Advanced Algorithms and Models for Computational Biology

Class Overview

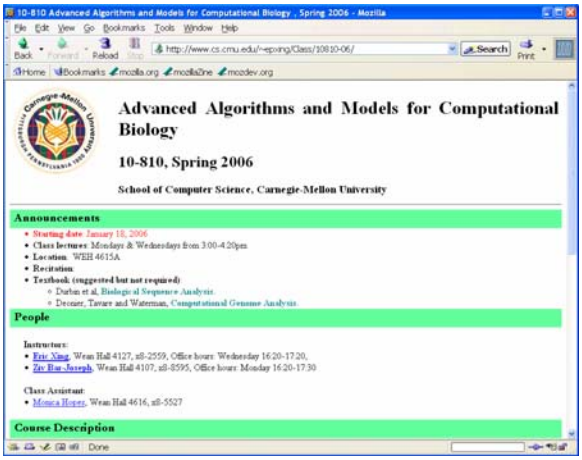
Eric Xing & Ziv Bar-Joseph

Lecture 1, January 18, 2005

Reading: Chap. 1, DTM book

Logistics

- Class webpage:
  - http://www.cs.cmu.edu/~epxing/Class/10810-06/



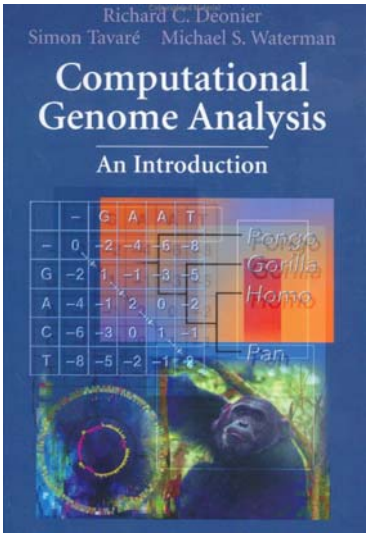
Logistics

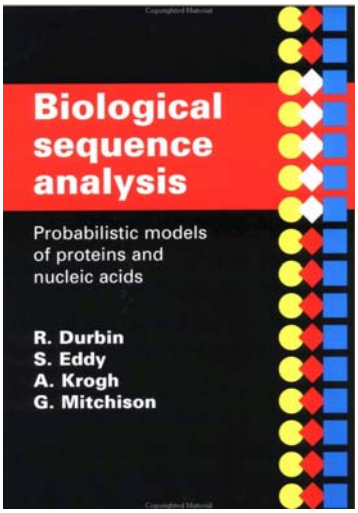
- 4 homework: 40% of grade
  - Theory exercises
  - Implementation exercises
- Final project: 40% of grade
  - Applying what you learned in the class to a realistic, non-trivial CompBio problem
    - Sequence analysis, network modeling, microarray mining, genetic polymorphism, evolution …
    - If on your own current research, higher expectation, talk to us first …
  - Expected outcome
    - A deliverable software based on an new/extended model or algorithm
    - Addressing a biological problem via a thorough analysis of a realistic dataset
    - A mini-paper …
  - Collaboration policies …
    - Tow persons per team at most
- Class participation and reading: 20% of grade

Logistics

- No required text books, but suggested reading will be announced prior every class from:
  - Durbin et al, **Biological Sequence Analysis.**
  - Deonier, Tavare and Waterman, **Computational Genome Analysis.**
  - Selected papers
- Mailing Lists:
  - Send email to eric.xing@gmail.com with: full name, department, register/audit
  - To contact the instructors: 10810-instr@cs.cmu.edu
  - Class announcements list: 10810-announce@cs.cmu.edu
- Class Assistant:
  - Monica Hopes, Wean Hall 4616, x8-5527

Books

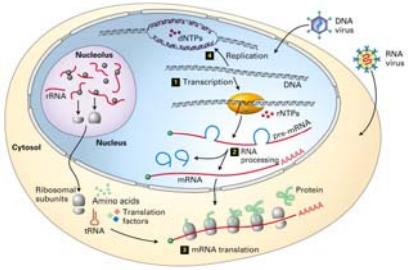




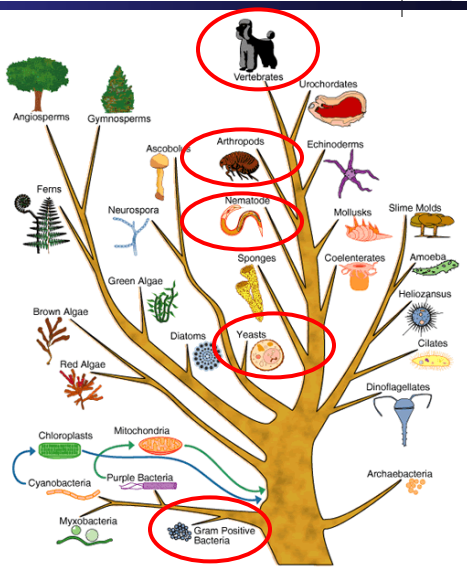
Class Plan

- Introduction (1week)
- Biological Sequence Analysis (2.5 weeks)
- Gene Expression Analysis (3 weeks)
- Population Genetics (2 weeks)
- Evolution and Phylogeny (2 weeks)
- Systems Biology (4 weeks)

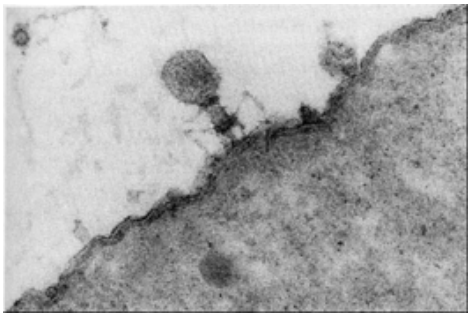
Introduction to cell biology, functional genomics, development, etc.

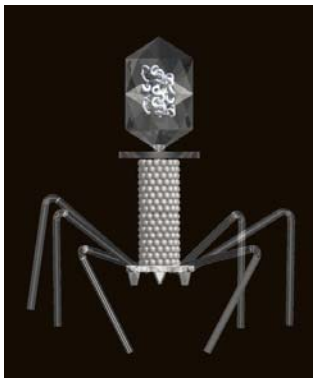


Model Organisms

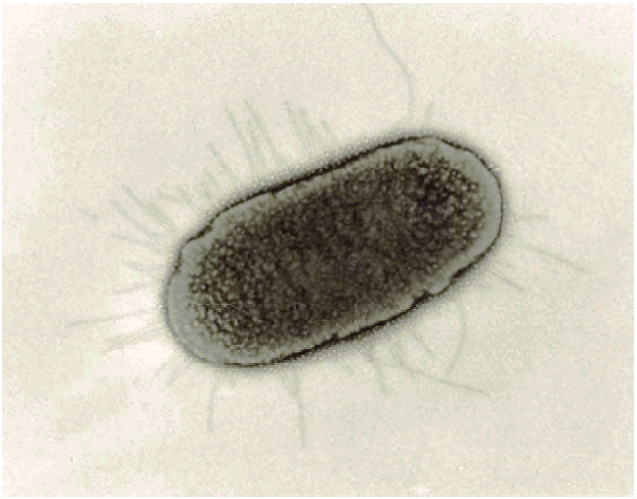


Bacterial Phage: T4

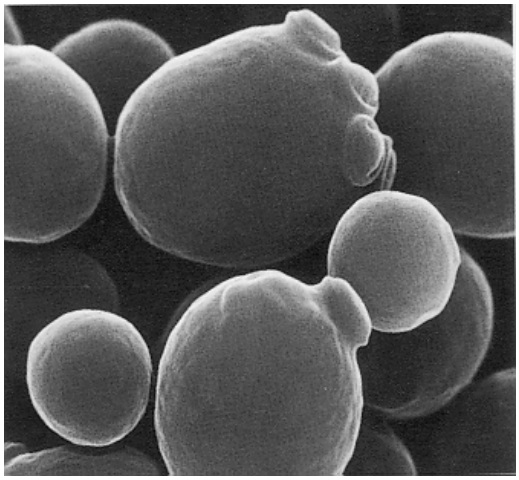




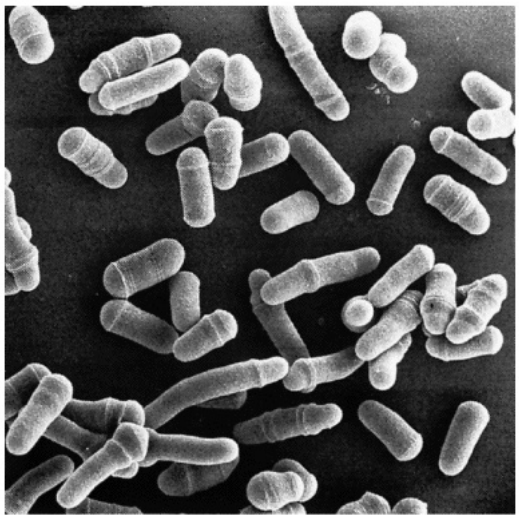
Bacteria: E. Coli



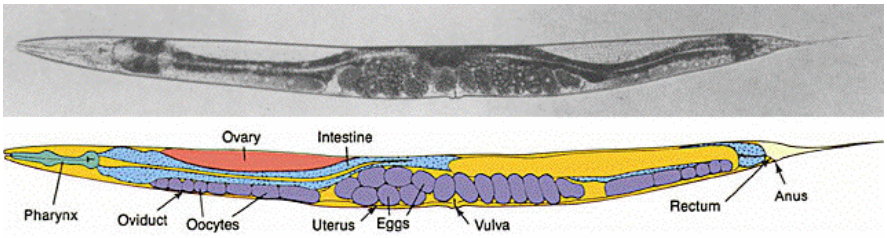
The Budding Yeast: *Saccharomyces cerevisiae*



The Fission Yeast: *Schizosaccharomyces pombe*



The Nematode: *Caenorhabditis elegans*



- **SMALL: ~ 250 µm**
- **TRANSPARENT**
- **959 CELLS**
- **300 NEURONS**
- **SHORT GENERATION TIME**
- **SIMPLE GROWTH MEDIUM**
- **SELF- FERTILIZING HERMAPHRODITE**
- **RAPID ISOLATION AND CLONING OF MULTIPLE TYPES OF MUTANT ORGANISMS**

The Fruit Fly: Drosophila Melanogaster

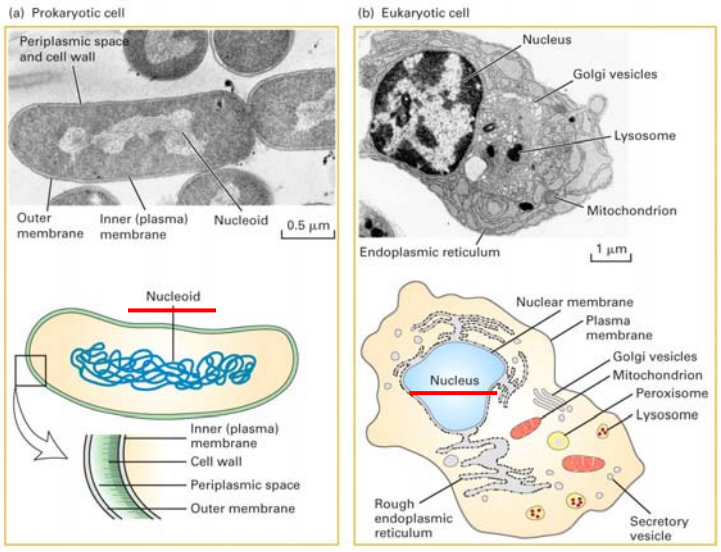
Normal

Ubx mutant

The Mouse

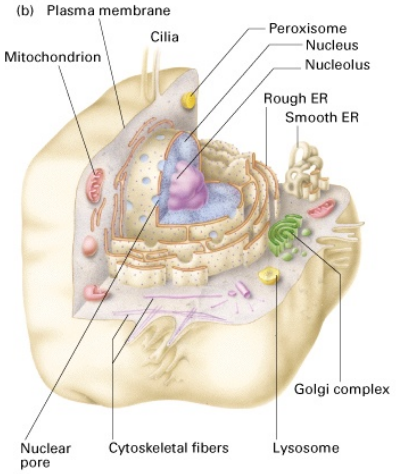
transgenic for human growth hormone

Prokaryotic and Eukaryotic Cells

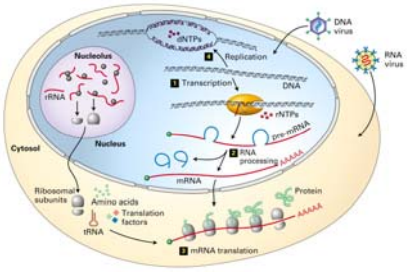


A Close Look of a Eukaryotic Cell

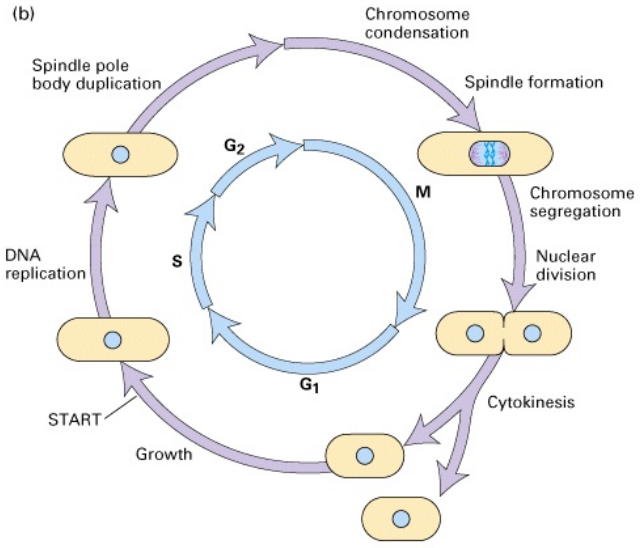
The structure:



The information flow:

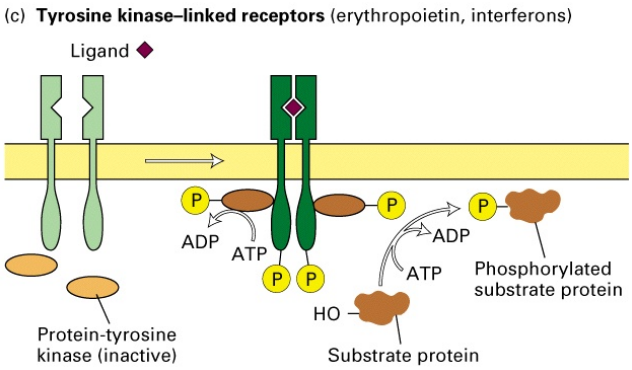


Cell Cycle

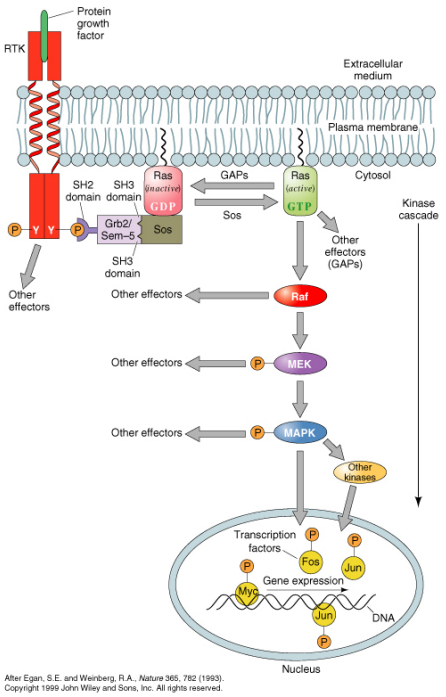


Signal Transduction

- **A variety of plasma membrane receptor proteins bind extracellular signaling molecules and transmit signals across the membrane to the cell interior**

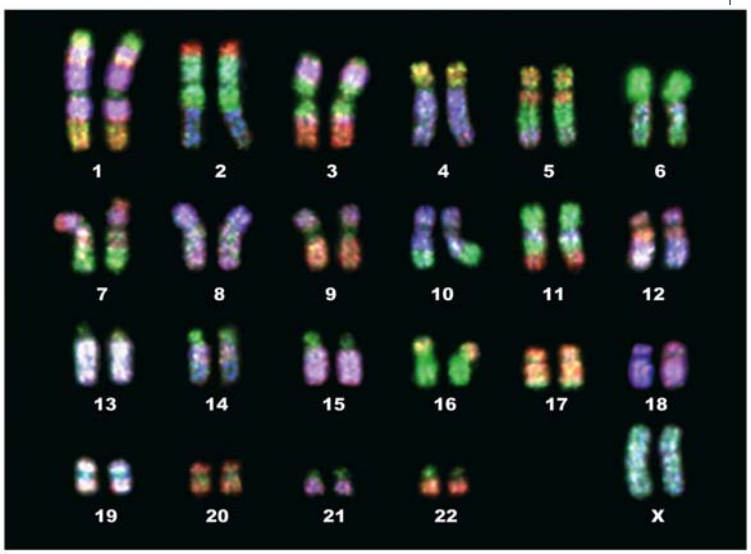


Signal Transduction Pathway

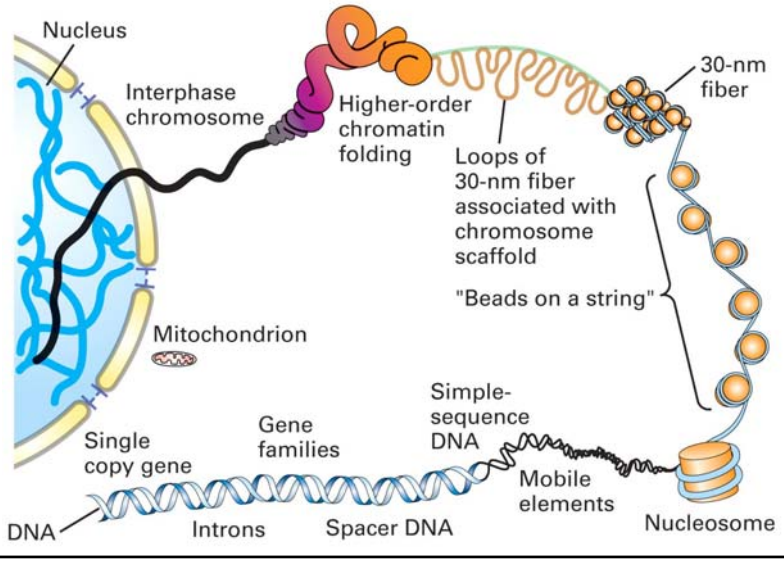


After Egan, S.E. and Weinberg, R.A., *Nature* 365, 782 (1993).

Functional Genomics and X-omics



A Multi-resolution View of the Chromosome



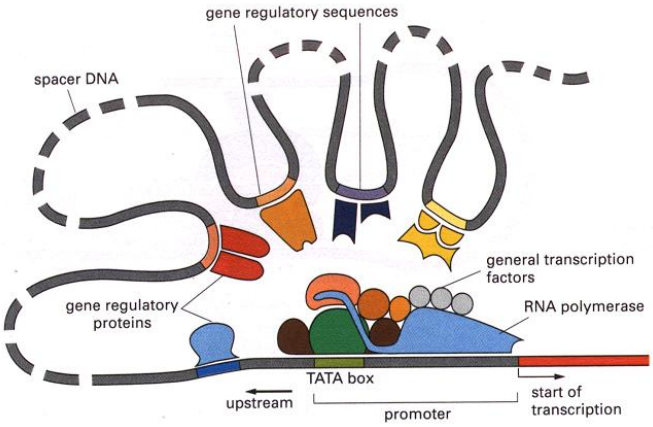
DNA Content of Representative Types of Cells

Organism	Number of base pairs (millions)	Number of encoded proteins	Number of chromosomes
PROKARYOTIC			
Mycoplasma genitalum (Bacterium)	0.58	470	1
Helicobacter pylori (Bacterium)	1.67	1590	1
Haemophilus influenza (Bacterium)	1.83	1743	1
EUKARYOTIC			
Saccharomyces cerevisiae (yeast)	12	5885	17
Drosophila melanogaster (insect)	165	13,601	4
Caenorhabditis elegans (worm)	97	19,099	6
Homo sapiens (human)	2900	30,000 TO 40,000	23
Arabidopsis thaliana (plant)	125	25,498	10

Functional Genomics

- The various **genome projects** have yielded the complete DNA sequences of many organisms.
  - E.g. human, mouse, yeast, fruitfly, etc.
  - Human: 3 billion base-pairs, 30-40 thousand genes.
- Challenge: **go from sequence to function**,
  - i.e., define the role of each gene and understand how the genome functions as a whole.

Regulatory Machinery of Gene Expression



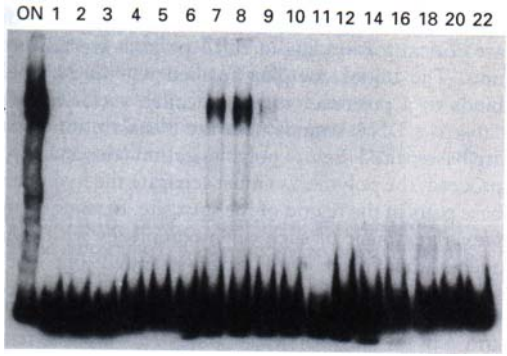
Classical Analysis of Transcription Regulation Interactions

"Gel shift": electorphoretic mobility shift assay ("EMSA") for DNA-binding proteins

Protein-DNA complex

Free DNA probe

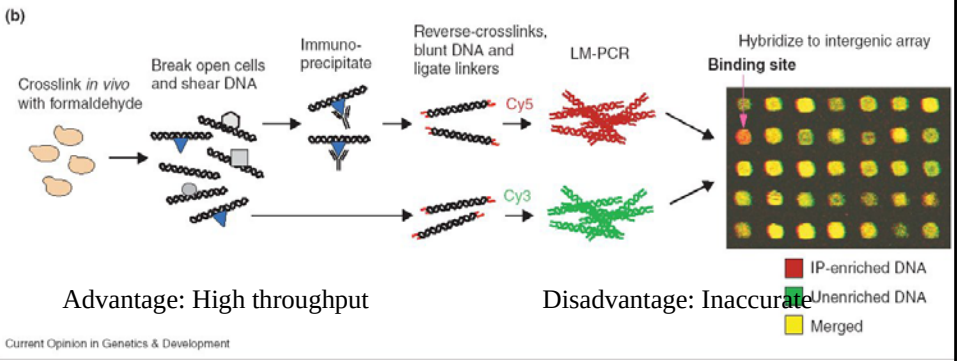
Advantage: sensitive



Disadvantage: requires stable complex; little "structural" information about which protein is binding

Modern Analysis of Transcription Regulation Interactions

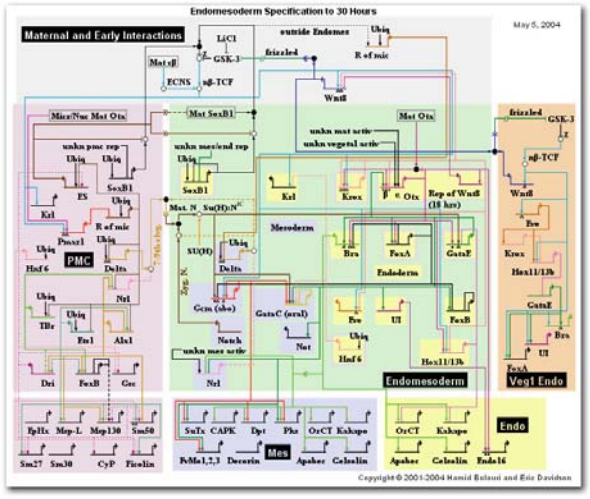
- **Genome-wide Location Analysis (ChIP-chip)**



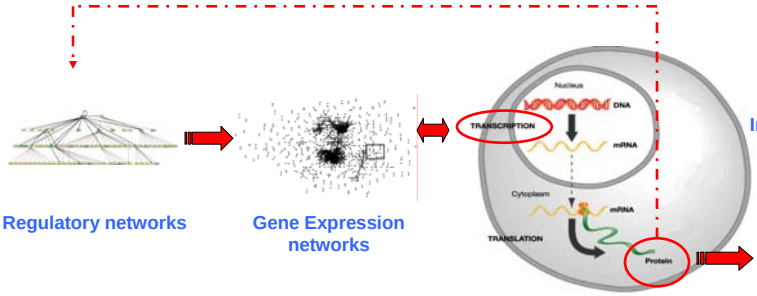
Advantage: High throughput

Disadvantage: Inaccurate

Gene Regulatory Network



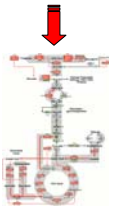
Biological Networks and Systems Biology



Regulatory networks

Gene Expression networks

Protein-protein Interaction networks



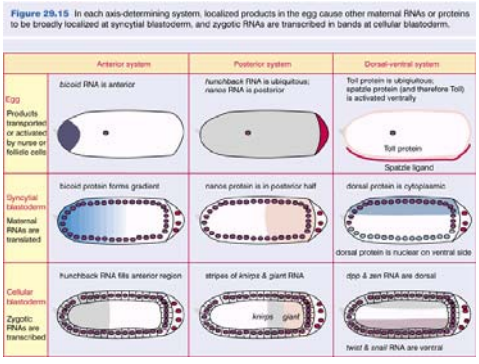
Metabolic networks

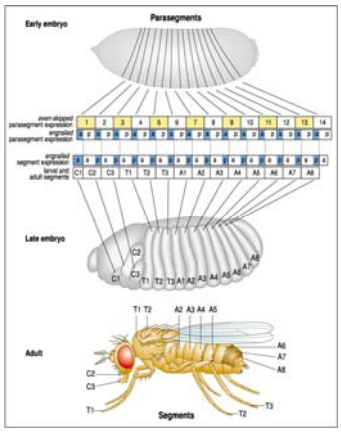
Systems Biology:

understanding cellular event under a system-level context

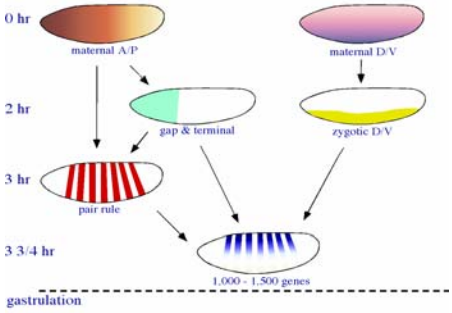
Genome + proteome + lipome + …

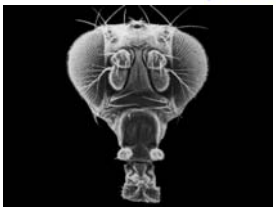
Gene Regulatory Functions in Development

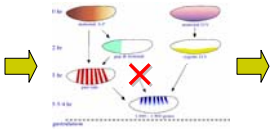




Temporal-spatial Gene Regulation and Regulatory Artifacts



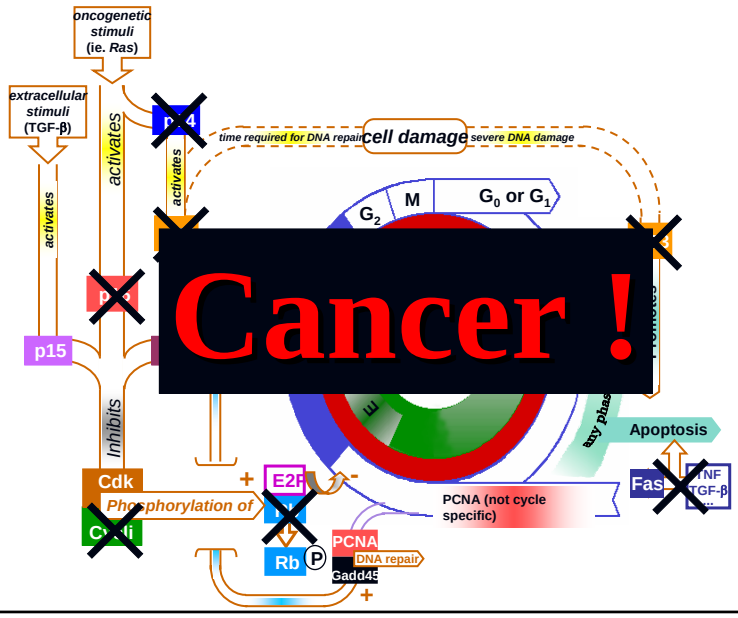




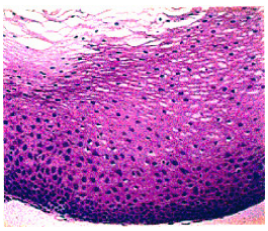
A normal fly

Hopeful monster?

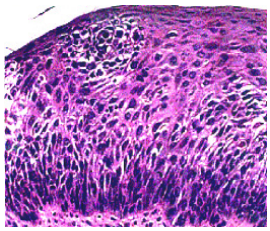
Gene Regulation and Carcinogenesis



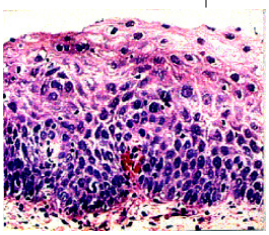
The Pathogenesis of Cancer



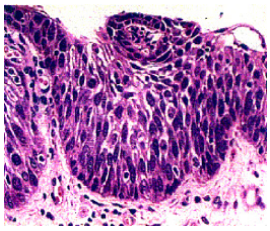
Normal



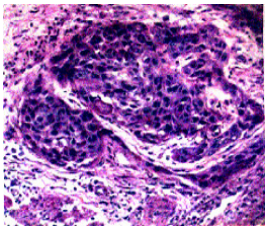
BCH



DYS



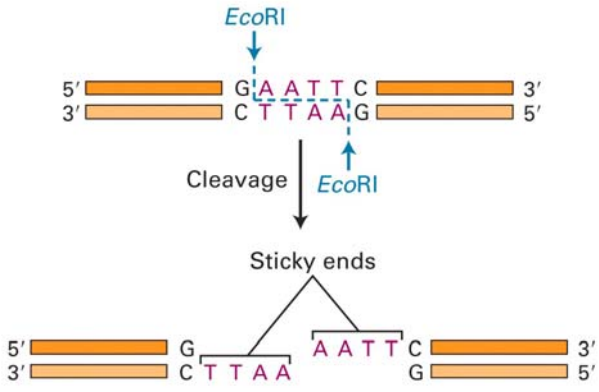
CIS



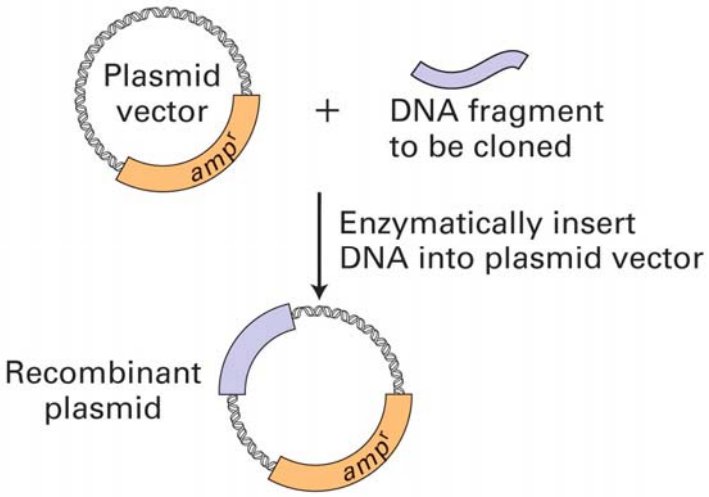
SCC

Bio-technology: Manipulating the Genome

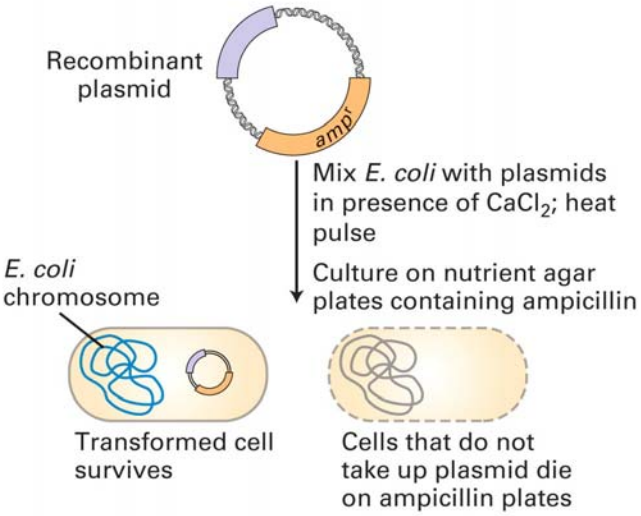
- Restriction Enzymes, naturally occurring in bacteria, that cut DNA at very specific places.



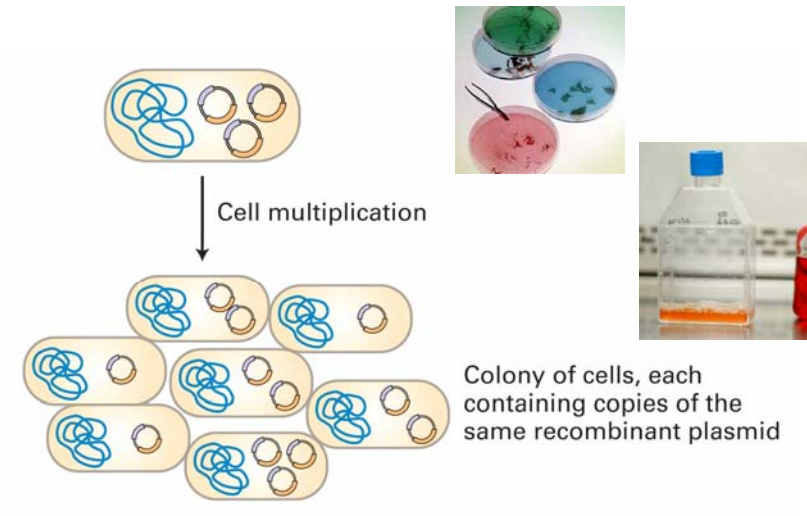
Recombinant DNA



Transformation

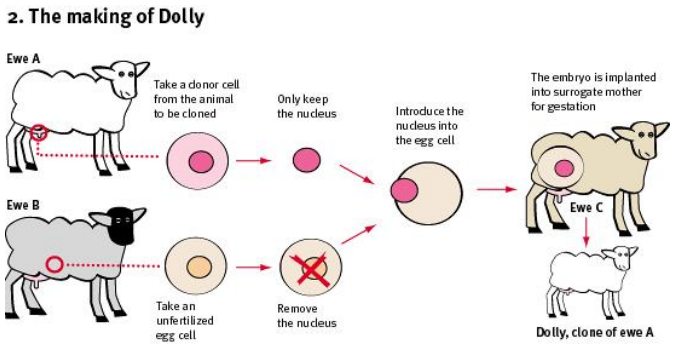


Formation of Cell Colony



How was Dolly cloned?

- Dolly is an exact genetic replica of another sheep.



Definitions

- Recombinant DNA: Two or more segments of DNA that have been combined by humans into a sequence that does not exist in nature.
- Cloning: Making an exact genetic copy. A **clone** is one of the exact genetic copies.
- Cloning vector: Self-replicating agents that serve as vehicles to transfer and replicate genetic material.

Software and Databases

- NCBI/NLM Databases Genbank, PubMed, PDB
  - DNA
  - Protein
  - Protein 3D
  - Literature

Entrez