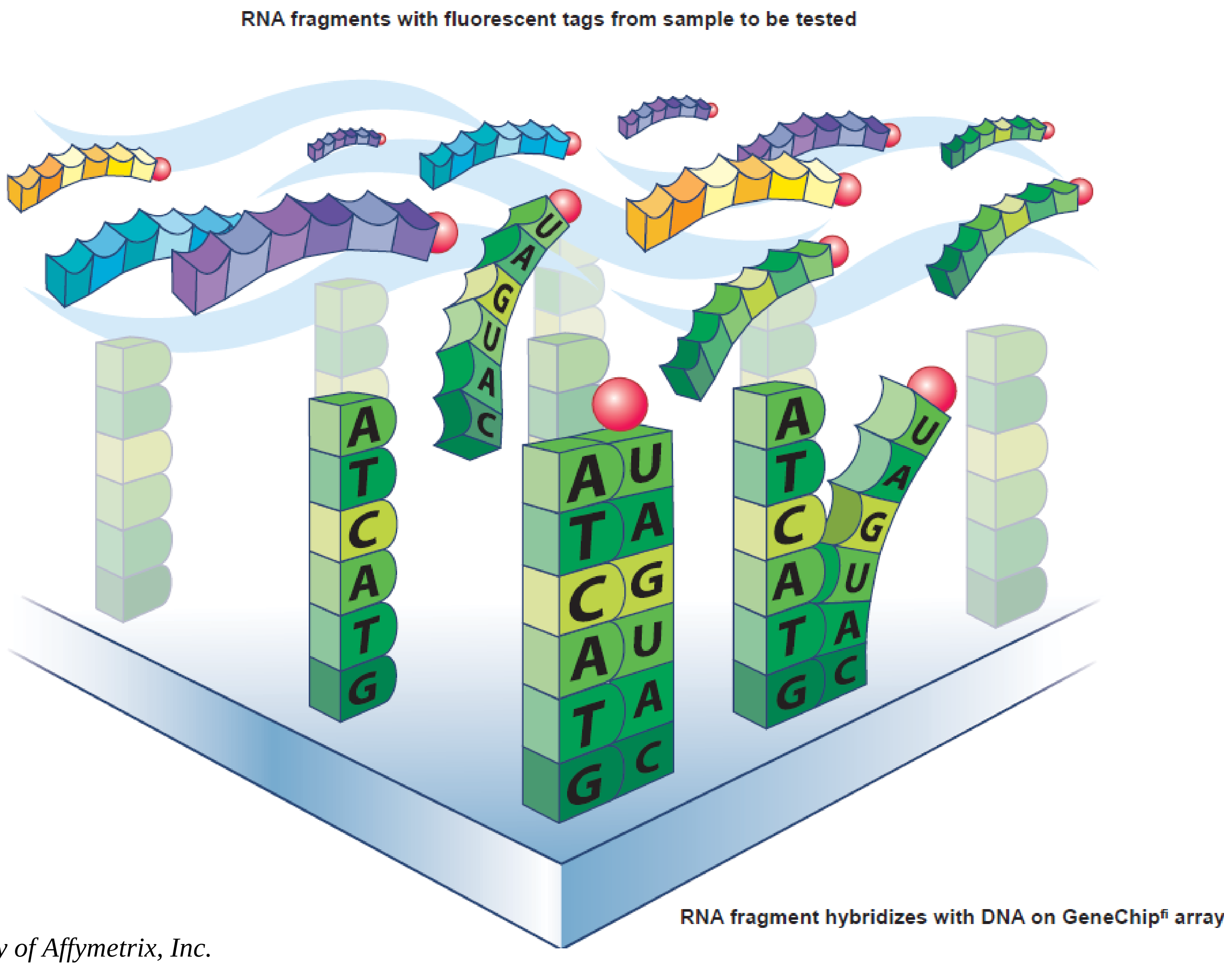


Image courtesy of Affymetrix, Inc.

Hybridized DNA Visualization

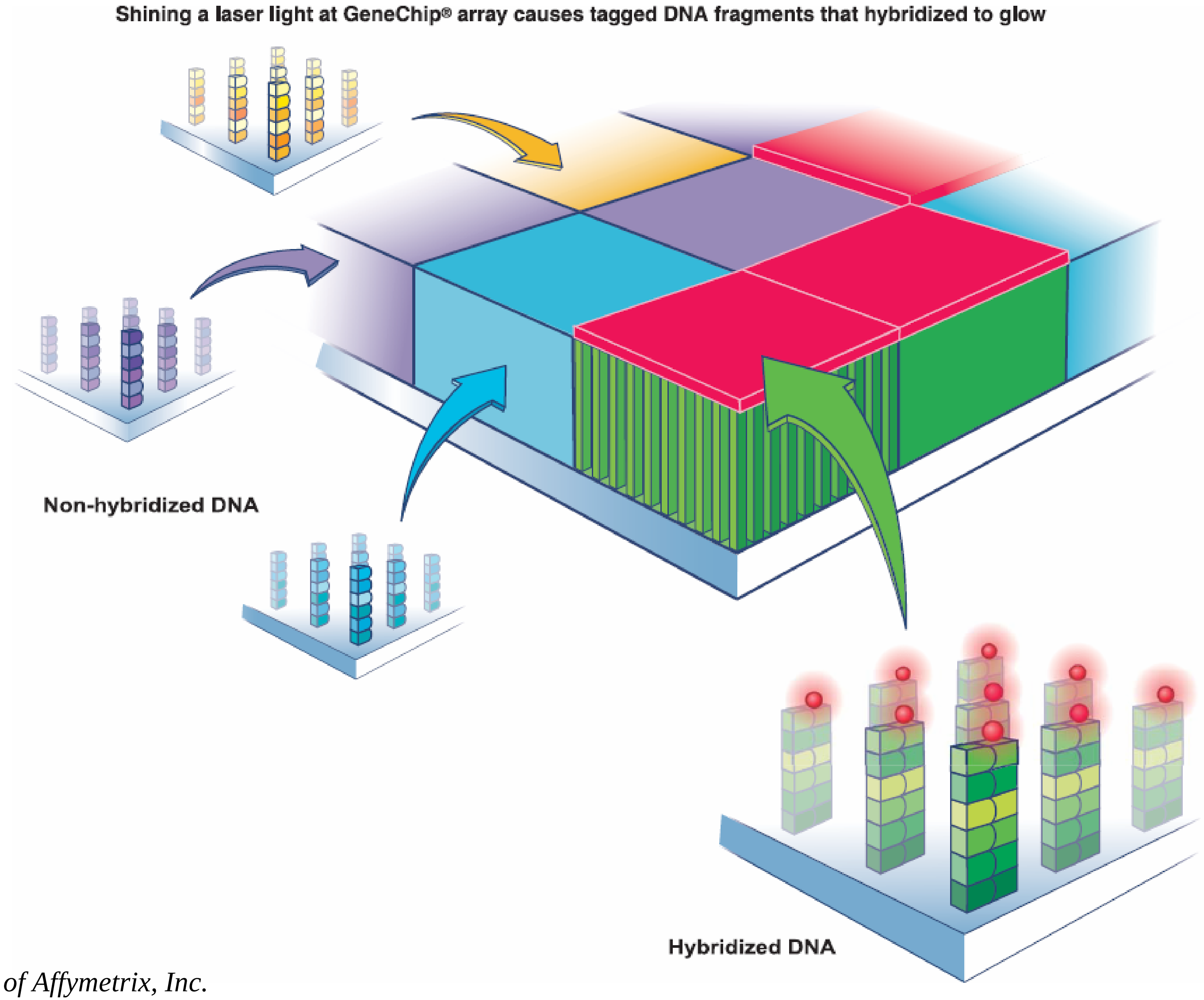
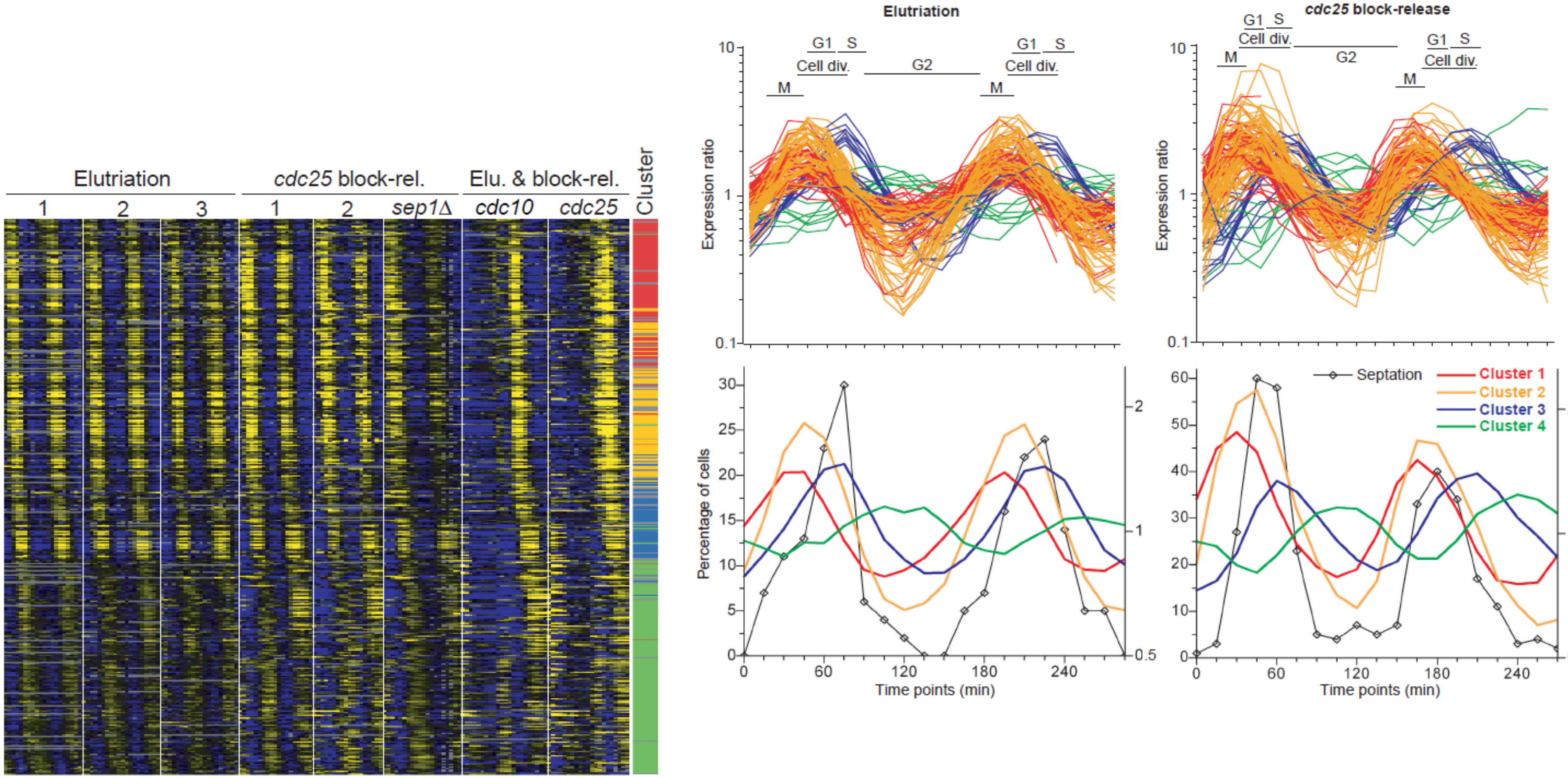


Image courtesy of Affymetrix, Inc.

Time-Course Experimental Data

- Microarrays enable to derive time-course experimental data, with a desired sampling time



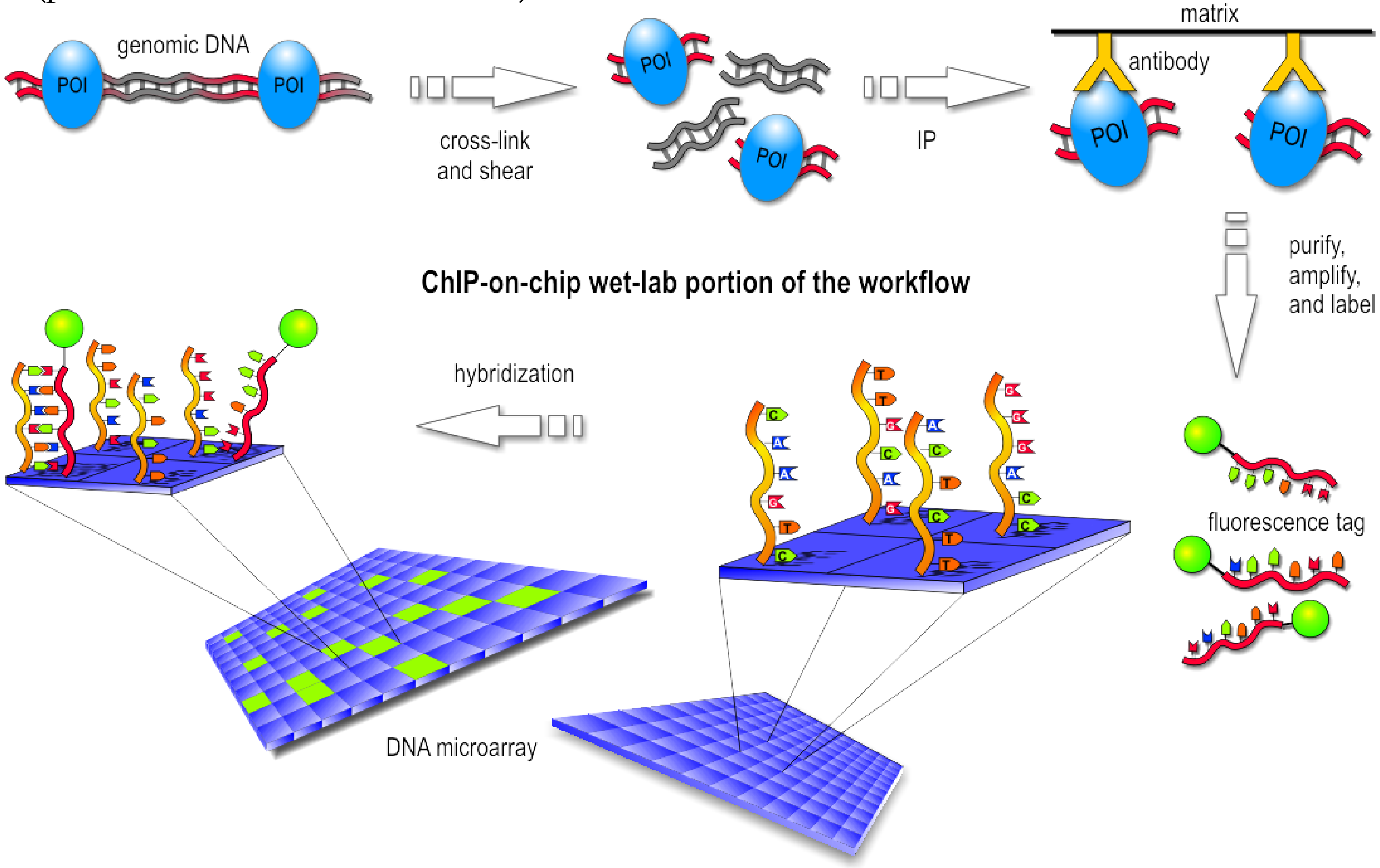
Rustici et al, Nature Genetics 36(8), 2004

Protein Microarrays

- The function of a gene is realized by the coded protein, not by its mRNA
- This fact addresses a main flaw in the DNA microarray approach
- Solution: protein microarrays! Unfortunately…
  - Proteins are not as uniform as DNA
  - It is not (yet) possible to generate the amount of recombinant protein needed for high-throughput experiments
  - Optimal interaction conditions (in terms of temperature, ionic strenght, pH) are largely varying among proteins
- Nonetheless, research on protein chips is rapidly progressing

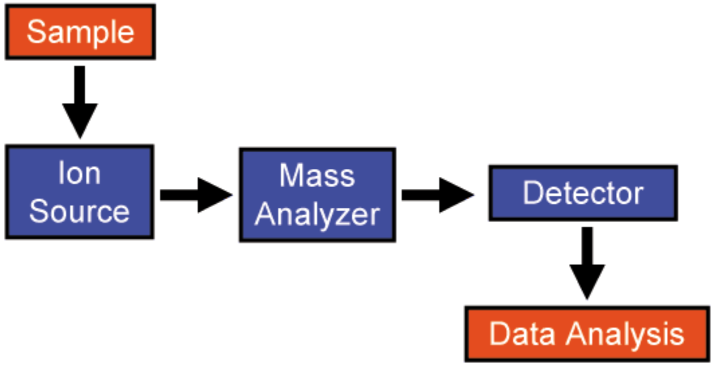
ChIP on Chip

- Chromatin Immuno–Precipitation is used to discover transcription factors (protein-DNA interactions)



Mass Spectrometry

- Mass spectrometry is used to identify unknown compounds and quantify known compounds in a solution
  - A small sample of compounds is ionized, usually to cations by loss of an electron
  - The ions are sorted and separated according to their mass and charge
  - The separated ions are then detected and tallied, and the results are displayed on a chart



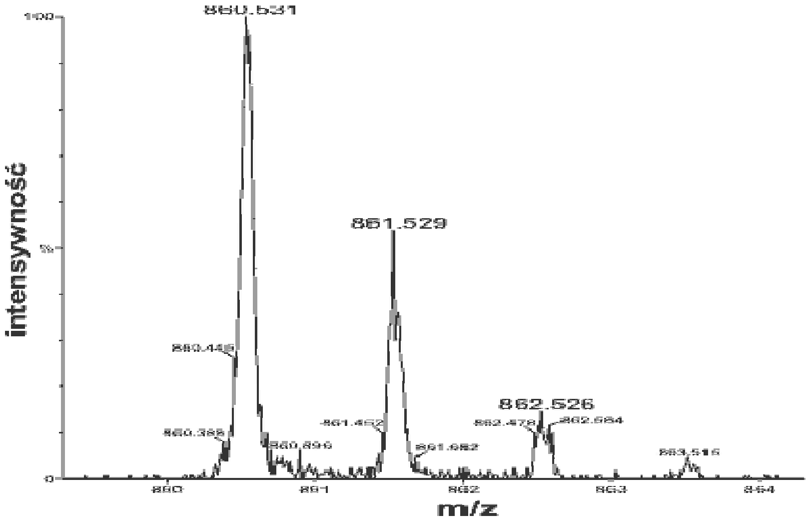
Sample

Ion Source

Mass Analyzer

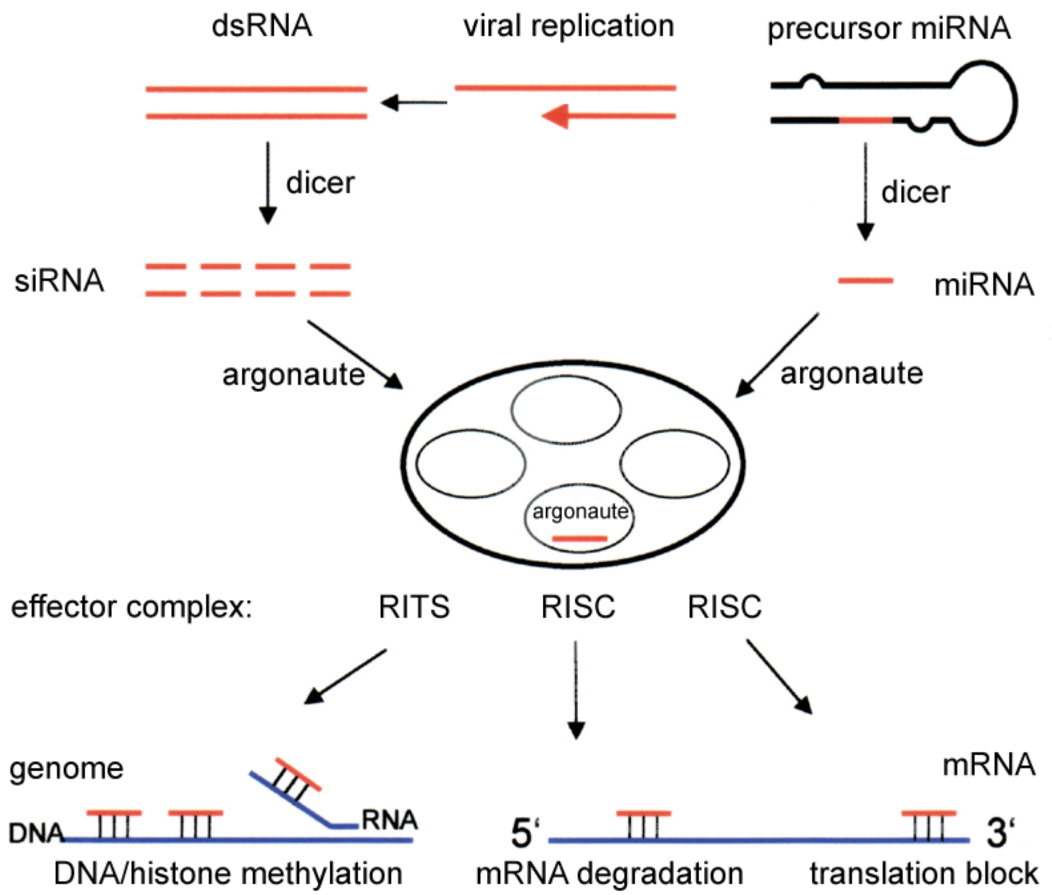
Detector

Data Analysis



RNA Interference (RNAi)

- Used for transiently down – regulating arbitrary genes



References

- Klipp et al, Systems Biology in Practice, Wiley-VCH, 2005
- DL Nelson, MM Cox, *Lehninger Principles of Biochemistry*, WH Freeman, 2004
- Lodish et al, *Molecular Cell Biology*, WH Freeman, 2007