

Pursuing Graduate Studies (in Bioinformatics)



Chau-Wen Tseng

**Department of Computer Science
University of Maryland, College Park**

What You Learn

- ◆ **Bachelor (B.A.)**
 - **Basic principles & problem-solving techniques**
 - **Apply techniques to solve classroom problems**
- ◆ **Master (M.S.)**
 - **State-of-the-art principles & techniques**
 - **Apply techniques to solve semi-real problems**
- ◆ **Doctorate (Ph.D.)**
 - **Independently discover new principles & techniques**
 - **Apply techniques to solve cutting-edge problems**

What Type of Job

- ◆ **Bachelor (B.A.)**
 - **Greatest variety / number of jobs**
 - **May become boring over time**
 - **Less control over responsibilities / goals**
- ◆ **Master (M.S.)**
 - **Large variety / number of jobs**
 - **Greater potential for management**
- ◆ **Doctorate (Ph.D.)**
 - **Limited variety / number of jobs**
 - **Research & teaching positions**
 - **More ability to set your own responsibilities & goals**

Graduate Studies Tradeoffs

◆ Advantages

- More interesting job opportunities
- More flexibility & control over job / career
- Possibly higher income
- Prestige

◆ Disadvantages

- Many additional years of school (5-8+ for Ph.D.)
- May not recover lost income
- Limited career options (especially with Ph.D.)
- Stressful on family & relationships
- Not suitable for everyone
 - Only ~50% graduate students complete Ph.D.

What To Study?

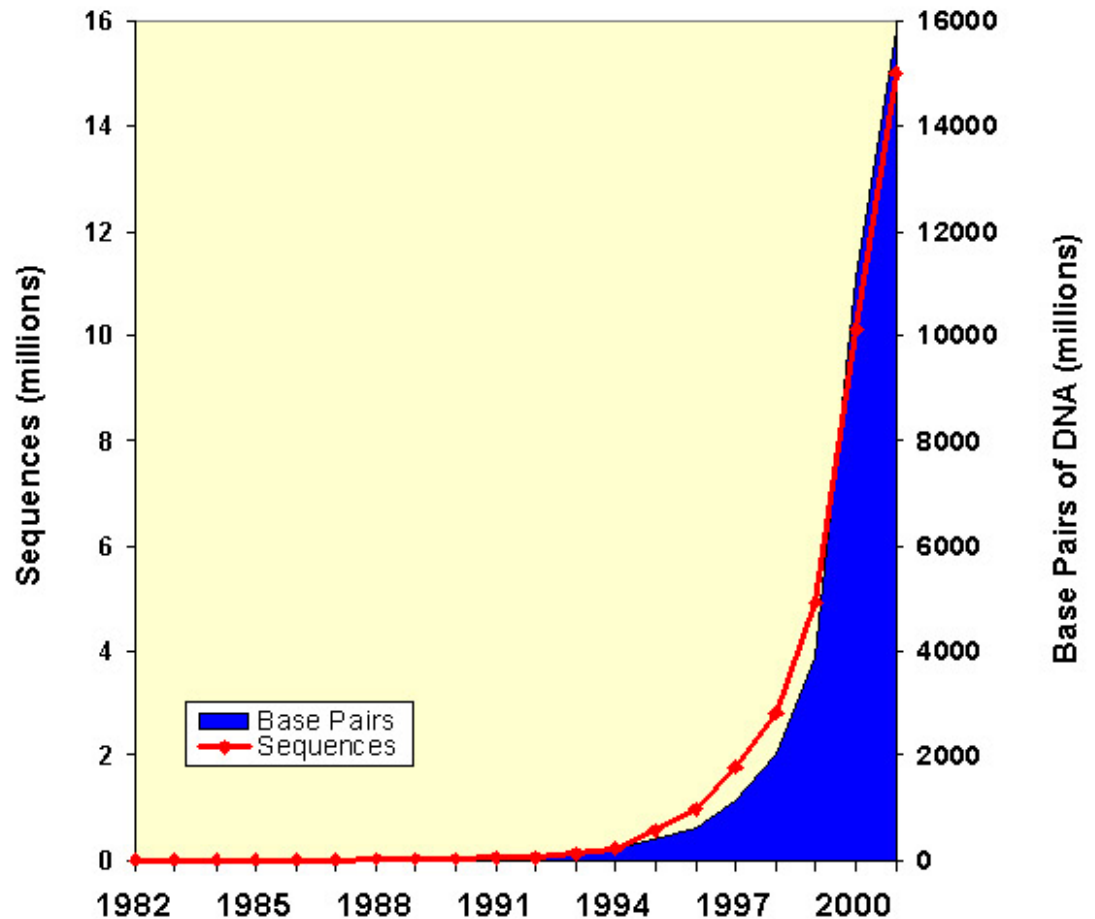
◆ Bioinformatics

- The creation and development of advanced information and computational techniques for solving problems biology

◆ Motivation

- Techniques for inexpensively measuring large amounts of biological data

Growth of GenBank



Graduate Studies in Bioinformatics

- ◆ **Exciting area**
 - **Biologists & doctors taking advantage of new technology & computing power**
 - **Area with lots of potential**

- ◆ **Requires wide range of skills**
 - **Molecular biology**
 - **Statistics**
 - **Algorithms**
 - **Artificial intelligence (machine learning)**
 - **Databases**

Some Warnings

◆ Bioinformatics jobs

- Highly specialized position
- Area has been (over) hyped by media
- Not many actual positions available in industry
- Monster.com (Oct. 19, 2004)

Job Listing Keyword	Listings in US	Listings in MD (Montgomery County)
Computer Security	5000+	199
Software Engineering	5000+	196
Bioinformatics	119	9 (4 biostatistics)

Bioinformatics Topics

- ◆ **Sequence alignments**
 - Find similarity between DNA / protein (amino acid) sequences
- ◆ **Genome assembly**
 - Combining genomic fragments to form whole genome
- ◆ **Gene identification & annotation**
 - Identify and classify genes on the genome
- ◆ **Microarrays & gene expression analysis**
 - Use DNA microarray (gene chip) to measure mRNA
- ◆ **Protein folding**
 - Compute 3D protein structure ↔ protein sequence
- ◆ **Phylogenetic analysis**
 - Find genetic relationships between sequences / species

Sequence Alignment

◆ Pairwise sequence alignment

- Mutual arrangement of sequences

G A T

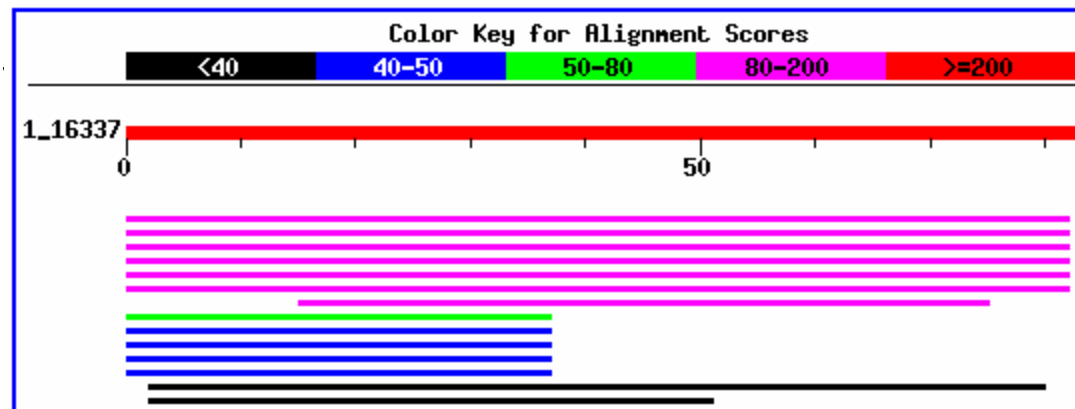
C A T C A

- Similar sequence → similar function?

G G T

C - T C A

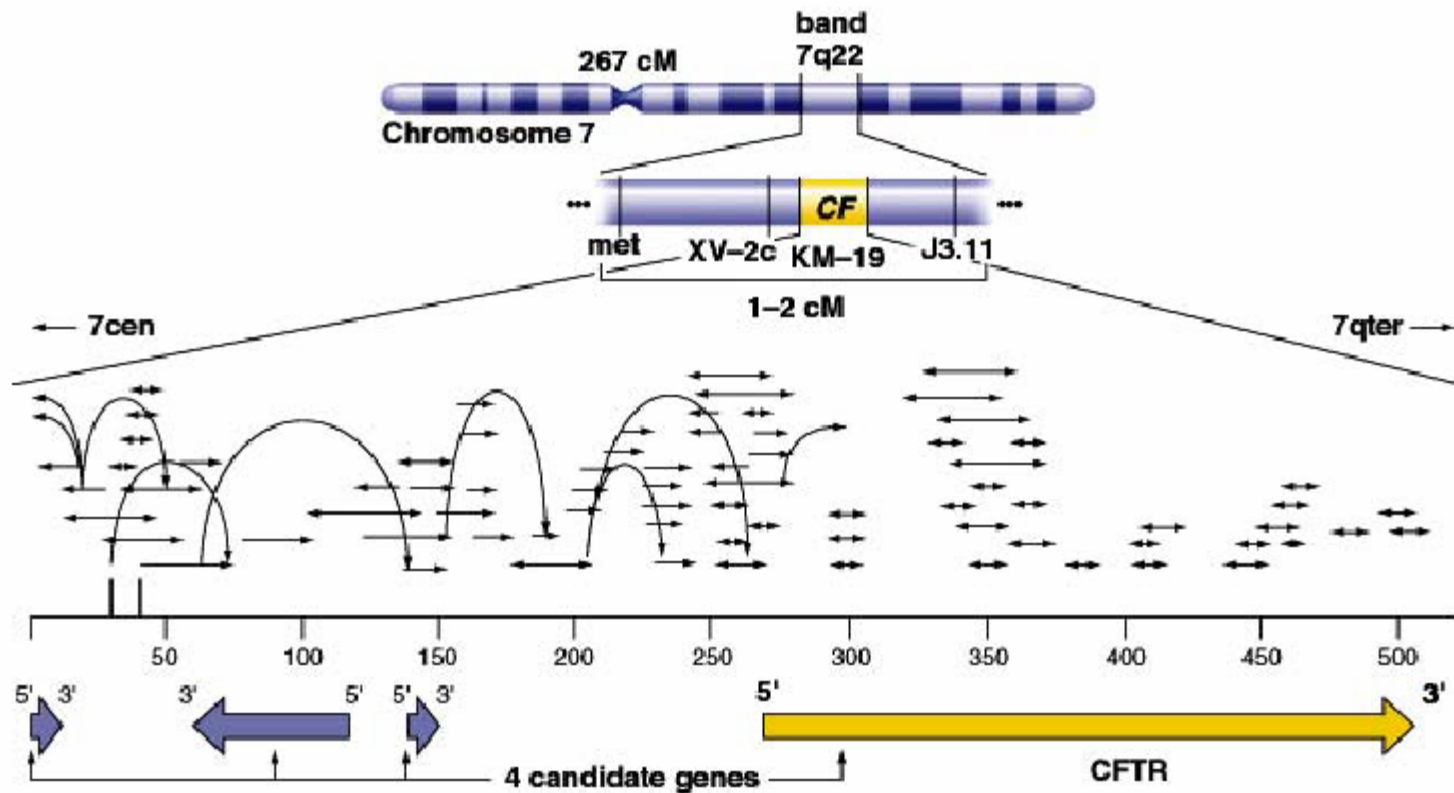
◆ Look for



Sequences producing significant alignments:	Score	E Value	
	(bits)		
gi 30584271 gb AAP36384.1 Homo sapiens DNA fragmentation f...	133	8e-31	
gi 4758148 ref NP_004392.1 DNA fragmentation factor, 45 kD...	133	8e-31	L
gi 14043461 gb AAH07721.1 DNA fragmentation factor, 45 kD,...	131	3e-30	L
gi 24158730 pdb 1IYR A Chain A, Nmr Structure Ensemble Of D...	122	2e-27	S

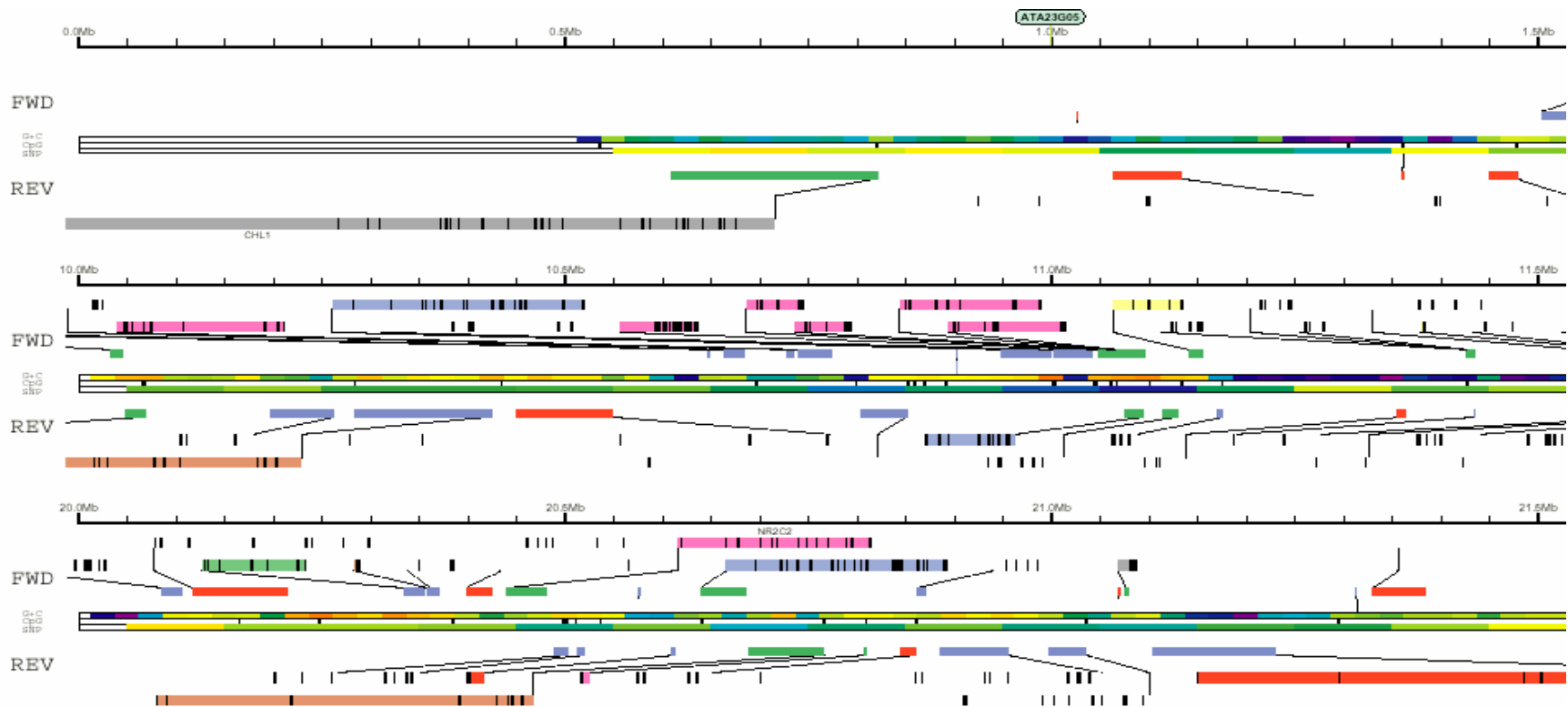
Sequence Assembly

- ◆ Combining short mRNA / DNA sequences (reads)
- ◆ Produce full transcript / genome



Gene Prediction

◆ Predicting expressed DNA (genes)



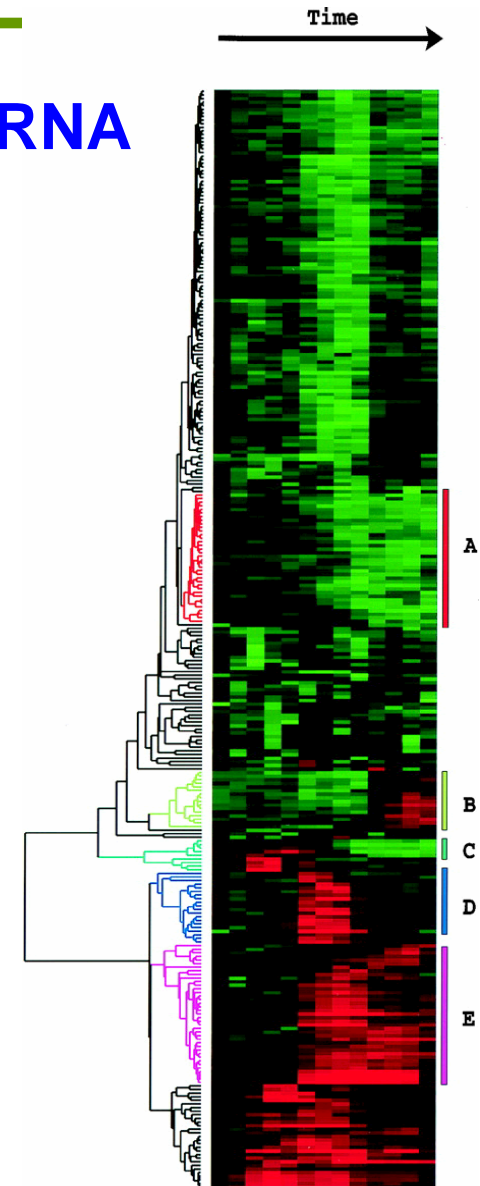
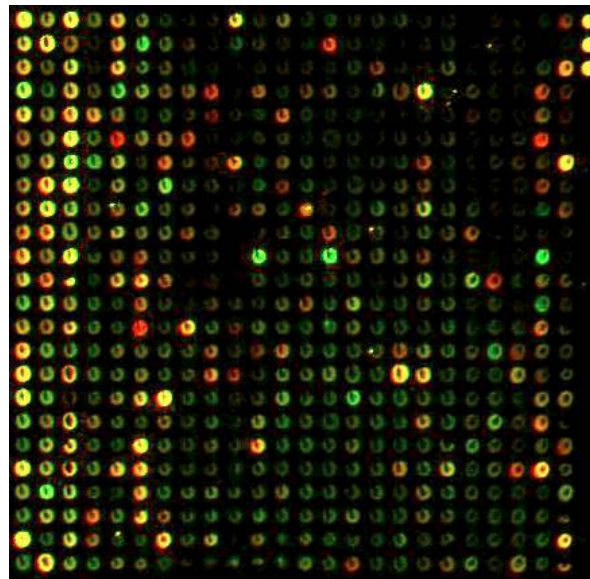
Measuring Gene Expression & Regulation

- ◆ Detecting DNA (genes) expressed as mRNA



Affymetrix DNA
microarray

500,000 oligomers
in 1.28 cm^2



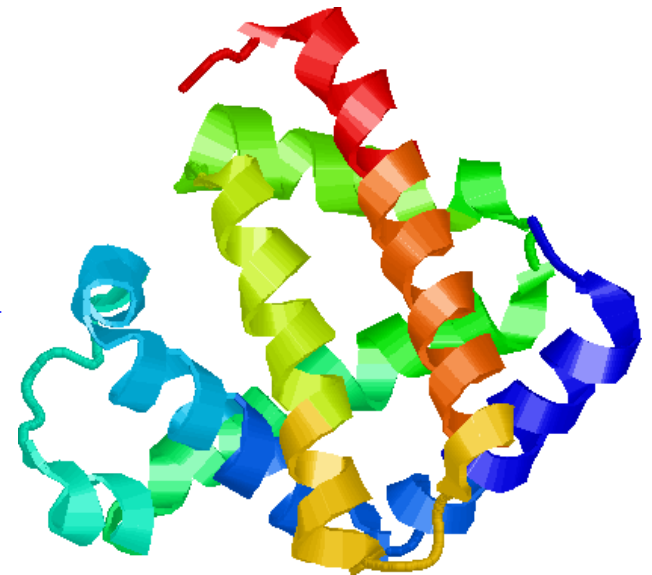
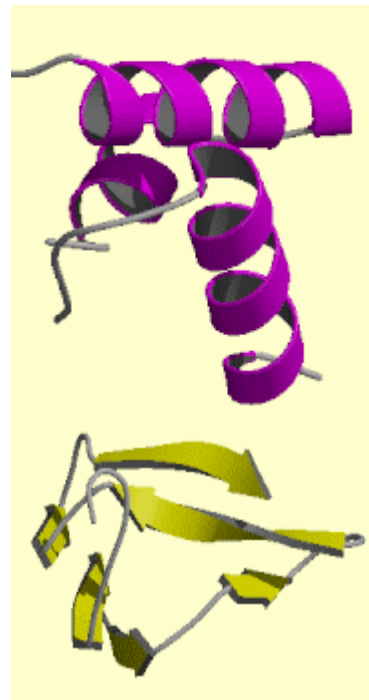
Protein Structure Prediction

- ◆ Predict protein 3D structure from (amino acid) sequence
- ◆ Predict interactions between proteins and ligands

5' atgcccaagctgaat ... 3'

atg ccc aag ctg aat ...

M P K L N ...



Phylogenetic Analysis

- ◆ Study of evolutionary relationships
- ◆ Infer evolutionary relationship from shared features

seqA	N	•	F	L	S
seqB	N	•	F	-	S
seqC	N	K	Y	L	S
seqD	N	•	Y	L	S

