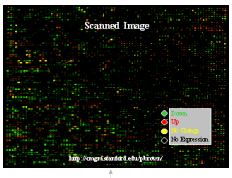
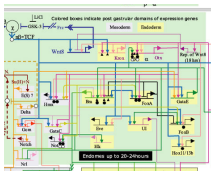
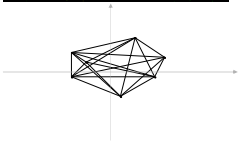
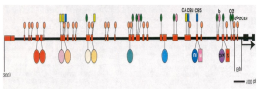
ECS 289A- Introduction to Computational Functional Genomics





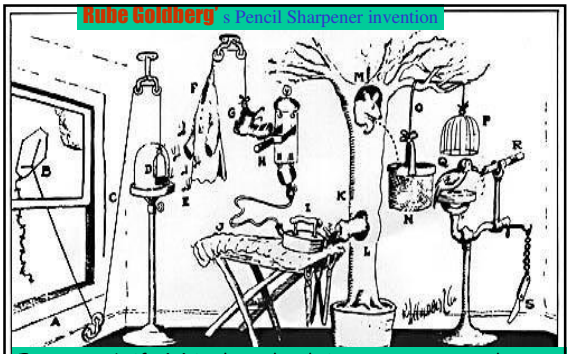




ECS 289A

ECS 289A

Rube Goldberg's Pencil Sharpener invention



Emergency knife (S) is always handy in case opossum or the woodpecker gets sick and can't work.

ECS 289A

- Prof. Vladimir Filkov
  3023 Engineering II
  filkov@cs.ucdavis.edu
- Office Hours: Tuesday, 2:00pm-4:00pm

ECS 289A

Administrativia

- ECS289A, 4 credits
- CRN 73358
- Tue/Thu 4:40-6pm, 70 SocSci
- http://www.cs.ucdavis.edu~/filkov/289a/
- No text required, papers and handouts
- Grading:
  - 60% project
  - 30% presentation(s)
  - 10% class participation

ECS 289A

Projects

- Projects will ideally combine biological knowledge and computational methods
- Some projects and reading lists out this week
  - Project proposals are due **April 14**
  - Progress reports are due **May 17**
  - Project presentations during last 2 weeks of classes
  - Final reports are due **Jun 14**
- If you have an idea come and talk to me asap

ECS 289A

Presentations

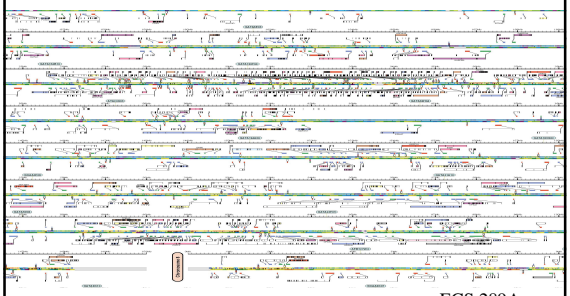
- Presentations can be of papers or software
- Each student will give 2 presentations during the class (40 min each)
- Need 2 volunteers to give first presentations on (1) Genomic Databases (GenBank, SGD, etc.) and (2) Microarray Resources (tools etc.)

Computational Biology

- Computational Scientists working on Molecular Biology Problems
- Different Scientific Cultures: CS vs Biology vs Statistics

- No unique definition: bioinformatics, computational biology, etc.
- Sub-areas in Comp Biology:
  - Genomics
  - Functional Genomics
  - Proteomics
  - Phylogenetics
  - Etc.

What good is Comp. Bio?



ECS 289A

What This Course is About

- Computational Functional Genomics
  - Genomics = molecular biology + robotics + informatics
    - Depends on tools and techniques of recombinant DNA technology (e.g. DNA sequencing)
    - The technologies used are high-throughput (e.g. sequencing, microarraying, etc.)
    - Data proceesed and analyzed by computers
  - Functional = exploration of gene and protein function in the cell processes (e.g. regulation, interactions, pathways, networks)
  - Computational = using computational/algorithmic methods

ECS 289A

Course Overview

Intro: Biology, Biotechnologies, Computational Methods

A. Central Dogma of Molecular Biology
B. High-Throughput technologies: DNA sequencing, Gene Expression Arrays, Protein-DNA, and Protein-Protein Interactions
C. CS for Biologists

Systems Modeling

A. Modeling, Simulation, Inference
B. Levels of modeling of genomic systems
C. Large-Scale Data Modeling

Bioinformatics and Data Mining of Large-Scale Data

A. Gene Expression analysis (statistics, classification, clustering)
B. Sequence analysis (promoter region analysis)
C. TF-DNA and Protein-Protein interactions analysis (topological properties and comparison)

ECS 289A

Course Overview, contd.

Gene Network Inference
- A. Graph Models
- B. Boolean Networks
- C. Bayesian Networks
- D. Linear Additive Models

Combining Heterogeneous Data Sources
- A. Sequence + gene expression
- B. Gene Expression + protein-protein interactions
- C. Methods for general data integration

ECS 289A

Lecture 1: Biological Preliminaries

- Life is survival of information
- Properties of life:
  - Information exchange (communication)
  - Procreation (passing on information)
  - Evolution (change)

- A machine that's set in motion and never stops

ECS 289A

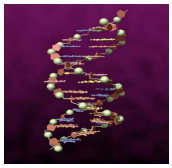
Preliminaries

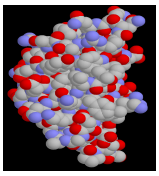
- Top-down Organization of life
  - Social groups etc.
  - Organisms
  - Species, etc.
  - Organs, Tissues
  - Cells: units of life

ECS 289A

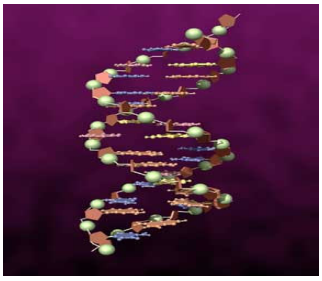
Preliminaries

- Inheritable information: Chromosomes (DNA)
- Day-to-day footwork: Proteins
- Both are complex polymer molecules





DNA



- DNA: String over the alphabet {A,C,G,T}
- A,T and C,G are complementary bases
- DNA codes for genes (proteins): genomics
- DNA codes for regulation: functional genomics

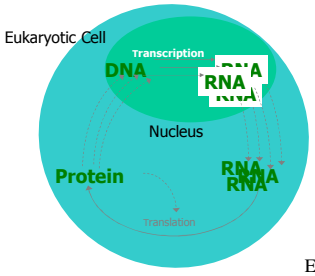
Genes and Proteins

- Nucleodtides, i.e. bases, A, C, G, and T
- Genes: units of inheritability
- Genes code for proteins!

ECS 289A

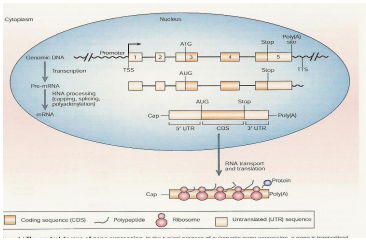
Genes to Proteins

Central Dogma: DNA ⇒ RNA ⇒ Proteins



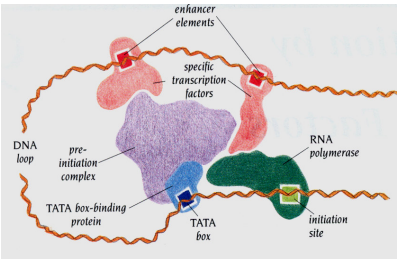
ECS 289A

Central Dogma



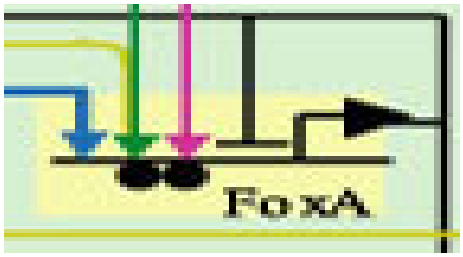
ECS 289A

Regulating Gene Expression

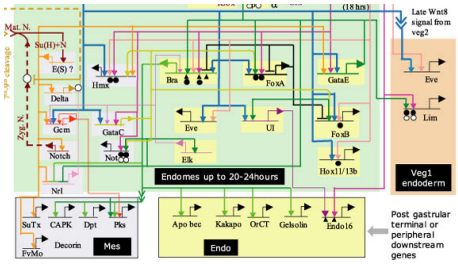


ECS 289A

Gene Regulation Simplified



ECS 289A



ECS 289A

Transcription

- Protein Binding
- Initiation
- Elongation
- Copying
- Termination

Translation

Second Base of Codon

First Base of Codon	U		C		A		G		Third Base of Codon
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC	phe	UCC	ser	UAC	tyr	UGC	cys	C
	UUA	leu	UCA	ser	UAA	STOP	UGA	STOP	A
	UUG	leu	UCG	ser	UAG	STOP	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC	leu	CCC	pro	CAC	his	CGC	arg	C
	CUA	leu	CCA	pro	CAA	gln	CGA	arg	A
	CUG	leu	CCG	pro	CAG	gln	CGG	arg	G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC	ile	ACC	thr	AAC	asn	AGC	ser	C
	AUA	ile	ACA	thr	AAA	lys	AGA	arg	A
	AUG	met*	ACG	thr	AAG	lys	AGG	arg	G
		* = START							
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC	val	GCC	ala	GAC	asp	GGC	gly	C
	GUA	val	GCA	ala	GAA	glu	GGA	gly	A
	GUG	val	GCG	ala	GAG	glu	GGG	gly	G