

# Bioinformatics and Genomics

MCDB 4520/5520; CSCI 4830/7000



"From the growth of the Internet through to the mapping of the human genome and our understanding of the human brain, the more we understand, the more there seems to be for us to explore."

--Martin Rees

Class: MWF 9:30-10:20 am

Location: JSCBB B115

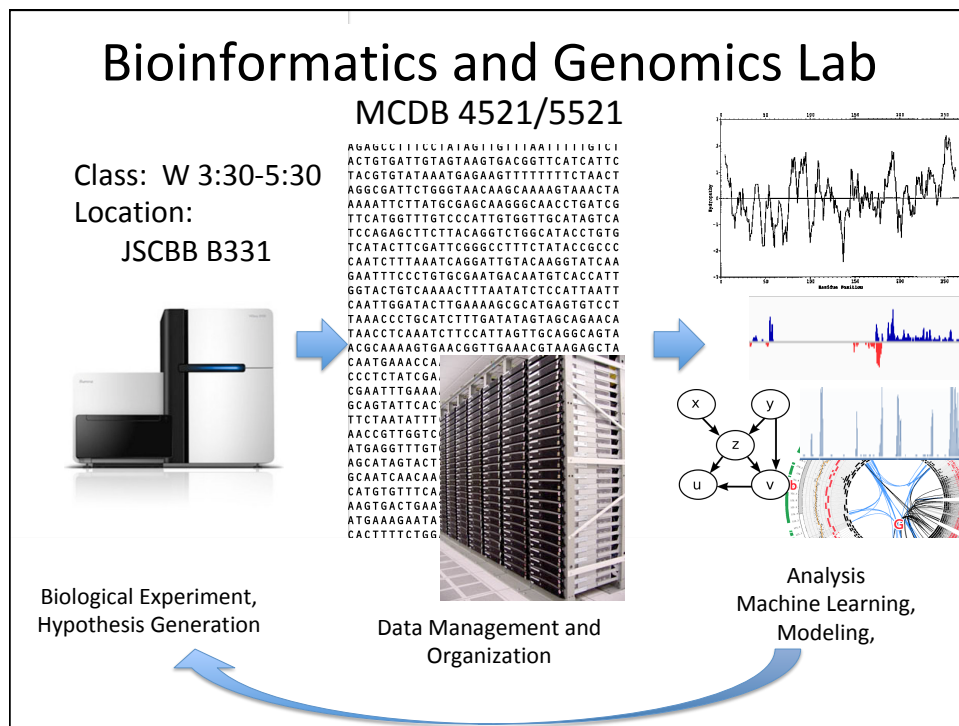
Dr. Robin Dowell

Office: JSCBB B416

Email: [robin.dowell@colorado.edu](mailto:robin.dowell@colorado.edu)

Office Hours: Tues 1-2 pm; or by appointment  
(just email me!)

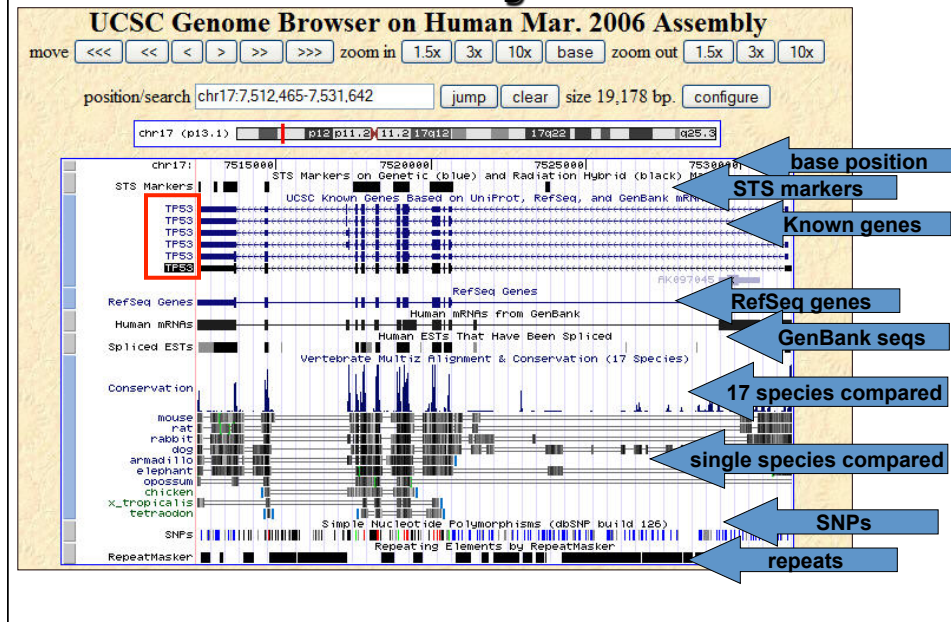
Course materials on Canvas!



## About me

- Degrees
  - B.S. in Computer Engineering
  - B.S. in Genetics
  - M.S. in Computer Science
  - D.Sc. in Biomedical Engineering
- Associate Professor, MCD Biology
- Co-founder, Arpeggio Biosciences

## Sample Genome Viewer image, TP53 region



## Learning Goals

- Have a general understanding of genomics.
- Understand basic computational thinking and how to utilize it to analyze big data.
- Recognize how experiments influence data analysis.
- Identify assumptions in an algorithm and how they influence your analysis.
- Determine the expectation of an analysis in order to determine if something is statistically significant.



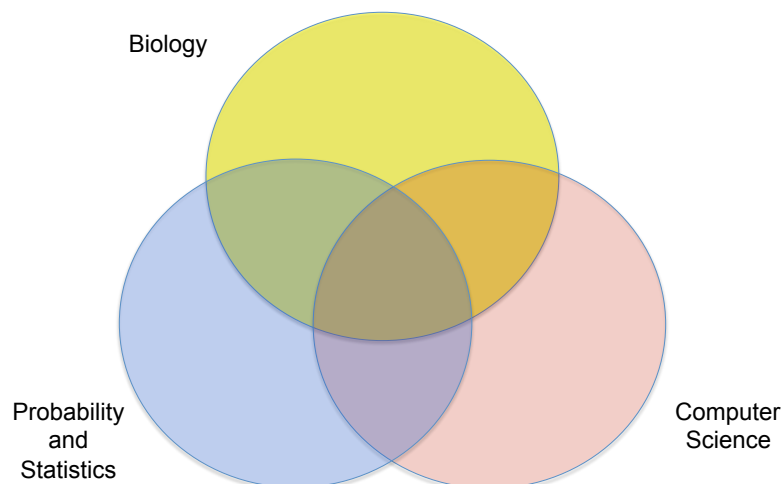
Please.

What is Bioinformatics?

## What Is Bioinformatics?

- **Bioinformatics** is the unified discipline formed from the combination of biology, computer science, and statistics.
- **Bioinformatics** is the science of managing and analyzing biological data using advanced computing techniques.
- **Bioinformatics** seeks to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles can be discerned.

## What is Bioinformatics?



## Basic principles

1. Begin with an end in mind. (Question driven)
2. Know your data.
3. Do science (e.g. experiments with data).
4. It is never perfect.
5. Assess whether what you find is interesting.

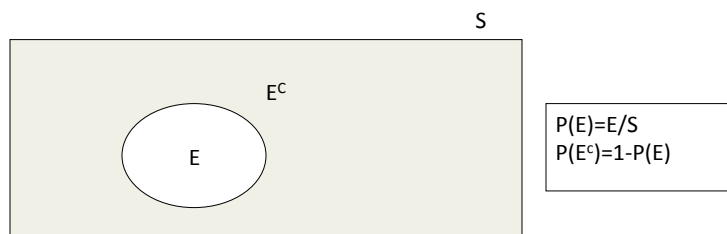
Assignment #1: Read Sean Eddy's Antedisciplinary Science (available on Canvas) before Monday Jan 22.



“The most important questions of life are  
...really only problems of probability.”

-Pierre Simon, Marquis de Laplace (1749-1827)

Probability: Computing the chance of a particular outcome of an experiment





“There Are Three Kinds of Lies:  
Lies, Damned Lies, and Statistics”

-Benjamin Disraeli (1804-1881)

Statistics: drawing inferences from the results  
of an experiment

Correctly understanding and applying statistics **REQUIRES**  
understanding the underlying probabilistic model.



“All models are wrong,  
but some are useful.”

-George Box

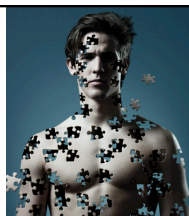




# What is Genomics?

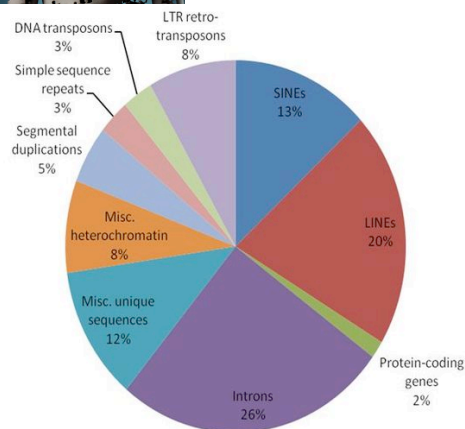
# What is Genomics?

“Book of Life” - NY Times (2001)  
 “Code of Life” – PBS (2002)  
 “Map of Life” – Science World (1999)  
 “Blueprint of Humanity” - BBC (2000)  
 “Instruction Set” – Bob Waterston (2003)  
 “Evolution’s Notebook” – Eric Lander (2002)



## What do we now know about a typical human genome?

- Nearly ALL genes have alternative isoforms
- 65-80% of the genome is transcribed
- ENCODE claims 80% of genome has some function



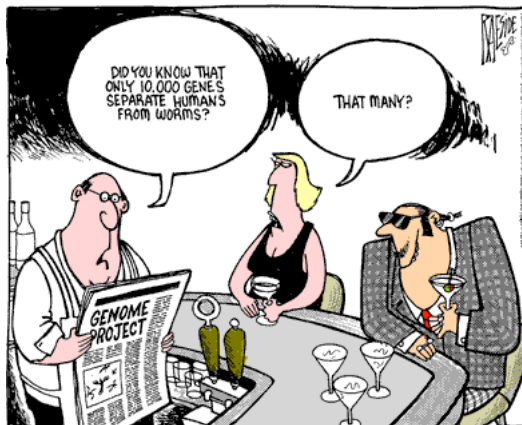
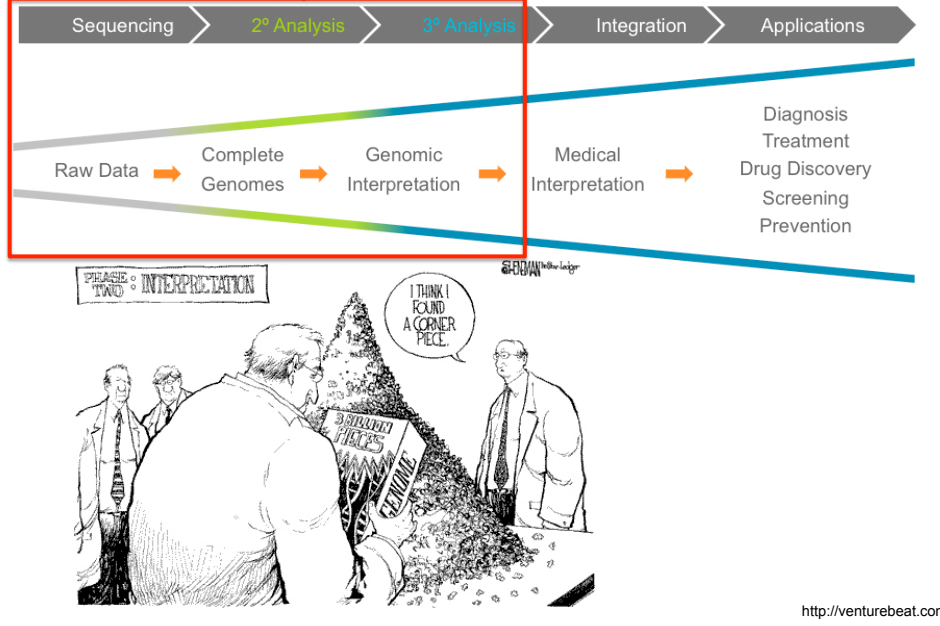
Components of the Human Genome

### COMPARE TWO INDIVIDUALS:

- 99.9% identical
- ~3 million variations (single nucleotide variations, insertions, deletions, copy number variations)

<http://www.technologyreview.com/featuredstory/422140/the-human-genome-a-decade-late>

## Genomics Landscape



On Friday we will discuss the basics of genomes and begin building our first computational model.