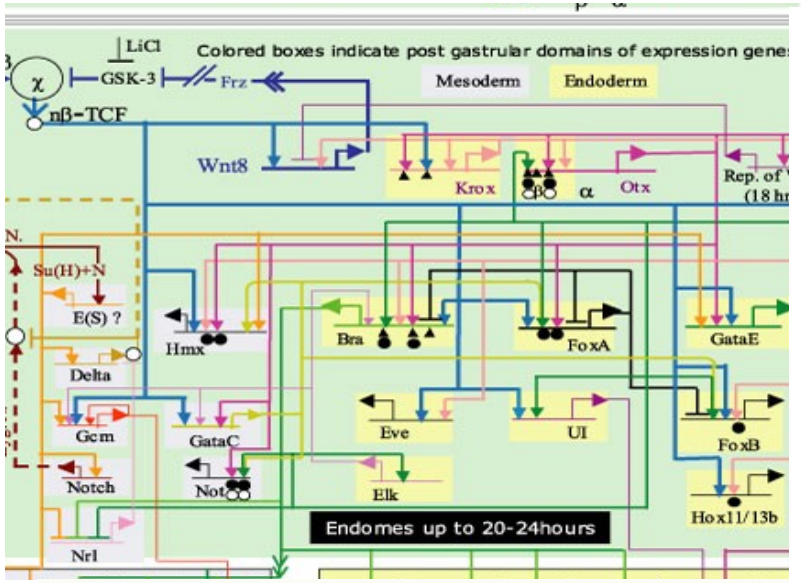
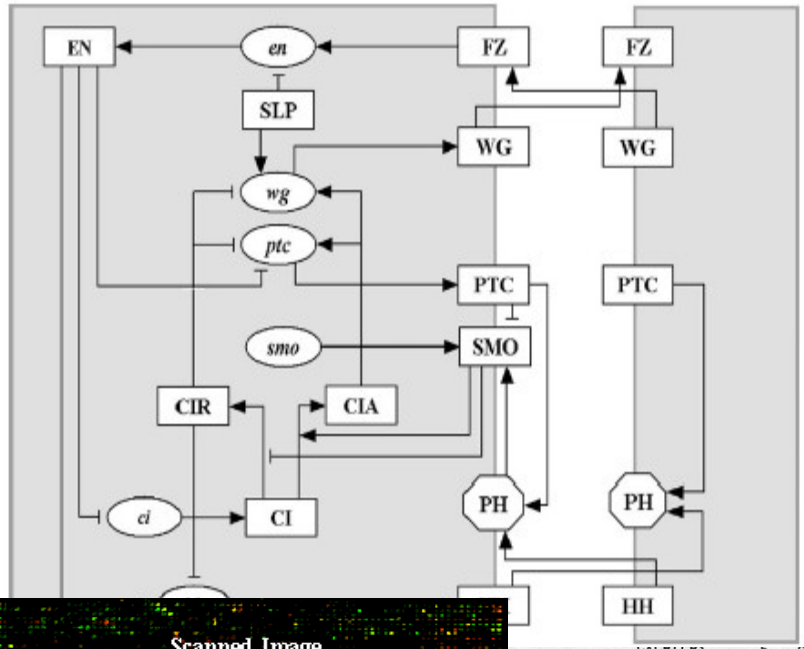
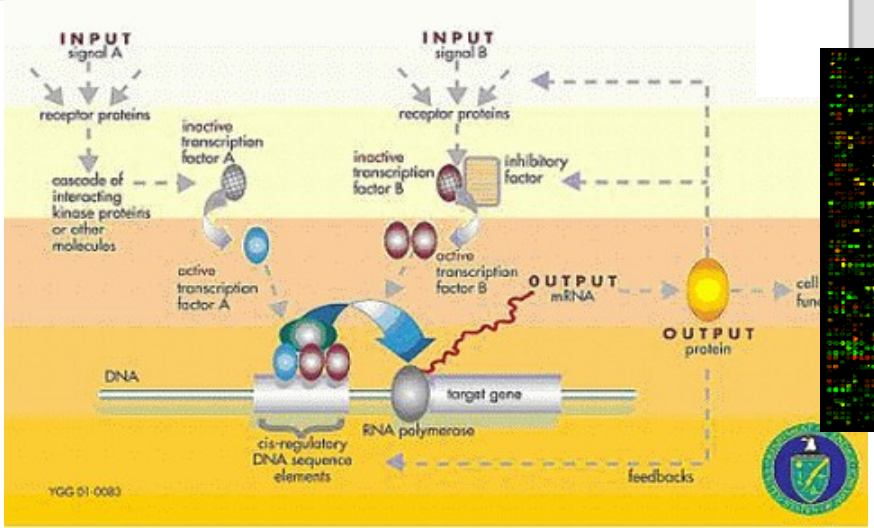
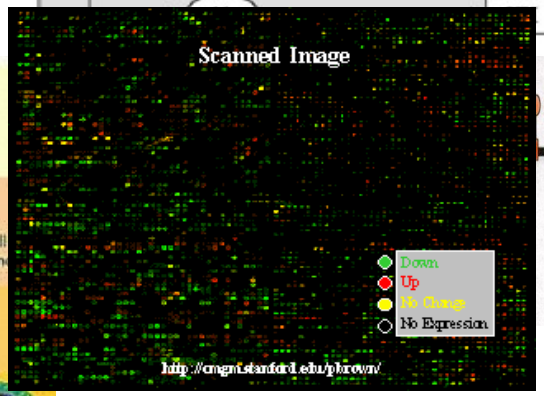
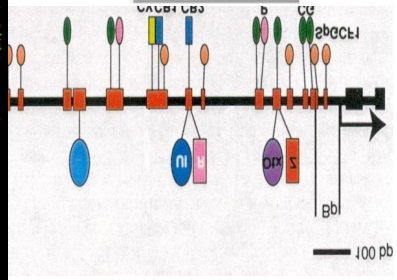
ECS 234: Introduction to Computational Functional Genomics











ECS 234

Administrativia

- Prof. Vladimir Filkov
  3023 Kemper
  filkov@cs.ucdavis.edu

- Appts:
  - Office Hours: Wednesday, 1:30-3p
  - Ask me or email me any time for appt

- ECS 234, 4 credits, CRN: 54135
- http://www.cs.ucdavis.edu~/filkov/234/
- No text required, papers and handouts
- Grading:
  - 60% project
  - 30% presentations
  - 10% class participation

Projects

- Groups should be formed by **Jan 16**
- Projects will ideally combine biological topics and computational methods
  - Project proposals are due **Jan 22**
  - Progress reports are due **Feb 17**
  - Project presentations **March 12**
  - Final reports are due **March 17**
- If you have an idea write to me and/or come and talk to me asap

Presentations

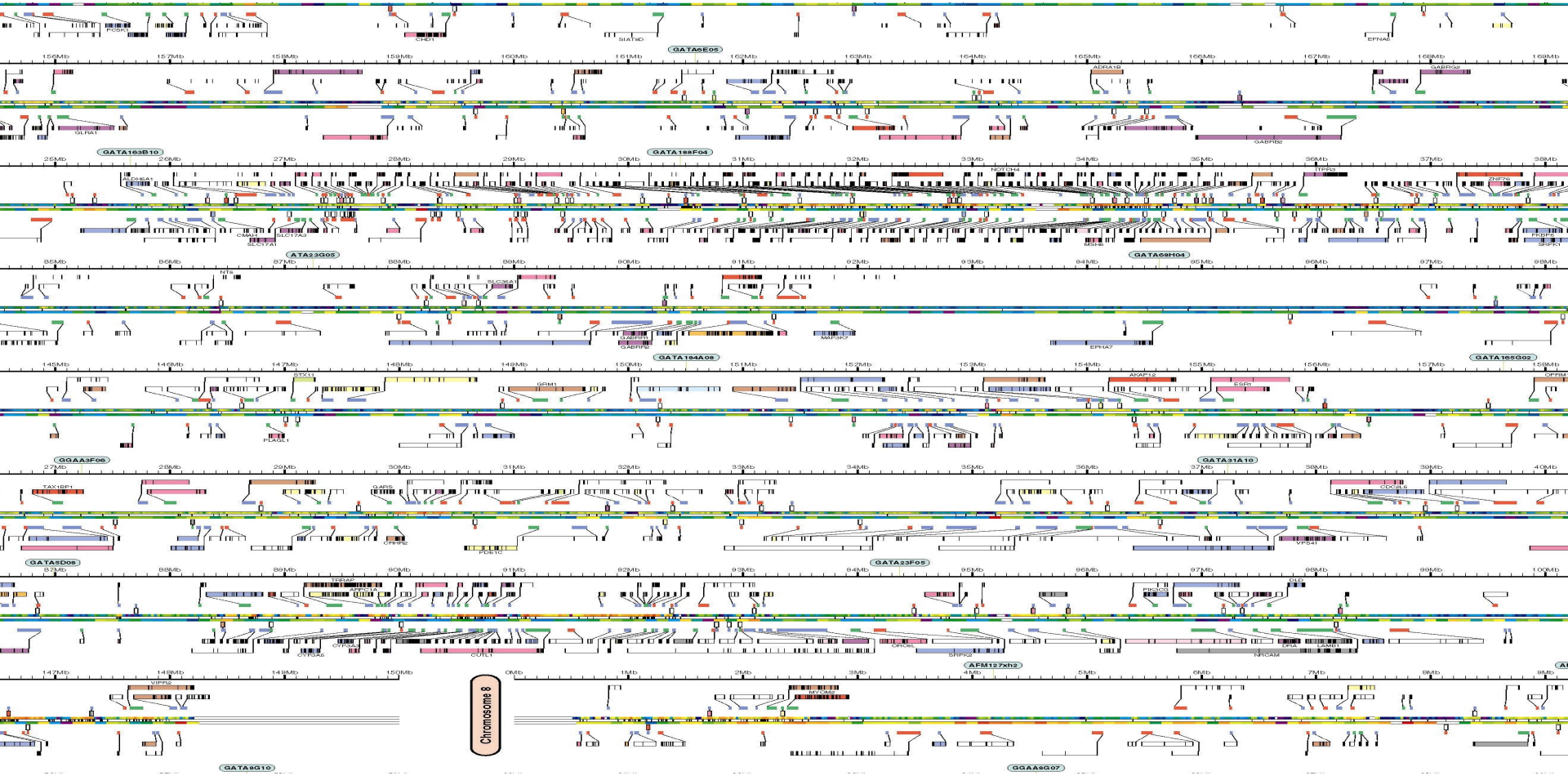
- Presentations: paper and/or software
- Each presentation will be during class time, 40-45 min long and 10-20min for discussion
- Need 2 volunteers for presentations on
  - Normalizing microarray data (different methods comparison),
  - Biclustering of microarray data,
  - or your choice.

Computational Biology and Bioinformatics

- Computational Biology: Computational scientists researching and developing methods for solving molecular biology problems
- Bioinformatics: More practical, deals with the appropriate use of tools to solve genomics problems
- Necessary background: molecular biology+computer science+statistics

- Sub-areas in Computational Biology/Bioinformatics:
  - Genomics
  - Functional Genomics
  - Proteomics
  - Phylogenetics
  - Bionetworks
  - etc.

What good is Comp. Bio?



ECS 234

This Course: Computational Functional Genomics

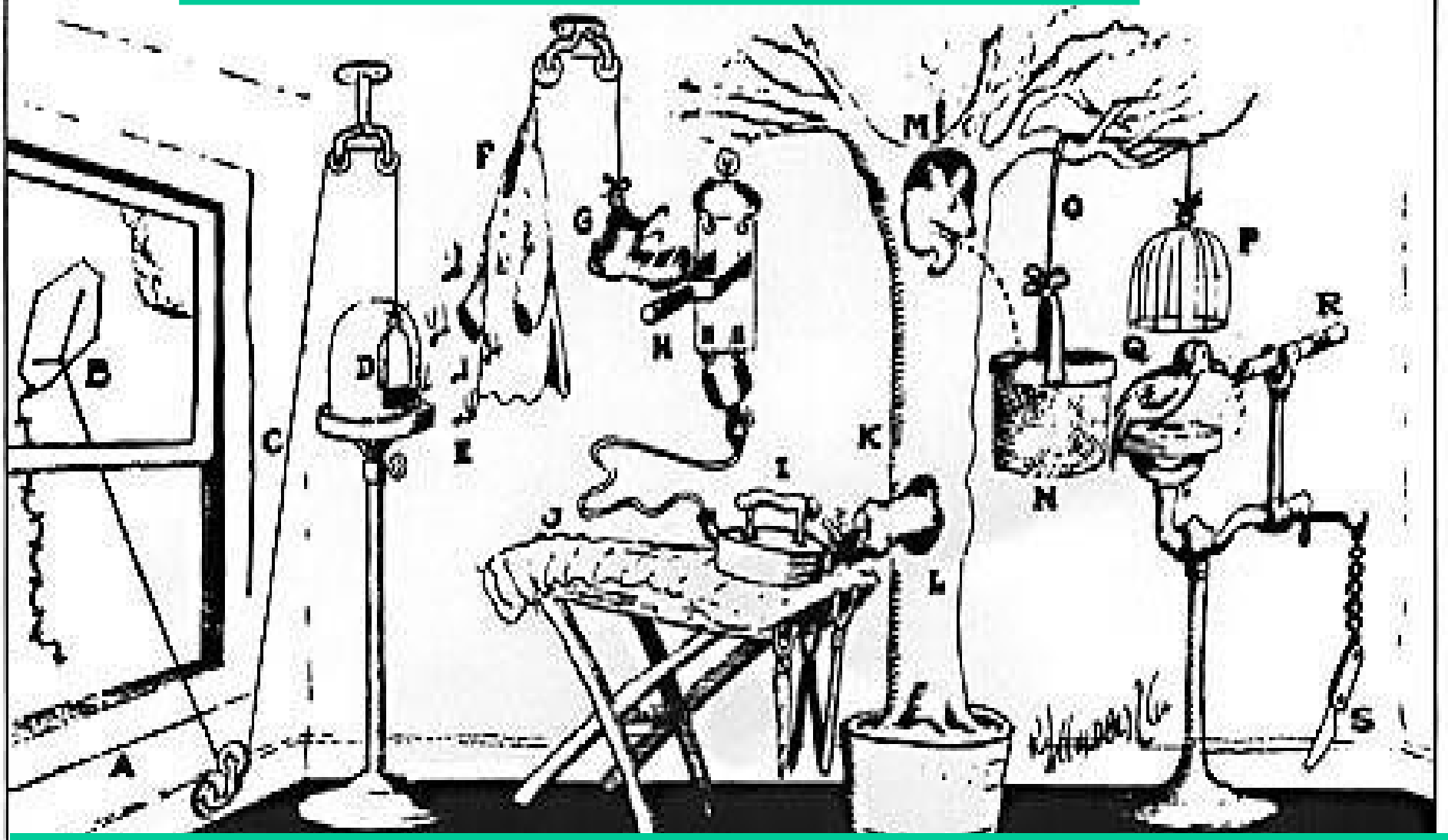
- Genomics = molecular biology + robotics + informatics
  - Depends on tools and techniques of recombinant DNA technology (e.g. DNA sequencing)
  - The technologies used are high-throughput (e.g. sequencing, microarraying, etc.)
  - Data proceesed and analyzed by computers
- Functional = exploration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

How Is Function Discovered?

- Observe Nature, analyze data, make and test hypotheses, repeat
- Small scale: gene by gene
  - Manually curated databases (e.g. MIPS for yeast)
- Large scale: robotics + computational methods
  - Sequence
  - Gene Expression
  - Protein Interactions

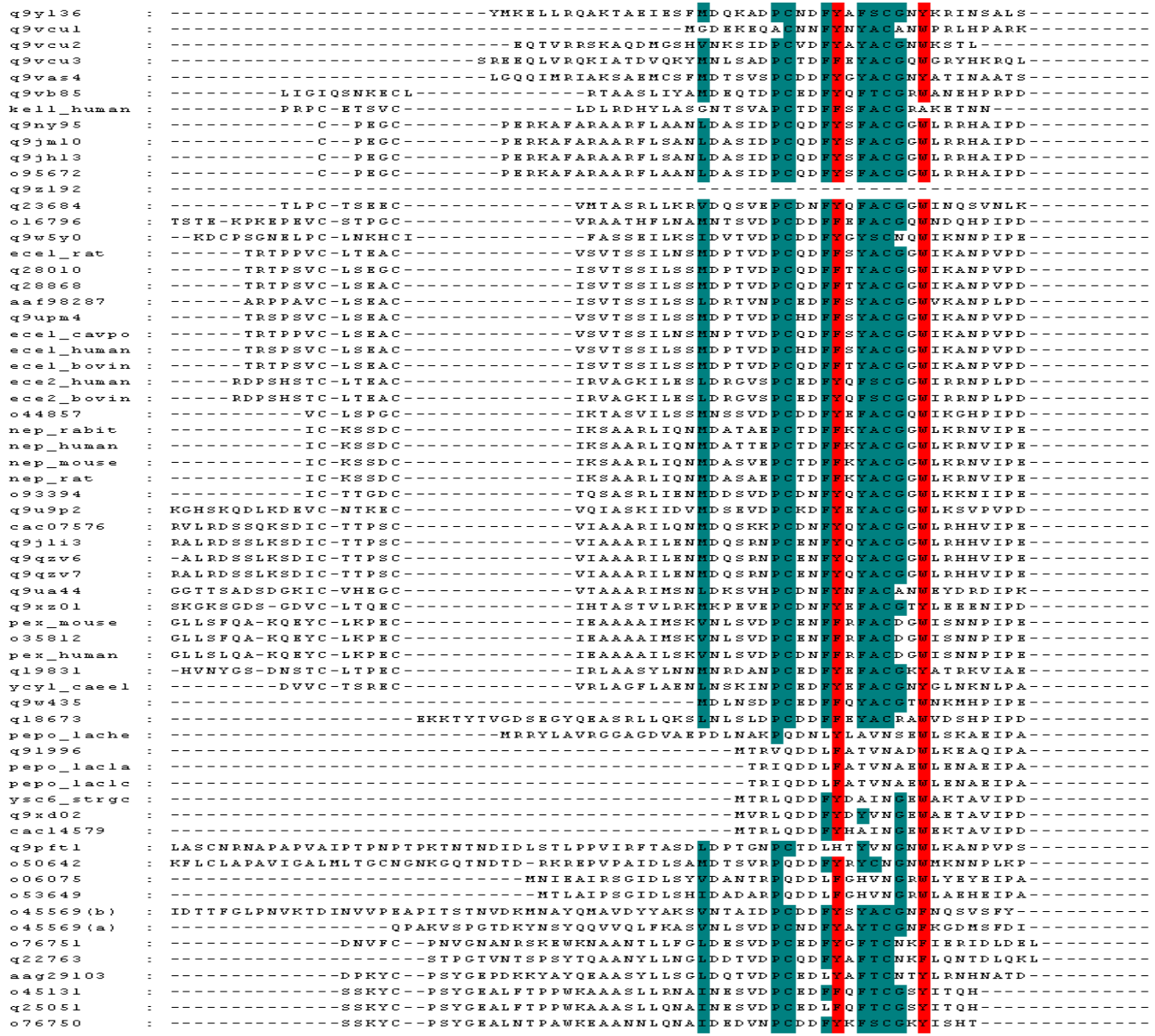
The System and Its Parts

Rube Goldberg's Pencil Sharpener invention



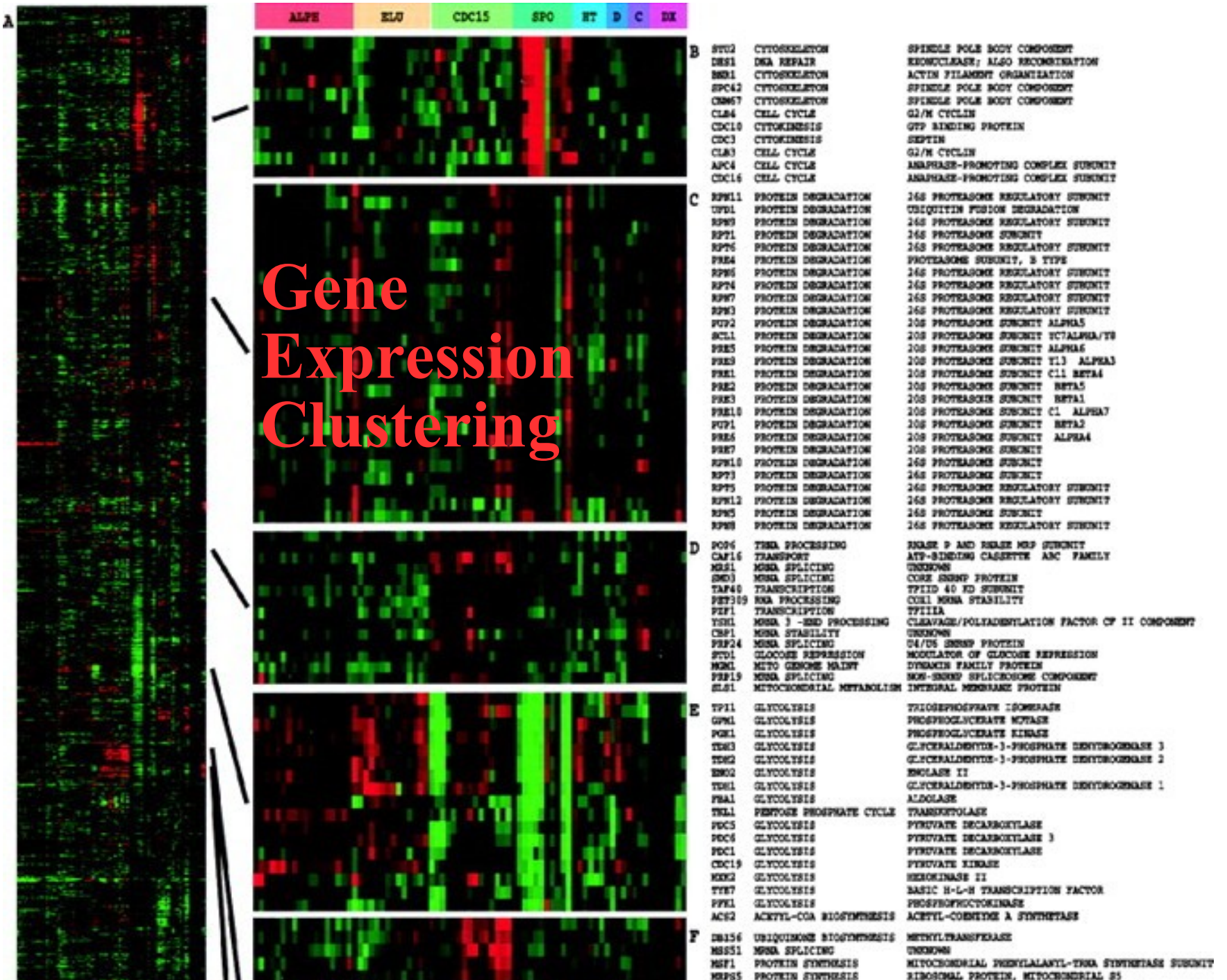
Emergency knife **(S)** is always handy in case opossum or the woodpecker gets sick and can't work.

Examples of Genomic Data and Computational Tools for Functional Discovery in Genomics

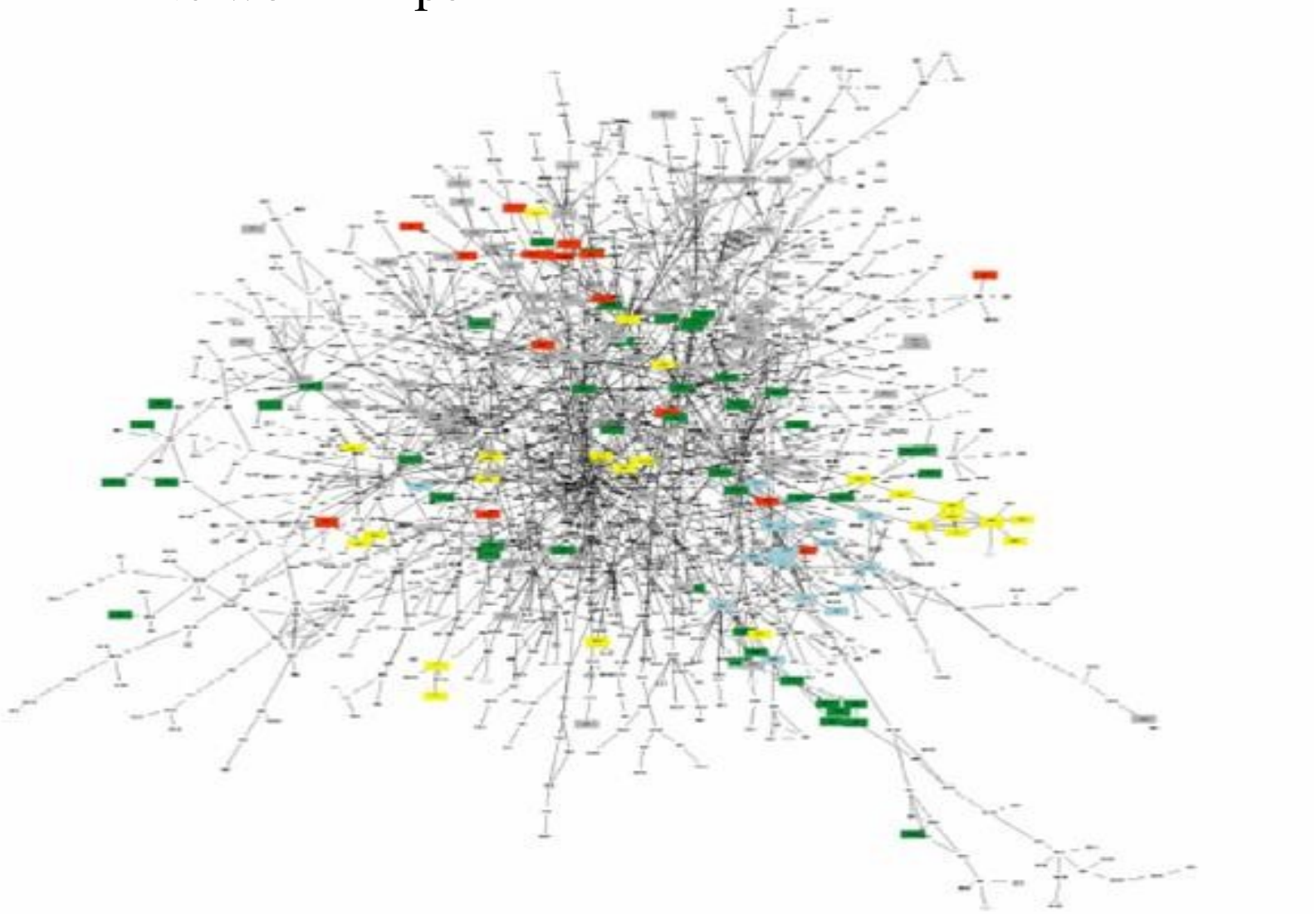


Sequence Homology (multiple alignment)

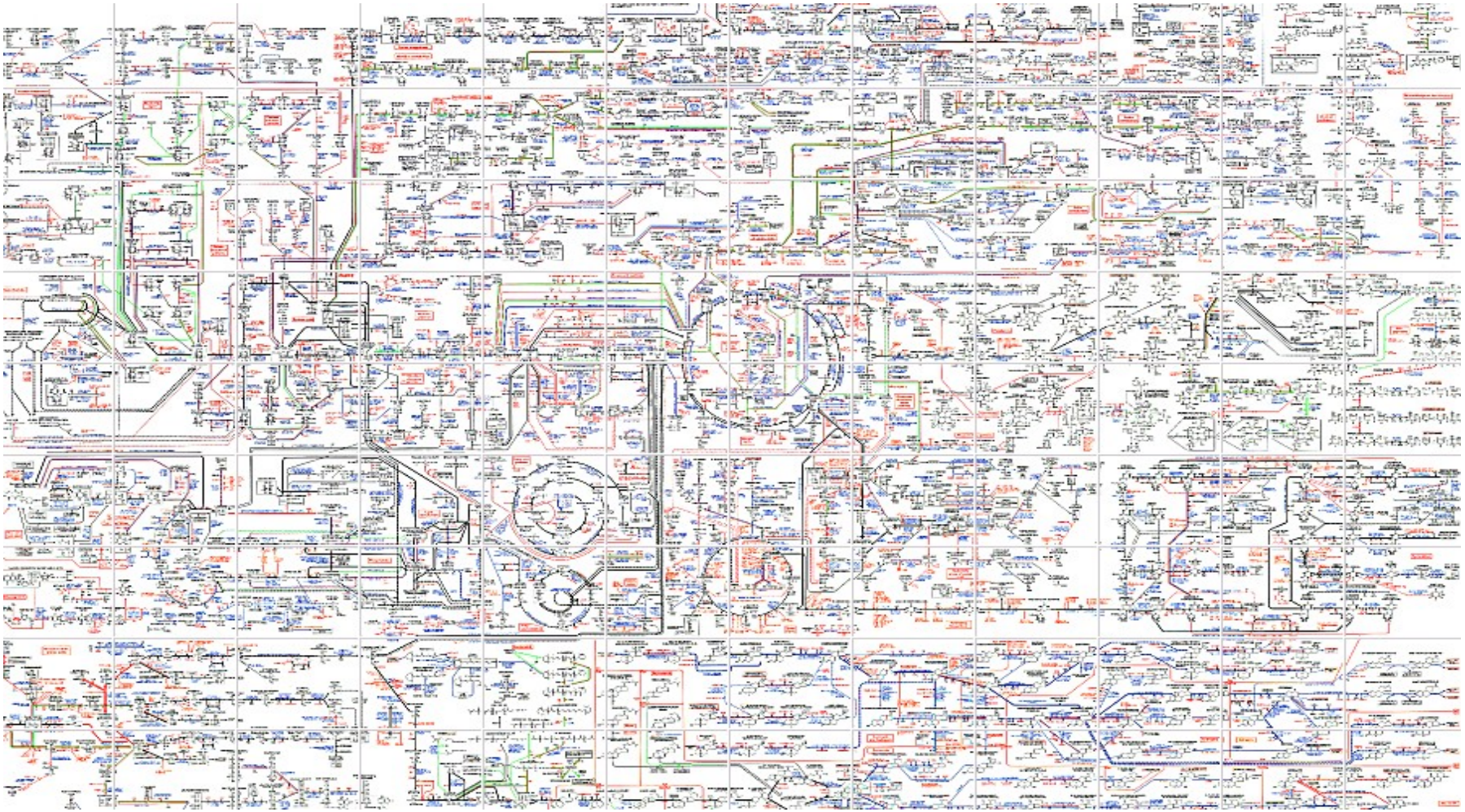
ECS 234


A
ALPH
ELU
CDC15
SPO
HT
D
C
DX
B
STU2 CYTOSKELETON SPINDLE POLE BODY COMPONENT
DES1 DNA REPAIR RHONUCLEASE; ALSO RECOMBINATION
BNR1 CYTOSKELETON ACTIN FILAMENT ORGANIZATION
SPC42 CYTOSKELETON SPINDLE POLE BODY COMPONENT
CNM67 CYTOSKELETON SPINDLE POLE BODY COMPONENT
CLB6 CELL CYCLE G2/M CYCLIN
CDC10 CYTOKINESIS GTP BINDING PROTEIN
CDC3 CYTOKINESIS SEPTIN
CLB3 CELL CYCLE G2/M CYCLIN
APC4 CELL CYCLE ANAPHASE-PROMOTING COMPLEX SUBUNIT
CDC16 CELL CYCLE ANAPHASE-PROMOTING COMPLEX SUBUNIT
Gene
Expression
Clustering
C
RPN11 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
UFD1 PROTEIN DEGRADATION UBIQUITIN FUSION DEGRADATION
RPN9 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPT1 PROTEIN DEGRADATION 26S PROTEASOME SUBUNIT
RPT6 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
PRE4 PROTEIN DEGRADATION PROTEASOME SUBUNIT, B TYPE
RPN6 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPT4 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPN7 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPN3 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
PUP2 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT ALPHA5
SCL1 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT YC7ALPHA/Y8
PRE5 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT ALPHA6
PRE9 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT Y13 ALPHA3
PRE1 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT C11 BETA4
PRE2 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT BETA5
PRE3 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT BETA1
PRE10 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT C1 ALPHA7
PUP1 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT BETA2
PRE6 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT ALPHA4
PRE7 PROTEIN DEGRADATION 20S PROTEASOME SUBUNIT
RPN10 PROTEIN DEGRADATION 26S PROTEASOME SUBUNIT
RPT3 PROTEIN DEGRADATION 26S PROTEASOME SUBUNIT
RPT5 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPN12 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
RPN5 PROTEIN DEGRADATION 26S PROTEASOME SUBUNIT
RPN8 PROTEIN DEGRADATION 26S PROTEASOME REGULATORY SUBUNIT
D
POP6 TRNA PROCESSING RNASE P AND RNASE MRP SUBUNIT
CAF16 TRANSPORT ATP-BINDING CASSETTE ABC FAMILY
MRS1 MRNA SPLICING UNKNOWN
SMD3 MRNA SPLICING CORE SNRNP PROTEIN
TAF40 TRANSCRIPTION TFIID 40 KD SUBUNIT
PET309 RNA PROCESSING COX1 MRNA STABILITY
PEF1 TRANSCRIPTION TFIIIA
YSH1 MRNA 3 -END PROCESSING CLEAVAGE/POLYADENYLATION FACTOR CF II COMPONENT
CBP1 MRNA STABILITY UNKNOWN
PRP24 MRNA SPLICING U4/U6 SNRNP PROTEIN
STD1 GLUCOSE REPRESSION MODULATOR OF GLUCOSE REPRESSION
MGM1 MITO GENOME MAINT DYNAMIN FAMILY PROTEIN
PRP19 MRNA SPLICING NOS-SNRNP SPLICEOSOME COMPONENT
SLS1 MITOCHONDRIAL METABOLISM INTEGRAL MEMBRANE PROTEIN
E
TPI1 GLYCOLYSIS TRIOSEPHOSPHATE ISOMERASE
GPM1 GLYCOLYSIS PHOSPHOGLYCERATE MUTASE
PGK1 GLYCOLYSIS PHOSPHOGLYCERATE KINASE
TDH3 GLYCOLYSIS GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE 3
TDH2 GLYCOLYSIS GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE 2
ENO2 GLYCOLYSIS ENOLASE II
TDH1 GLYCOLYSIS GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE 1
FBA1 GLYCOLYSIS ALDOLASE
TKL1 PENTOSE PHOSPHATE CYCLE TRANSKETOLASE
PDC5 GLYCOLYSIS PYRUVATE DECARBOXYLASE
PDC6 GLYCOLYSIS PYRUVATE DECARBOXYLASE 3
PDC1 GLYCOLYSIS PYRUVATE DECARBOXYLASE
CDC19 GLYCOLYSIS PYRUVATE KINASE
HXK2 GLYCOLYSIS HEXOKINASE II
TYE7 GLYCOLYSIS BASIC H-L-H TRANSCRIPTION FACTOR
PFK1 GLYCOLYSIS PHOSPHOFRUCTOKINASE
ACS2 ACETYL-COA BIOSYNTHESIS ACETYL-COENZYME A SYNTHETASE
F
COQ5 UBIQUINONE BIOSYNTHESIS METHYLTRANSFERASE
MSS51 MRNA SPLICING UNKNOWN
MSF1 PROTEIN SYNTHESIS MITOCHONDRIAL PHENYLALANYL-TRNA SYNTHETASE SUBUNIT
MRPS5 PROTEIN SYNTHESIS RIBOSOMAL PROTEIN, MITOCHONDRIAL S5

PPI Network Maps



Data Integration



ECS 234

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

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)

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