Introduction to Bioinformatics

Biostatistics & Medical Informatics 576

Computer Sciences 576

Fall 2017

Irene Ong

Irene.ong@wisc.edu

www.biostat.wisc.edu/bmi576/

Goals for today

- **Administrivia**
- Course Topics
- Short survey of interests/background

Course Web Site

- www.biostat.wisc.edu/bmi576
- syllabus
- readings
- tentative schedule
- lecture slides in PDF/PPT
- homework
- link to Piazza discussion board
- etc.

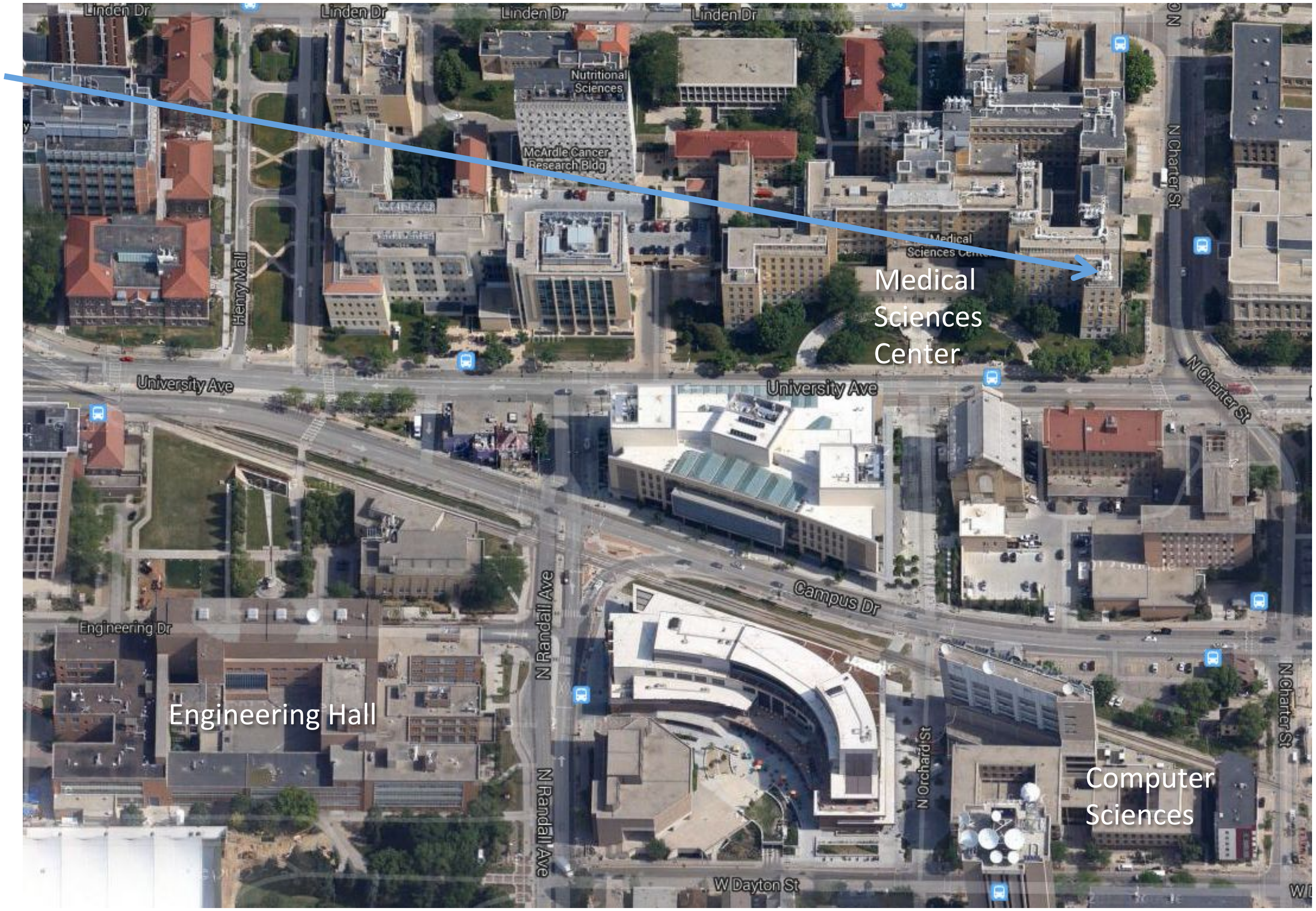
Your Instructor: Irene Ong

- email: irene.ong@wisc.edu
- website: www.biostat.wisc.edu/~ong/
- office: 4710 Medical Sciences Center

- Assistant Professor in the Department of Obstetrics & Gynecology and Biostatistics & Medical Informatics
- Research interests: machine learning and probabilistic modeling applied to biological and clinical data

Finding My Office: 4710 Medical Sciences Center

my office



- slightly confusing building(s)
- best bet: use Charter Street entrance

Course TAs

- Wei Zhang
  - wzhang336@wisc.edu
  - Office: 6397 Computer Sciences
- Ankit Pensia
  - pensia@wisc.edu
  - Office: 1302 Computer Sciences

Office Hours

- To be announced
- Will begin next week
- Poll to determine a good office hour schedule for TAs and me
  - Please fill out poll to increase the likelihood that our office hours will work for you!
  - With a class of this size we have limited ability to accommodate appointments outside of office hours
- You are encouraged to visit our office hours!

Expected Background

- CS 367 (Intro to Data Structures) or equivalent
  - Arrays
  - Hash tables
  - Trees
  - Graphs
- Statistics: good if you've had at least one course, but not required
  - Continuous/Discrete probability distributions
  - Conditional and joint distributions
- Molecular biology: no knowledge assumed, but an interest in learning some basic molecular biology is mandatory

Course grading

- 6 or so homework assignments: 55%
  - Programming problems
  - Written exercises
- midterm exam: 20%
- final exam: 20%
- participation: 5%

Homework assignments

- All homework (written and programming) is to be done <u>individually</u>
- For programming exercises, you should use one of:
  - Python, Java, C/C++
  - Perl (somewhat discouraged, Perl is often difficult to read)
  - R (somewhat discouraged, not general-purpose)
  - Matlab (somewhat discouraged, not general-purpose)
- These are the most commonly used languages in bioinformatics
- Use a language not on this list at your own risk
- All homework will be submitted electronically
- You are strongly encouraged to typeset your written work (e.g., with LaTeX or Word)
- 5 free late days except HWs before midterm and final
- Issues related to homework grades must be resolved within 1 week of distribution of grade

Computing Resources for the class

- UNIX workstations in Dept. of Biostatistics & Medical Informatics
  - accounts will be created soon
  - two machines
    - mi1.biostat.wisc.edu
    - mi2.biostat.wisc.edu
- UNIX tutorial: http://pages.cs.wisc.edu/~deppeler/tutorials/UNIX/

Exams

- Midterm: October 31st, in class
  - will cover first three modules
- Final: December 21st, 12:25pm-2:25pm
  - will cover last three modules

Participation

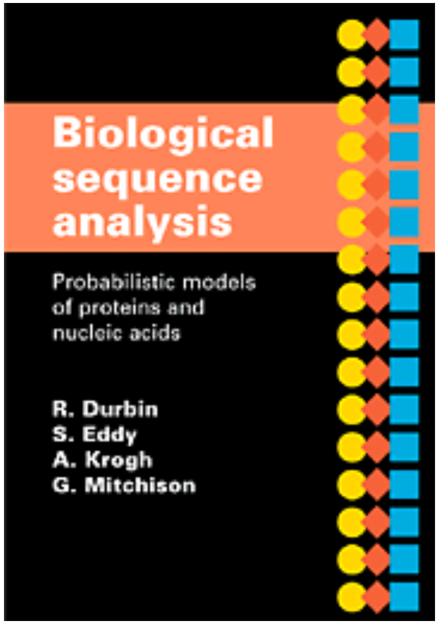
- Attending lectures is required
- A significant amount of material is not in the slides (e.g., board work)
- Questions are welcome during class
- Piazza

Piazza Discussion Forum

- Instead of a mailing list
- http://piazza.com/wisc/fall2017/bmics576/home
- Please consider posting your questions to Piazza first, before emailing the instructor or TAs
- Consider answering your classmates' questions!
- Quick announcements will also be posted to Piazza
- Email instructor or TAs with questions inappropriate for Piazza
  - Expect email response within 24 hours

Course readings

- Readings assigned for each lecture – please read these ahead of time
- Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. Cambridge University Press, 1998.



- Articles from the primary literature (scientific journals, etc.)

Reading assignment for Sep 12th

- ***Life and Its Molecules A Brief Introduction*** by Lawrence Hunter
  - http://www.biostat.wisc.edu/bmi576/papers/hunter04.pdf

Goals for today

- Administrivia
- **Course Overview**
- Short survey of interests/background

Learning goals of this class

- Gain an overview of different problem areas in bioinformatics
- Understanding of significant & interesting algorithms
- Ability to apply the computational concepts to related problems in biology and other areas
- Ability to understand scientific articles about more cutting-edge approaches
- Foundation to enable independent learning and deeper study of related topics

What is Bioinformatics?

- The term Bioinformatics was coined in the 1970s
- Very close cousin: Computational Biology
- An interdisciplinary field rooted in computer and information sciences and life sciences.
- Draws from other areas such as
  - Math, statistics, machine learning, physics, genetics, evolutionary biology, biochemistry
- Definitions from the National Institute of Health
  - Bioinformatics: Research, development, or application of computational tools and approaches to make the vast, diverse and complex life sciences data more understandable and useful.
  - Computational biology: The development and application of mathematical and computational approaches to address theoretical and experimental questions in biology

Why Bioinformatics?

- Biology is a data-driven field
  - By far the richest types and sources of data
  - Biological systems are complex and noisy
- Need informatics tools to
  - Store, manage, mine, visualize biological data
  - Model biological complexity
  - Generate testable hypotheses
- Many biological questions translate naturally into a computational problem
  - Pattern extraction
  - Search
  - Inferring function of bio-chemical entities
  - Finding relationships among entities

Bioinformatics then and now

- 1990s: Mostly data storage, search and retrieval of sequence data, and databases to store biological knowledge
- Now: abstract knowledge and principles from large-scale data, to present a complete representation of cells and organisms, and to make computational predictions of systems of higher complexity such as cellular interaction networks and global phenotypes

Kanehisa and Bork, 2003

A few important dates

Year	Biological landmarks	Computational advances
1953	DNA's double helix structure	
1967	Availability of protein sequences	First database of protein sequences by Margaret Dayhoff
1970-81		Global and local alignment algorithms
1987		Swissprot: First indexed database
1990		BLAST, a fast program to search large databases for query sequences
1995-1998	Several whole genomes sequenced	HMMs for sequence analysis
1997	First DNA microarrays	Clustering of expression data
2000	Large collections of expression data	Probabilistic graphical models to analyze networks
2003	Human genome sequence published	
2005-	Growth of next-generation sequencing methods	Advanced statistical and machine learning methods for next-gen sequencing data

Overview of bioinformatics topics

- Sequence assembly
- Sequence alignment
- Phylogenetic trees
- Genome annotation
- Clustering and analysis of “omic” datasets
- Modeling and analysis of biological networks

Computer Science Topics

- Algorithms
  - Graphs
  - Exact
  - Greedy
  - Dynamic Programming
  - Branch and bound
  - Heuristics
- Computational Complexity

Statistics Topics

- Probability for discrete random variables
- Markov Chains
- Hidden Markov Models
- Maximum Likelihood
- Expectation-Maximization
- Bayesian networks

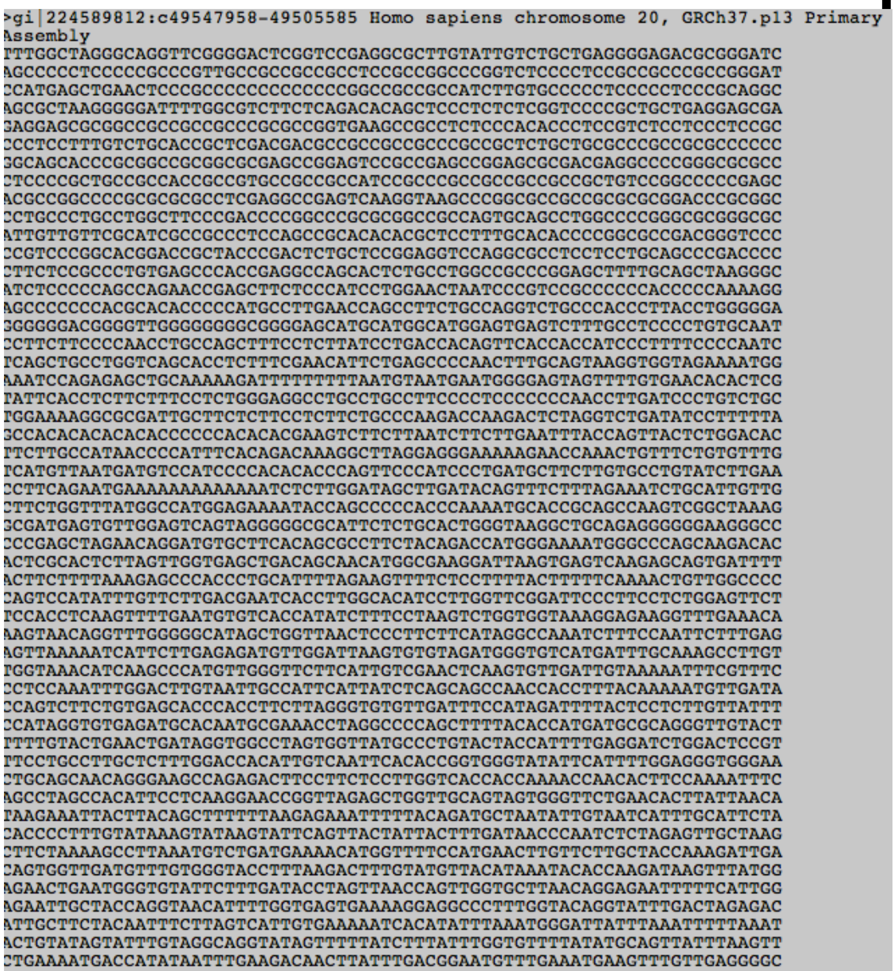
Sequence Assembly

- How do we determine the genome sequence of an organism?

Topics in sequence assembly

- Sequencing technologies
- Fragment assembly problem
- Spectral assembly problem
- Graph algorithms
- Assembly in practice

Sequence comparison: How similar are the sequences?



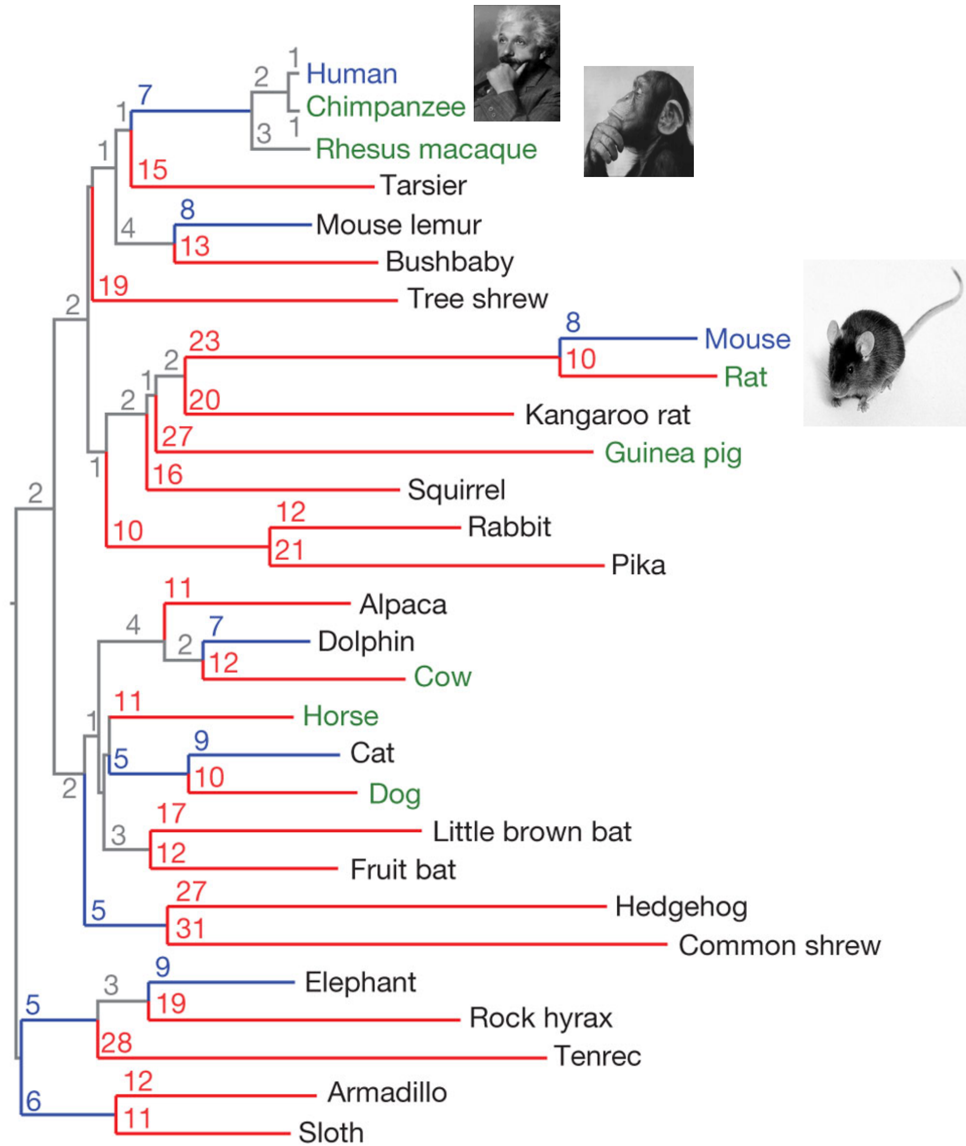
Human ADNP gene

Mouse ADNP gene

Topics in sequence alignment

- Pairwise alignment
  - Global alignment
  - Local alignment
- Multiple sequence alignment
- Scores and substitution matrices
- Practical algorithms for sequence alignment
  - BLAST
  - Progressive multiple alignment

How are these organisms related?



Toh et al, Nature, 2011

Topics in phylogenetic trees

- Reconstructing Phylogenetic trees
  - distance-based approaches
  - probabilistic methods
  - parsimony methods
- Inferring ancestral sequences
- Felsenstein's algorithm
- Neighbor Joining
- UPGMA

CCACACCACACCCACACACCCACACACCACACCACACACCACACCACACCCACACACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCTGGCCAACC
TGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTACCCTGTCCCATTCAACCATACCACTCCGAACCACCATCCATCCCTCTACTTACTACCACTCACCCAC
CGTTACCCTCCAATTACCCATATCCAACCCACTGCCACTTACCCTACCATTACCCTACCATCCACCATGACCTACTCACCATACTGTTCTTCTACCCACCATATTGAAA
CGCTAACAAATGATCGTAAATAACACACACGTGCTTACCCTACCACTTTATACCACCACCACATGCCATACTCACCCTCACTTGTATACTGATTTTACGTACGCACA
CGGATGCTACAGTATATACCATCTCAAACTTACCCTACTCTCAGATTCCACTTCACTCCATGGCCCATCTCTCACTGAATCAGTACCAAATGCACTCACATCATTATG
CACGGCACTTGCCTCAGCGGTCTATACCCTGTGCCATTTACCCATAACGCCCATCATTATCCACATTTTGATATCTATATCTCATTCGGCGGTCCCAAATATTGTATA
ACTGCCCTTAATACATACGTTATACCACTTTTGCACCATATACTTACCACTCCATTTATATACACTTATGTCAATATTACAGAAAAATCCCCACAAAAATCACCTAAA
CATAAAAATATTCTACTTTTCAACAATAATACATAAACATATTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGC
TTGAACGGATGCTATTTCAGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGG
TAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCATATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGT
ATGTGACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAA
TATAACTTATCAGCGGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATT
TCTAGTTACAGTTACACAAAAAACTATGCCAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAAC
TCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAAGCAGCGACTCATTTTTAT
TTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTCTACAATAGTGTAGAA
GTTTCTTTCTTATGTTCATCGTA … ATCTACGGTATTTATATC

Where are the genes in this genome?

ATCAAAAAAAAGTAGTTTTTTT … TTGCGATAGTGTAGATA
CCGTCCTTGGATAGAGCACTG … GAGCAATACCGGTCAAC
ATGGTGGTGAAGTCACCGTAGTTGAAAACGGCTTCAGCAACTTCGACTGGGTAGGTTTCAGTTGGGTGGGCGGCTTGGAACATGTAGTATTGGGCTAAGTGAGC
TCTGATATCAGAGACGTAGACACCCAATTCCACCAAGTTGACTCTTTCGTCAGATTGAGCTAGAGTGGTGGTTGCAGAAGCAGTAGCAGCGATGGCAGCGACAC
CAGCGGCGATTGAAGTTAATTTGACCATTGTATTTGTTTTGTTTGTTAGTGCTGATATAAGCTTAACAGGAAAGGAAAGAATAAAGACATATTCTCAAAGGCATAT
AGTTGAAGCAGCTCTATTTATACCCATTCCCTCATGGGTTGTTGCTATTTAAACGATCGCTGACTGGCACCAGTTCCTCATCAAATATTCTCTATATCTCATCTTTCA
CACAATCTCATTATCTCTATGGAGATGCTCTTGTTTCTGAACGAATCATAAATCTTTCATAGGTTTCGTATGTGGAGTACTGTTTTATGGCGCTTATGTGTATTCGTA
TGCGCAGAATGTGGGAATGCCAATTATAGGGGTGCCGAGGTGCCTTATAAAACCCTTTTCTGTGCCTGTGACATTTCCTTTTTCGGTCAAAAAGAATATCCGAATT
TTAGATTTGGACCCTCGTACAGAAGCTTATTGTCTAAGCCTGAATTCAGTCTGCTTTAAACGGCTTCCGCGGAGGAAATATTTCCATCTCTTGAATTCGTACAACAT
TAAACGTGTGTTGGGAGTCGTATACTGTTAGGGTCTGTAAACTTGTGAACTCTCGGCAAATGCCTTGGTGCAATTACGTAATTTTAGCCGCTGAGAAGCGGATGG
TAATGAGACAAGTTGATATCAAACAGATACATATTTAAAAGAGGGTACCGCTAATTTAGCAGGGCAGTATTATTGTAGTTTGATATGTACGGCTAACTGAACCTA
AGTAGGGATATGAGAGTAAGAACGTTCGGCTACTCTTCTTTCTAAGTGGGATTTTTCTTAATCCTTGGATTCTTAAAAGGTTATTAAAGTTCCGCACAAAGAACGC
TTGGAAATCGCATTCATCAAAGAACAACTCTTCGTTTTCCAAACAATCTTCCCGAAAAAGTAGCCGTTCATTTCCCTTCCGATTTCATTCCTAGACTGCCAAATTTTT
CTTGCTCATTTATAATGATTGATAAGAATTGTATTTGTGTCCCATTCTCGTAGATAAAATTCTTGGATGTTAAAAAATTAAAGGGACTATATCTAGTCAAGACGATA
CTGTCAGTAGCAGCGATGGCAGCGTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGCTTGAACGGATGCTATTT
CAGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGCTCAAACTCAT
GCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCATATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGTATGTGACACTACTCATA
CGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAATATAACTTATCAGCGGC
GTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATTTCTAGTTACAGTTACAC
AAAAAACTATGCCAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTAACTCGCACTGCCCTGATCT
GCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAAGCAGCGACTCATTTTTATTTAAGGACAAAGGTTG

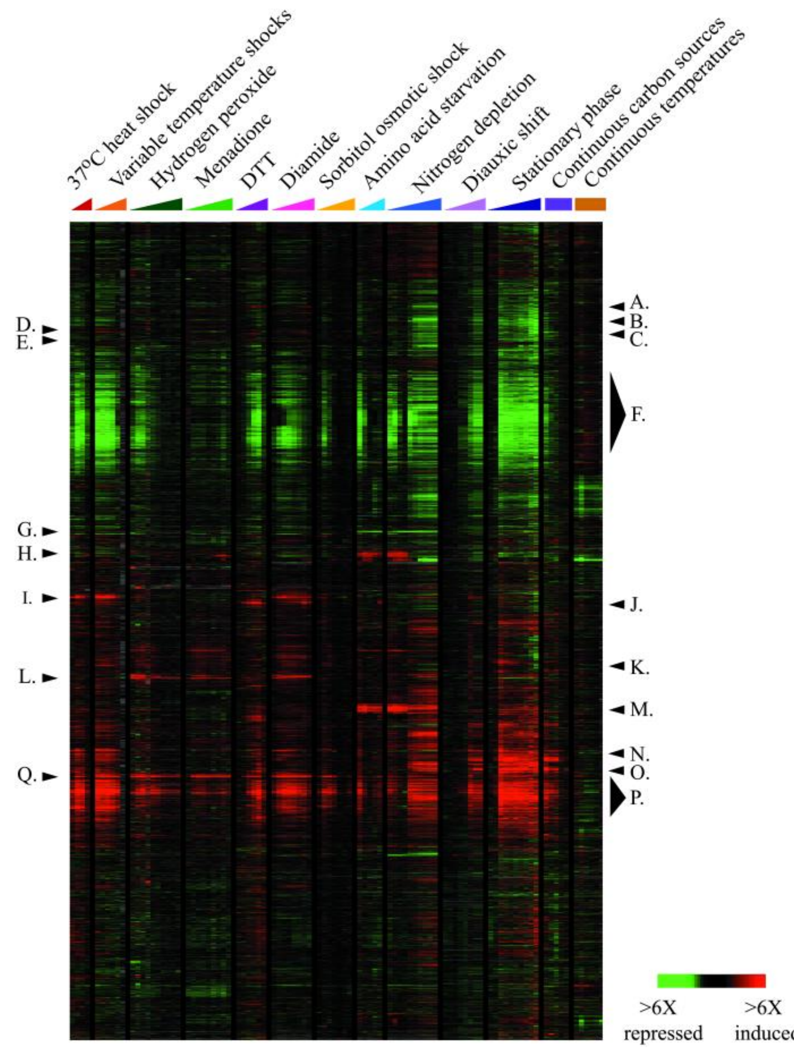
CCACACCACACCCACACACCCACACACCACACCACACACCACACCACACCCACACACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCTGGCCAACC
TGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTACCCTGTCCCATTCAACCATACCACTCCGAACCACCATCCATCCCTCTACTTACTACCACTCACCCAC
CGTTACCCTCCAATTACCCATATCCAACCCACTGCCACTTACCCTACCATTACCCTACCATCCACC**ATGACCTACTCACCATACTGTTCTTCTACCCACCATATTGAA**
ACGCTAACAAATGATCGTAAATAACACACACGTGCTTACCCTACCACTTTATACCACCACATGCCATACTCACCCTCACTTGTATACTGATTTTACGTACGC
ACACGGATGCTACAGTATATACCATCTCAAACTTACCCTACTCTCAGATTCCACTTCACTCCATGGCCCATCTCTCACTGAATCAGTACCAAATGCACTCACATCA
TTATGCACGGCACTTGCCTCAGCGGTCTATACCCTGTGCCATTTACCCATAACGCCCATCATTATCCACATTTTGATATCTATATCTCATTCGGCGGTCCCAAATAT
TGTATAACTGCCCTTAATACATACGTTATACCACTTTTGCACCATATACTTACCACTCCATTTATATACACTTATGTCAATATTACAGAAAATCCCCACAAAAATCA
CCTAAACATAAAAATATTCTACTTTTCAACAATAATACATAAACATATTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCA
ATTTGCTTGAACGGATGCTATTTCAGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATA
ATACGGTAGTGGCTCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCATATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTA
ATCTTGTATGTGACACTACTCATACGAAGGGACTATATCT Protein coding sequence ATTTAATAGGATCTATAACGAAATGTCAAATAATTTT
ACGGTAATATAACTTATCAGCGGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCA
GCTTTATTTCTAGTTACAGTTACACAAAAAACTATGCCAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGA
CTTGTAACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAAGCAGCGACT
CATTTTTATTTAAGGACAAAGGTTGCGAAGCCGCACATTTCCAATTTCATTGTTGTTTATTGGACATACACTGTTAGCTTTATTACCGTCCACGTTTTTCTACAATA
GTGTAGAAGTTTCTTTCTTATGTTCATCGTATTCATAAAATGCTTCACGAACACCGTCATTGATCAAATAGGTCTATAATATTAATATACATTTATATAATCTACGGT
ATTTATATCATCAAAAAAAAGTAGTTTTTTATTTTATTTTGTTCGTTAATTTTCAATTTCTATGGAAACCCGTTCGTAAAATTGGCGTTTGTCTCTAGTTTGCGATAG
TGTAGATACCGTCCTTGGATAGAGCACTGGAGATGGCTGGCTTTAATCTGCTGGAGTACCATGGAACACCGGTGATCATTCTGGTCACTTGGTCTGGAGCAATAC
CGGTCAACATGGTGGTGAAGTCACCGTAGTTGAAAACGGCTTCAGCAACTTCGACTGGGTAGGTTTCAGTTGGGTGGGCGGCTTGGAACATGTAGTATTGGGCT
AAGTGAGCTCTGATATCAGAGACGTAGACACCCAATTCCACCAAGTTGACTCTTTCGTCAGATTGAGCTAGAGTGGTGGTTGCAGAAGCAGTAGCAGCGATGGC
AGCGACACCAGCGGCGATTGAAGTTAATTTGACCATTGTATTTGTTTTGTTTGTTAGTGCTGATATAAGCTTAACAGGAAAGGAAAGAATAAAGACATATTCTCA
AAGGCATATAGTTGAAGCAGCTCTATTTATACCCATTCCCTCATGGGTTGTTGCTATTTAAACGATCGCTGACTGGCACCAGTTCCTCATCAAATATTCTCTATATCT
CATCTTTCACACAATCTCATTATCTCTATGGAGATGCTCTTGTTTCTGAACGAATCATAAATCTTTCATAGGTTTCGTATGTGGAGTACTGTTTTATGGCGCTTATGT
GTATTCGTATGCGCAGAATGTGGGAATGCCAATTATAGGGGTGCCGAGGTGCCTTATAAAACCCTTTTCTGTGCCTGTGACATTTCCTTTTTCGGTCAAAAAGAAT
ATCCGAATTTTAGATTTGGACCCTCGTACAGAAGCTTATTGTCTAAGCCTGAATTCAGTCTGCTTTAAACGGCTTCCGCGGAGGAAATATTTCCATCTCTTGAATTC
GTACAACATTAAACGTGTGTTGGGAGTCGTATACTGTTAGGGTCTGTAAACTTGTGAACTCTCGGCAAATGCCTTGGTGCAATTACGTAATTTTAGCCGCTGAGA
AGCGGATGGTAATGAGACAAGTTGATATCAAACAGATACATATTTAAAAGAGGGTACCGCTAATTTAGCAGGGCAGTATTATTGTAGTTTGATATGTACGGCTAA
CTGAACCTAAGTAGGGATATGAGAGTAAGAACGTTCGGCTACTCTTCTTTCTAAGTGGG Protein coding sequence TATTAAAGTTCCGCAC
AAAGAACGCTTGGAAATCGCATTCATCAAAGAACAACTCTTCGTTTTCCAAACAATCTTC ATTTCATTCCTAGACTG
CCAAATTTTTCTTGCTCATTTATAATGATTGATAAGAATTGTATTTGTGTCCCATTCTCGTAGATAAAATTCTTGGATGTTAAAAAATTAAAGGACTATATCTAGTC
AAGACGATACTGTCAGTAGCAGCGATGGCAGCGTGGCTTGTGGTAGCAACACTATCATGGTATCACTAACGTAAAAGTTCCTCAATATTGCAATTTGCTTGAACG
GATGCTATTTCAGAATATTTCGTACTTACACAGGCCATACATTAGAATAATATGTCACATCACTGTCGTAACACTCTTTATTCACCGAGCAATAATACGGTAGTGGC
TCAAACTCATGCGGGTGCTATGATACAATTATATCTTATTTCCATTCCCAT**ATGCTAACCGCAATATCCTAAAAGCATAACTGATGCATCTTTAATCTTGTATGTG**
ACACTACTCATACGAAGGGACTATATCTAGTCAAGACGATACTGTGATAGGTACGTTATTTAATAGGATCTATAACGAAATGTCAAATAATTTTACGGTAATA
TAACTTATCAGCGGCGTATACTAAAACGGACGTTACGATATTGTCTCACTTCATCTTACCACCCTCTATCTTATTGCTGATAGAACACTAACCCCTCAGCTTTATT
TCTAGTTACAGTTACACAAAAAACTATGCCAACCCAGAAATCTTGATATTTTACGTGTCAAAAAATGAGGGTCTCTAAATGAGAGTTTGGTACCATGACTTGTA
ACTCGCACTGCCCTGATCTGCAATCTTGTTCTTAGAAGTGACGCATATTCTATACGGCCCGACGCGACGCGCCAAAAAATGAAAAACGAAGCAGCGACTCATTTTT

Topics in sequence annotation

- Markov chains
- Hidden Markov models
- Inference and Parameter estimation
  - Forward, Backward, Viterbi algorithms
- Applications to genome segmentation

How do cells function under different conditions?

- Measure mRNA/protein levels under different environmental conditions
- Compare levels of genes under different conditions



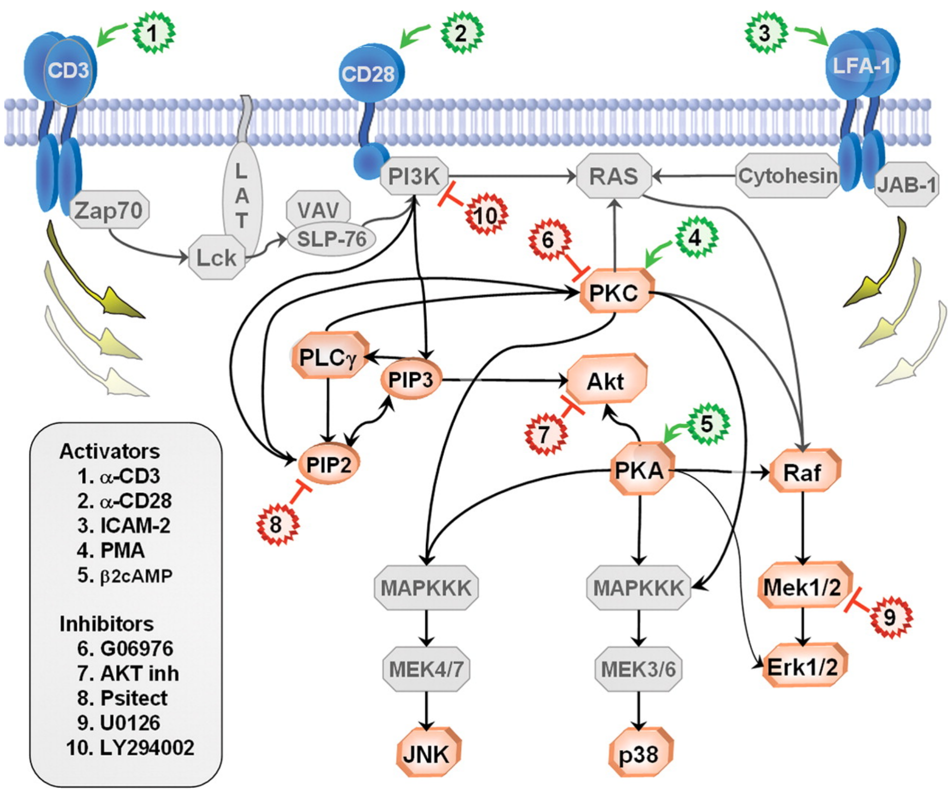
Topics in data analysis from high-throughput experiments

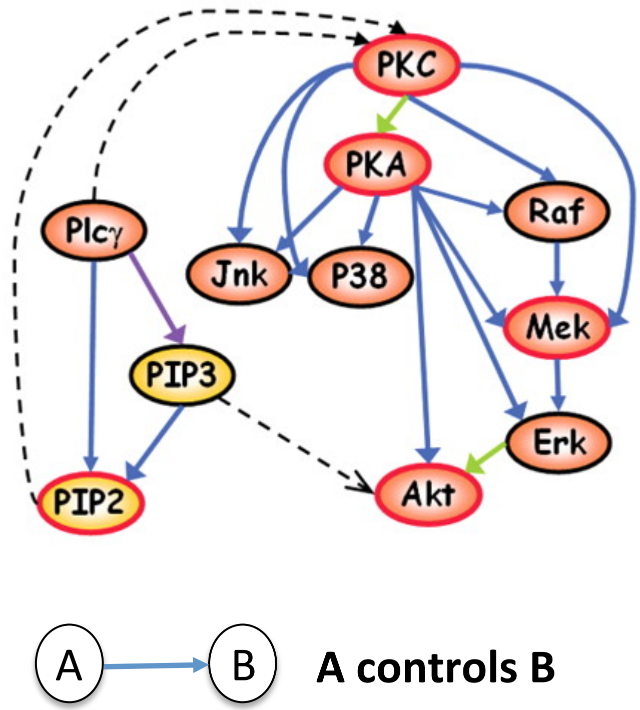
- Clustering algorithms
  - hierarchical clustering
  - k-means clustering
  - EM-based clustering
- Interpretation of clusters
- Evaluation of clusters

How do molecular entities interact within a cell?

Interactions within a cell

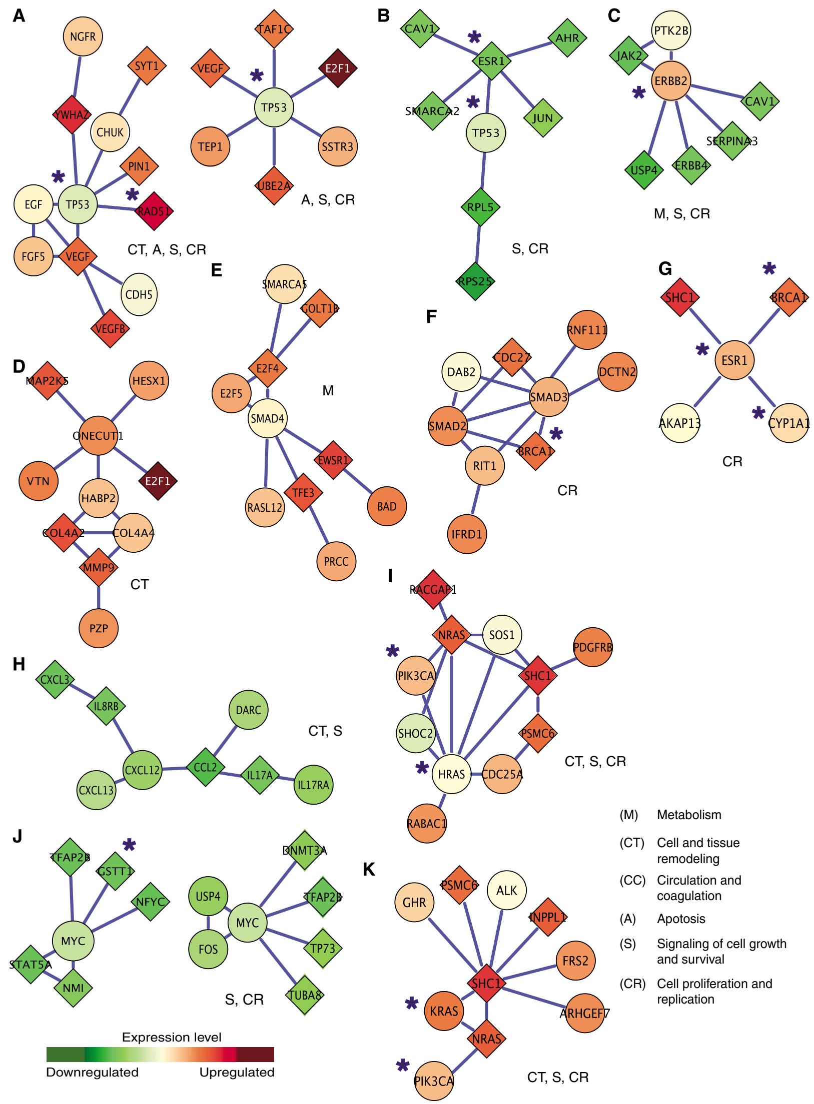
Network model





What networks get perturbed in a disease?

Subnetworks of genes predictive of cancer prognosis



Chuang et al., MSB 2007

Topics in network modeling

- Different types of biological networks
- Probabilistic graphical models for representing networks
- Algorithms of network inference
- Evaluating inferred networks
- Analysis of inferred networks

The Short-term Plan

- Tuesday (9/12)
  - “Molecular Biology 101” lecture
  - Optional for molecular biology students
- Thursday (9/14)
  - start on “Sequence Assembly”

Reminder: Reading assignment for Tuesday

- ***Life and Its Molecules A Brief Introduction*** by Lawrence Hunter
  - http://www.biostat.wisc.edu/bmi576/papers/hunter04.pdf

Goals for today

- Administrivia
- Course Overview
- **Short survey of interests/background**