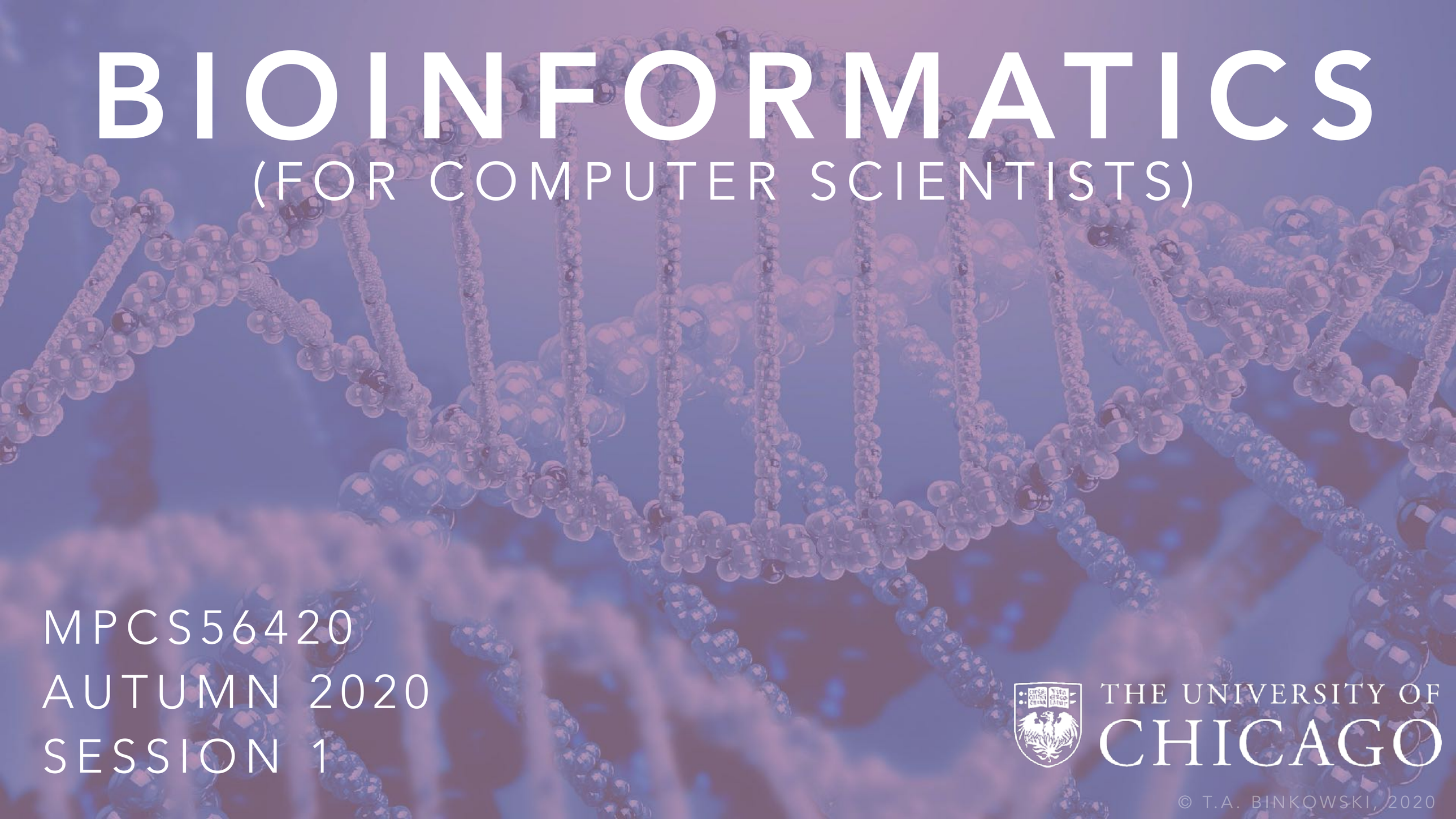
BIOINFORMATICS

(FOR COMPUTER SCIENTISTS)



MPCS56420
AUTUMN 2020
SESSION 1

THE UNIVERSITY OF CHICAGO


WELCOME TO
BIOINFORMATICS

COURSE LOGISTICS

- Prerequisites
  - Algorithms
  - Core Programming
- Consent of instructor
  - Requires departmental approval

COURSE LOGISTICS

- Flipped Classroom
  - Videos of lectures
  - Meet for discussions and office hours

COURSE LOGISTICS

- Regular Class Meeting
  - Thursday 5:30-6:30pm
- Lecture Video Recap/Office Hours
  - Monday 1:00 (See slack poll)
- Office hours
  - Tuesday 5:30 (Neal Conrad)
  - Monday 1:30pm (Andrew) (See slack poll)

Introduce upcoming weeks material
Introduce/update assignment
Open Q and A

Student presentations

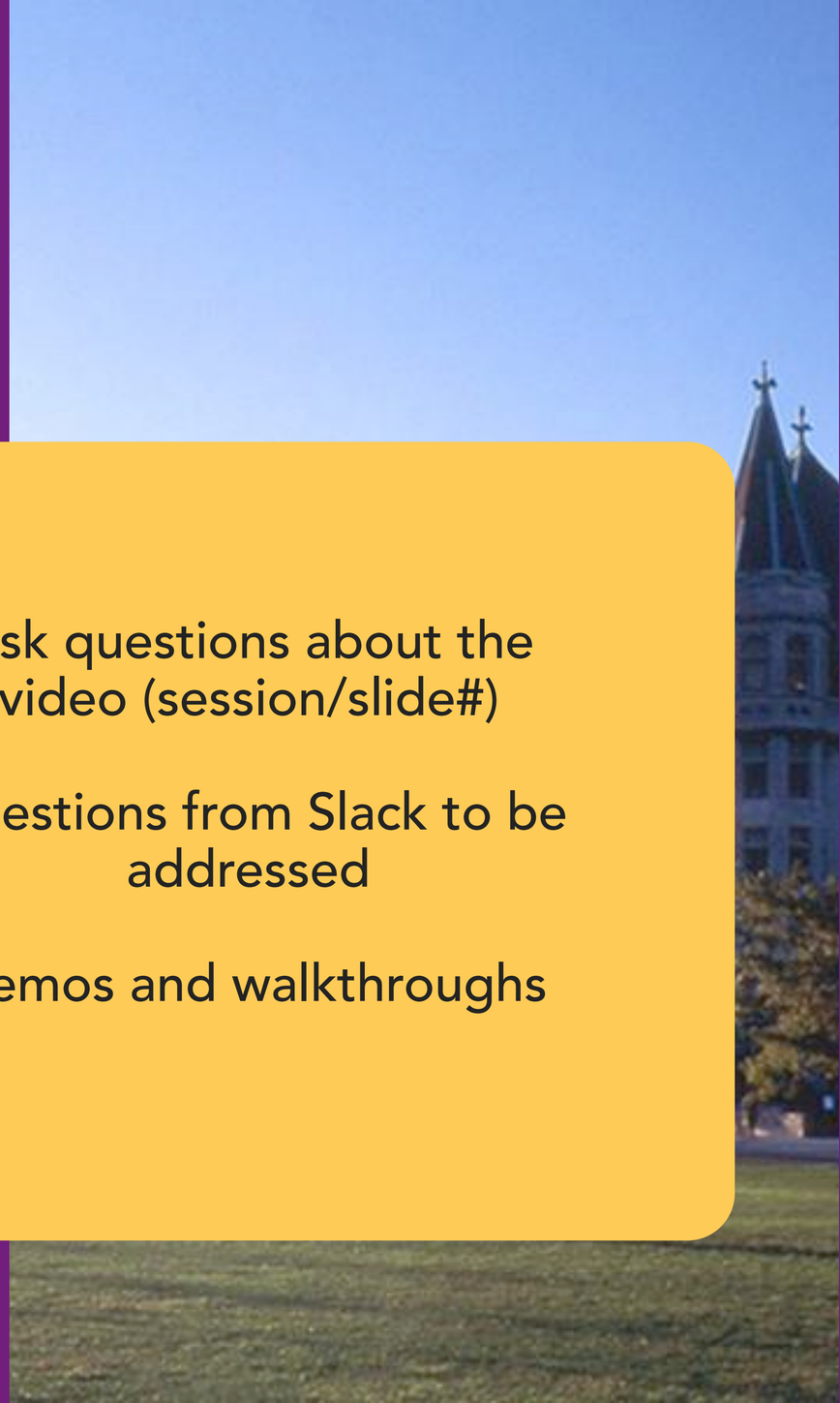
COURSE LOGISTICS

- Regular Class Meeting
  - Thursday 5:30-6:30pm
- Lecture Video Recap/Demos
  - Monday 1:00 (See slack poll)
- Office hours
  - Tuesday 5:30 (Neal Conrad)
  - Monday 1:30pm (Andrew) (See slack poll)

Ask questions about the video (session/slide#)

Questions from Slack to be addressed

Demos and walkthroughs



COURSE LOGISTICS

- Regular Class Meeting
  - Thursday 5:30-6:30pm
- Lecture Video Recap/Demos
  - Monday 1:00 (See slack poll)
- Office hours
  - Tuesday 5:30 (Neal Conrad)
  - Monday 1:30pm (Andrew) (See slack poll)

Please try and schedule a specific time to meet for office hours.

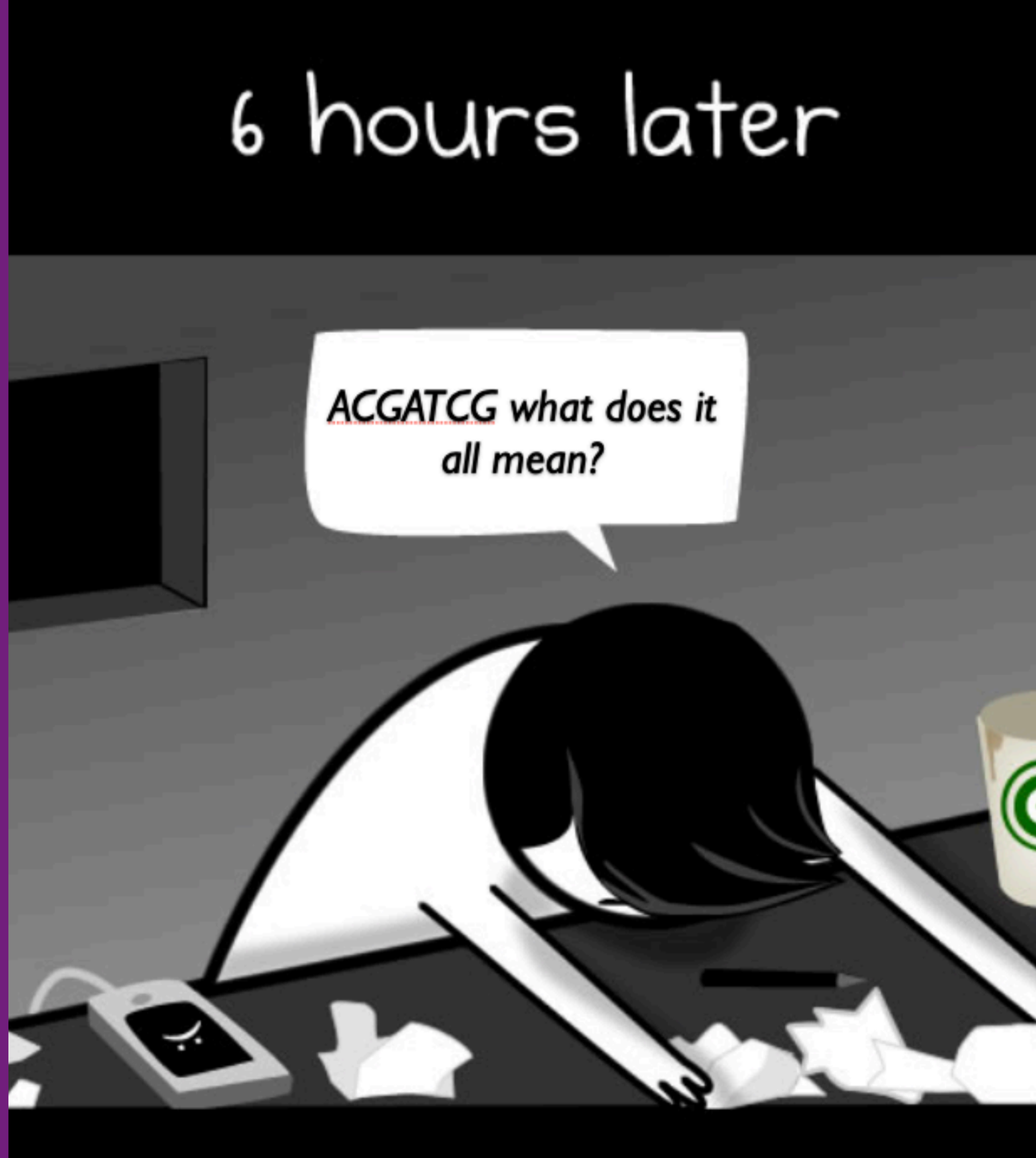
#office-hours

COURSE LOGISTICS

- No biology required
  - Weekly “need to know” basis
  - Class will adjust for student backgrounds
  - Online resources for learning biology
    - Khan Academy, EdX, Coursera, Udacity, YouTube

COURSE LOGISTICS

- Biology is a learn-as-you-go
- Researchers study one gene for a lifetime



COURSE DESIGN

INSTRUCTORS

- Andrew
  - The University of Chicago
    - Center for Structural Genomics of Infectious Diseases
    - 10 years teaching MPCS Program
  - ANL
    - Midwest Center for Structural Genomics, Leadership Computing Facility
    - Computation Institute

COURSE DESIGN

INSTRUCTORS

- Research programs
  - Biomolecular modeling; Drug Discovery
  - Protein Crystallography
  - HPC for bioinformatics pipelines
  - STEM Education (K-12)

COURSE DESIGN

INSTRUCTORS

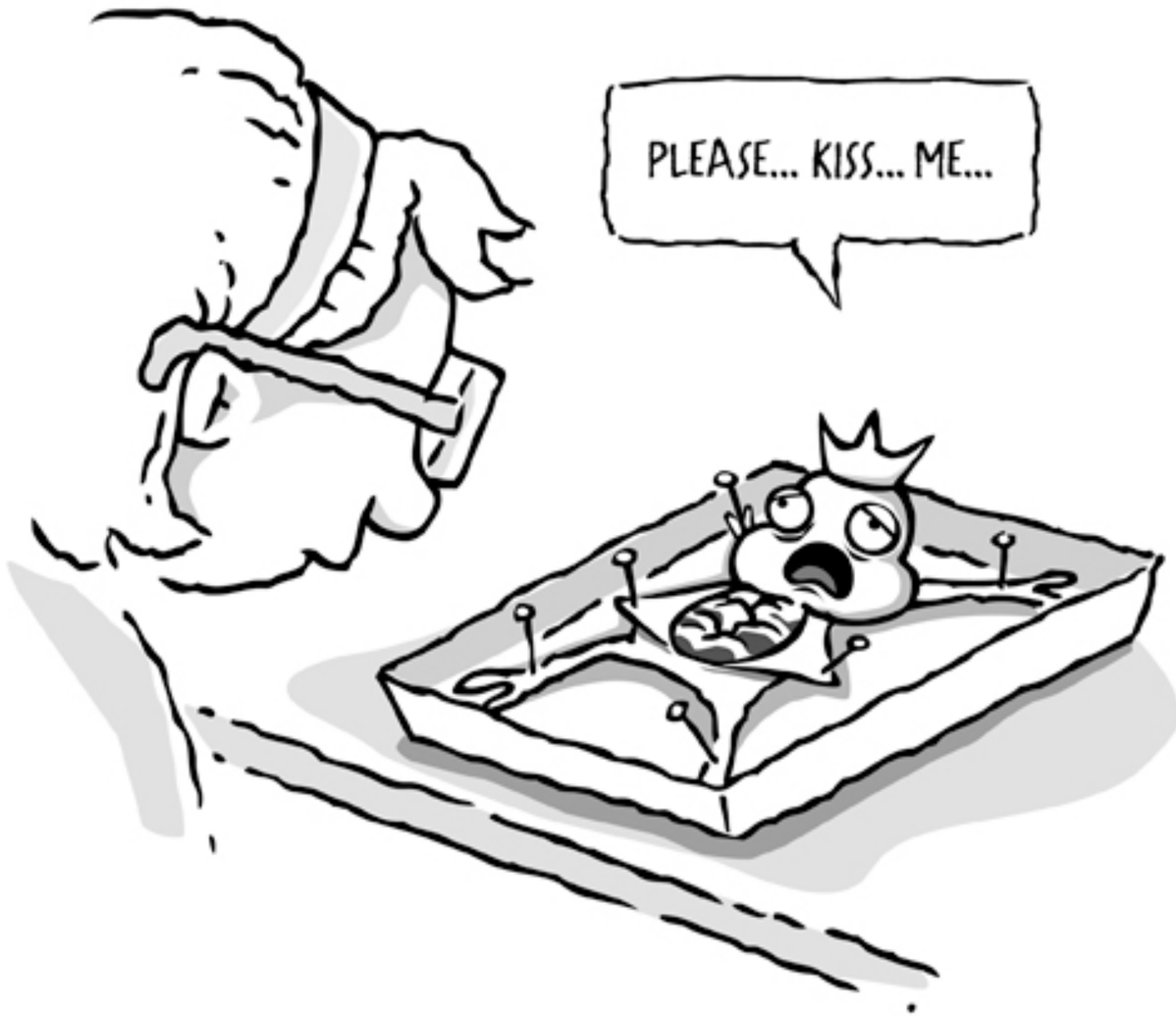
- Neal Conrad
  - Software Engineering Associate 2 at Argonne National Laboratory's Data Science and Learning Division
  - Graduated from MPCS in March, 2019
  - His work includes designing and building large-scale web applications, data visualization tools, and open-source libraries to support scientific research in bioinformatics.

Course Design

COURSE DESIGN

CHALLENGES

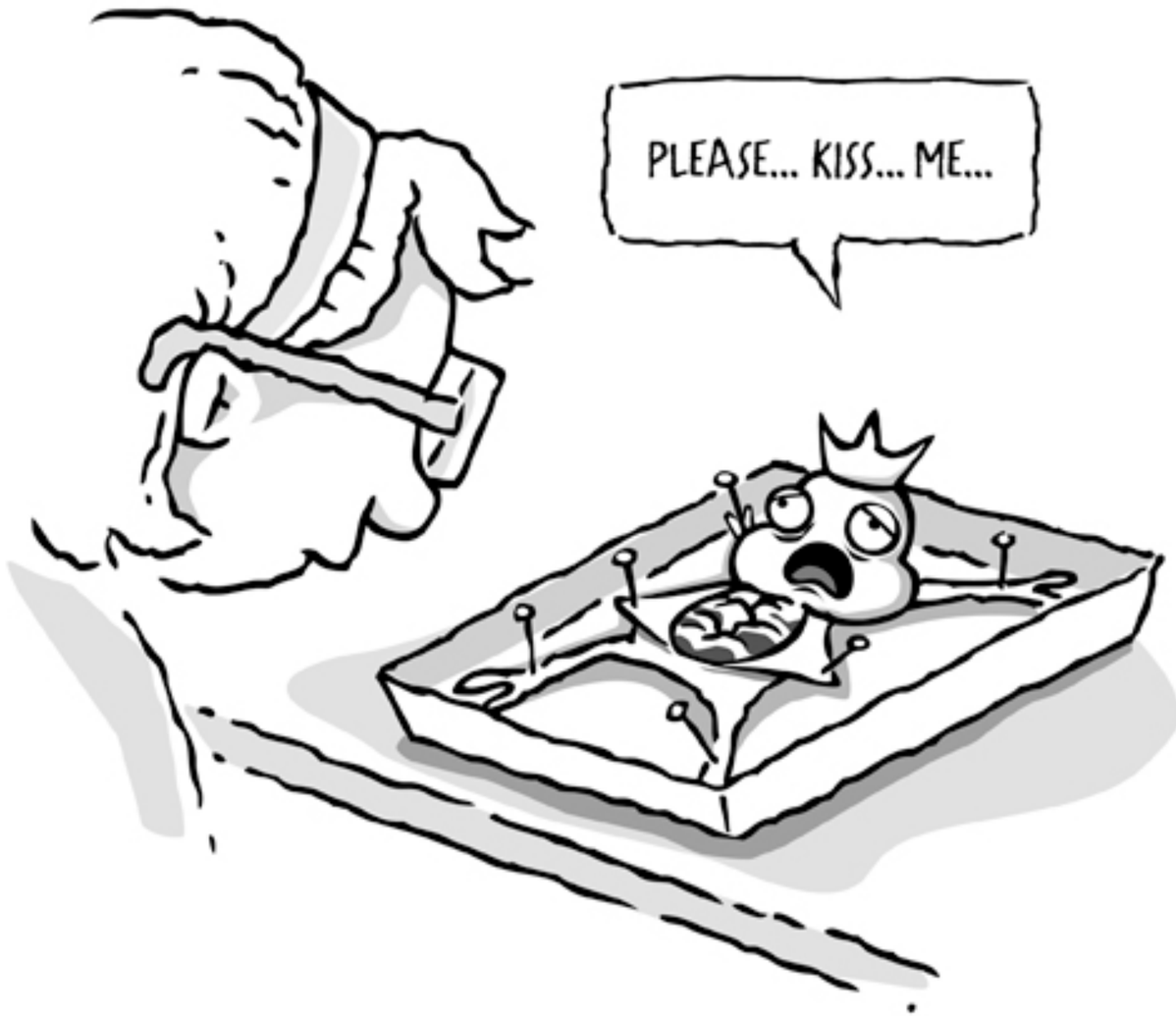
- People from different backgrounds
    - No biology since high school



COURSE DESIGN

CHALLENGES

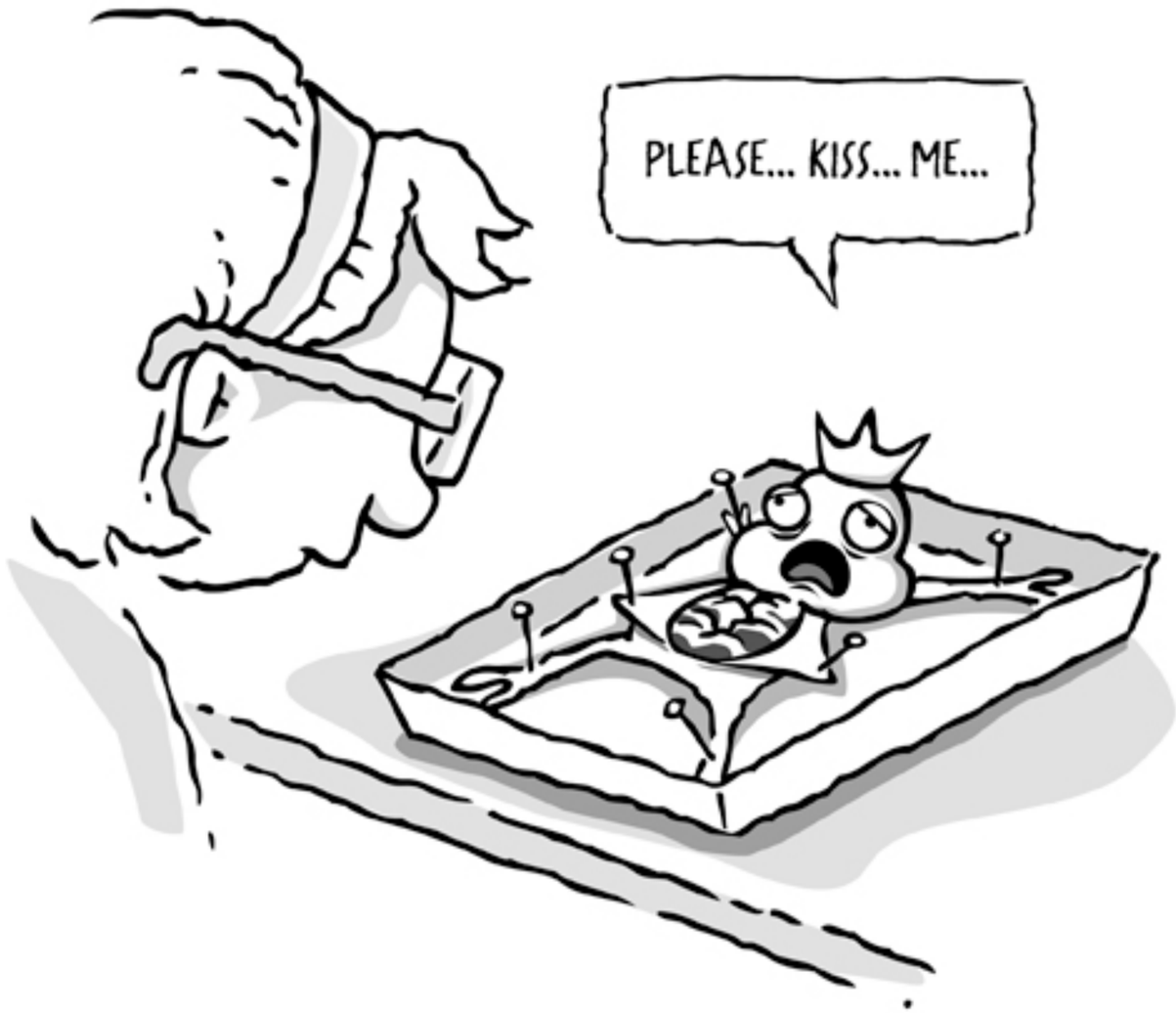
- Wide topic that touches many fields and sub-fields
  - Molecular biology, structural biology, genomics, medicine



COURSE DESIGN

CHALLENGES

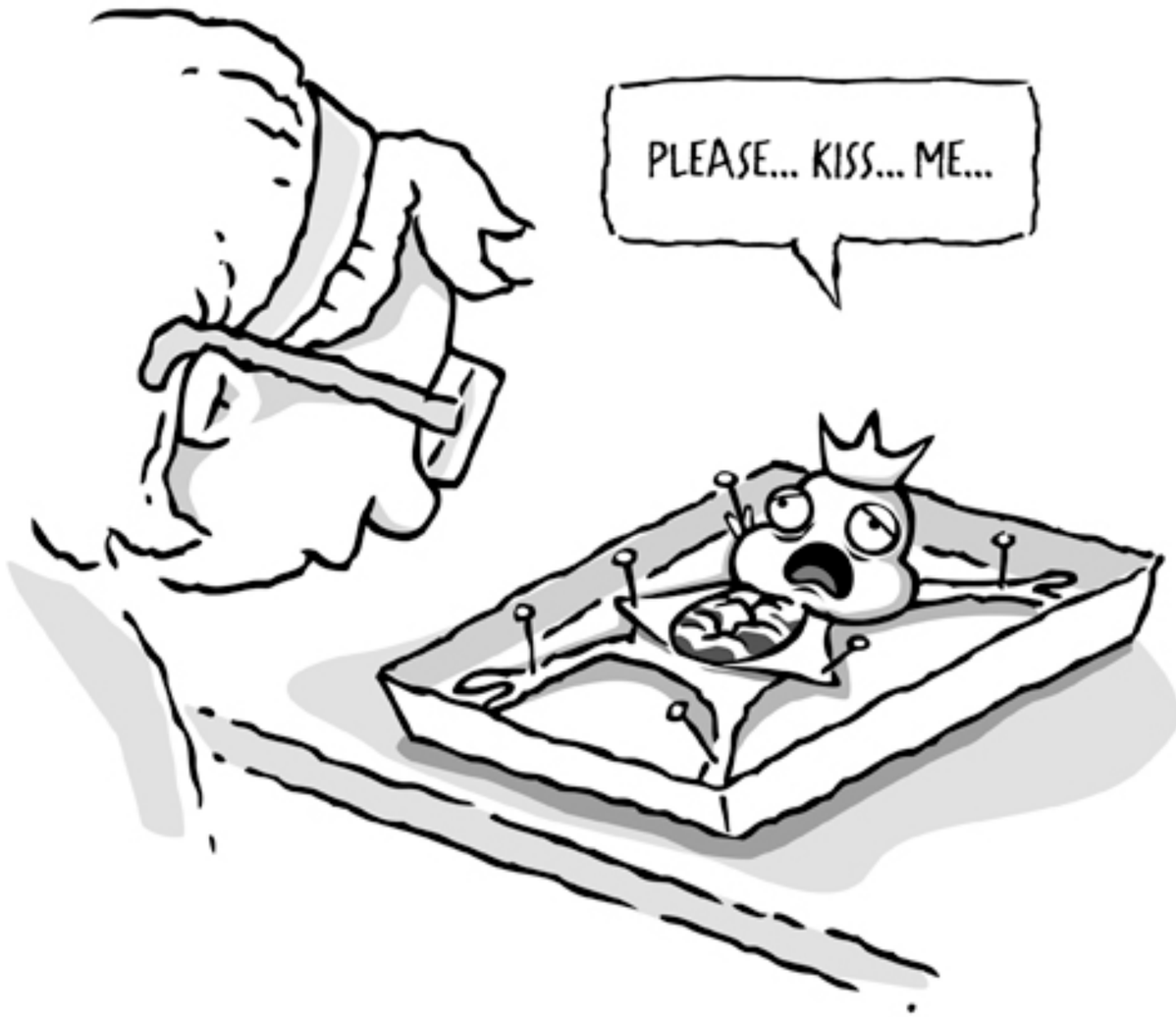
- Different programming backgrounds and skill sets
  - HPC, scripting, compiled languages, unix, web, etc.



COURSE DESIGN

CHALLENGES

- Different interests and goals for the course
    - Casual interest in biology
    - Elective credit



COURSE DESIGN

PROVIDE KNOWLEDGE AND TOOLS TO UNDERSTAND AND SOLVE A RESEARCH PROBLEM (IN BIOLOGICAL SPACE) WITH THE AID OF COMPUTATION

THINGS YOU LEARN HERE CAN BE UNIVERSALLY APPLIED

COURSE DESIGN

COURSE GOALS

- Introduce major techniques and bioinformatics algorithms
  - Implement some of them
- Develop applications for bioinformatic analysis
  - Make them accessible to other researchers
- Become proficient in most important bioinformatics databases
  - NCBI, EBI, PDB, etc.
- Provide practical experience in research

COURSE DESIGN

SECONDARY GOAL: BECOME "COCKTAIL PARTY" CONVERSANT IN BIOLOGY

Syllabus

COURSE DESIGN

- Week 1,2 - Biology and Biological Databases
- Week 3,4 - Sequence Alignment
- Week 5,6 - Protein Bioinformatics
- Week 7,8,9 - Discovery, Genomes and Human Variation

COURSE DESIGN

- Week 1: Genomics, Bioinformatics and Molecular Biology
  - A high-level view of increasingly important role of computing in the biological sciences will be presented.

- Week 2: Genomes, Sequences and Databases
  - A survey of the current state of the art in storing, organizing and analyzing large data sets will be discussed
  - The advantages and disadvantages of these methods will be explored in the context of academic and commercial research initiatives.

COURSE DESIGN

- Week 3: Sequence Alignment
  - Fast, reliable alignment of text strings started the bioinformatics revolution. This lecture will show how these seemingly simple strings form the basis of almost all bioinformatics research.
- Week 4: Analyzing DNA and Protein Sequences
  - Techniques for sequence analysis will be discussed

COURSE DESIGN

- Week 5: Protein Structure and Function
  - Proteins are central building blocks of all organisms
  - Take bioinformatics to the third-dimension, showcasing how the spatial assembly and interactions of proteins support life and cause of disease

- Week 6: Structure Analysis and Molecular Modeling
  - Understanding protein function holds the promise developing therapeutics and curing diseases, but the computational complexity of analyzing three-dimensional models presents obstacles that have been difficult to overcome. This lecture will discuss approaches to shape analysis and comparison that can be scaled to large data sets.

COURSE DESIGN

- Week 7: In-Silico Drug Discovery
  - Approaches to using computer models to develop new drugs will be presented. We will discuss how years of playing Tetris might be more useful than you thought in combating antibiotic resistant pathogens.

- Week 8: Student Presentations / The Human Genome and Disease
  - The cause of diseases can be as simple as a single misplaced letter in a DNA sequence. From gene to disease, we will trace the genetic origin of disease. We will explore different approaches to cataloging and analyzing these changes.

COURSE DESIGN

- Week 9: Personal Genomics and Drug Discovery
  - Personalized genomic analysis is being used by consumers to better understand their health and their ancestry. The technologies used to power these services will be introduced as well as the different approaches used to provide web services to analyze the data.
- Final Exam Week 10: Final Project Presentations
  - Students will present a research topic in bioinformatics of their own choosing

Course Work

COURSE WORK

- 6 assignments (10% each)
  - Split between practical problems and implementation problems
- Gene Study Presentations (15%)
- Final project (25%):
  - Implement a bioinformatic method of your choosing

WE WILL DISCUSS THIS IN DETAIL LATER ON

COURSE WORK

Week	Material	Homework	Final Project
1	Bioinformatics, Genomics, and Molecular Biology	Assignment 1	
2	Biological Databases	Assignment 2	
3	Sequence Alignment	Assignment 3	
4	Analyzing Sequences	Assignment 4	
5	Protein Structure and Function	Assignment 5	
6	Molecular Modelling	Assignment 6	
7	in-silico drug discovery	Gene Study	Proposals Due before class
8	Student Gene Presentations		
9	The Human Genome and Disease		
Final	Student Video Presentations		Projects Due

COURSE WORK

- Identify a gene of interest
  - Related to disease, biofuels, manufacturing, etc.
  - Family of genes, what pathway, what does it do?
  - Find a gene of unknown function, predict its function…
- Become an expert on that gene
  - Publications
  - Sequence, structure and function analysis on it
- Present to classs

K-RAS AND CANCER

AIN'T NUTHIN' BUT A G PROTEIN, BABY.

ASSIGNMENTS & PROJECTS

FINAL PROJECT

- Project of your choosing
- Examples
  - Implement a published bioinformatic method
  - Invent your own
  - Research a topic
  - Anything related to anything we've talked about...
- Present it to the class

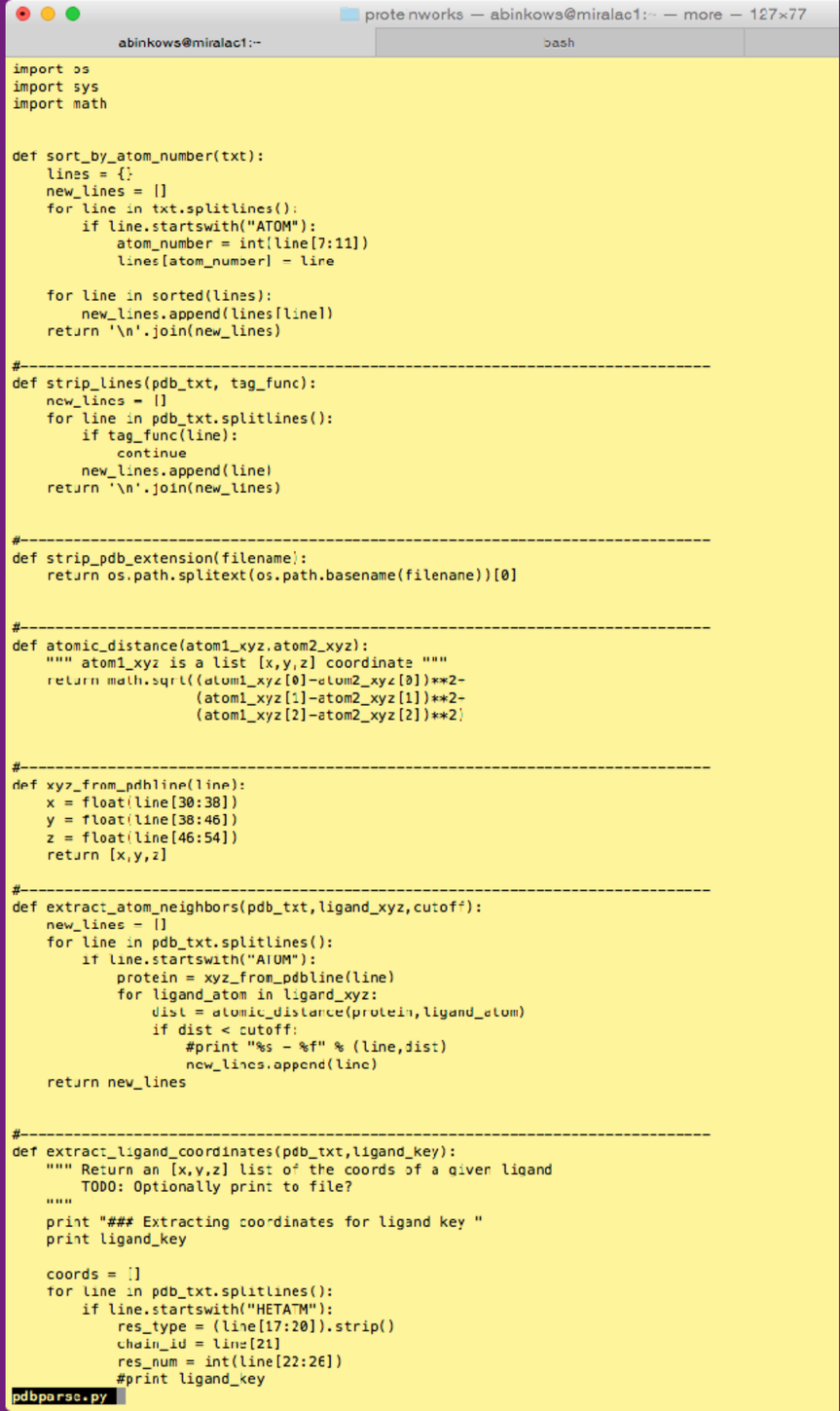
ASSIGNMENTS & PROJECTS

- Honor Code
  - All the assignments should be your own work
  - Department policies are strictly enforced
- Citing Resources
  - Cite any resources you use on homework, presentation and projects
  - Includes online resources
    - StackOverfow, BioStars, blogs, GitHub, etc.
- Third-party libraries and software
  - There are many great libraries for bioinformatics
  - Unless specifically stated, you should not use them

Course Technologies

COURSE TECHNOLOGIES

- Bioinformatics is focused on developing solutions to biological problems
  - Not on mastery of any particular language
  - Many bioinformatics resources violate all CS good design and implementation rules
  - Utility trumps all
    - Flat files are the lifeblood of many bioinformatic pipelines
    - The rise and fall of DoubleTwist



COURSE TECHNOLOGIES

- Bioinformatics requires aptitude in a variety of programming languages
  - Scripting (Python, Perl)
  - Command line (Bash, wget, curl, etc.)
  - Compiled languages (Fortran, C, C++)
  - Specialized languages (R, Matlab)
  - Web programming (Javascript, PHP, HTML)

```
import os
import sys
import math


def sort_by_atom_number(txt):
    lines = {}
    new_lines = []
    for line in txt.splitlines():
        if line.startswith("ATOM"):
            atom_number = int(line[7:11])
            lines[atom_number] = line

    for line in sorted(lines):
        new_lines.append(lines[line])
    return '\n'.join(new_lines)

#--------------------------------------------------------------------------------
def strip_lines(pdb_txt, tag_func):
    new_lines = []
    for line in pdb_txt.splitlines():
        if tag_func(line):
            continue
        new_lines.append(line)
    return '\n'.join(new_lines)


#--------------------------------------------------------------------------------
def strip_pdb_extension(filename):
    return os.path.splitext(os.path.basename(filename))[0]


#--------------------------------------------------------------------------------
def atomic_distance(atom1_xyz,atom2_xyz):
    """ atom1_xyz is a list [x,y,z] coordinate """
    return math.sqrt((atom1_xyz[0]-atom2_xyz[0])**2+
                     (atom1_xyz[1]-atom2_xyz[1])**2+
                     (atom1_xyz[2]-atom2_xyz[2])**2)


#--------------------------------------------------------------------------------
def xyz_from_pdbline(line):
    x = float(line[30:38])
    y = float(line[38:46])
    z = float(line[46:54])
    return [x,y,z]

#--------------------------------------------------------------------------------
def extract_atom_neighbors(pdb_txt,ligand_xyz,cutoff):
    new_lines = []
    for line in pdb_txt.splitlines():
        if line.startswith("ATOM"):
            protein = xyz_from_pdbline(line)
            for ligand in ligand_xyz:
                if atomic_distance(protein,ligand)<cutoff:
                    new_lines.append(line)
                    break
    return '\n'.join(new_lines)


#--------------------------------------------------------------------------------
def extract_ligand_coordinates(pdb_txt,ligand_key):
    """ Return an [x,y,z] list of the coords of a given ligand
        TODO: optionally print to file
    """
    print "### Extracting coordinates for ligand key "
    print ligand_key

    coords = []
    for line in pdb_txt.splitlines():
        if line.startswith("HETATM"):
            res_type = (line[17:20]).strip()
            chain_id = line[21]
```

MORE IMPORTANT THAN YOU MIGHT THINK

COURSE TECHNOLOGIES

- Command line tools
  - “Dirty little secret” of bioinformatics
  - You have to do something with you data before/after you big cluster runs

Command-line tools can be 235x faster than your Hadoop cluster

Sat 25 January 2014 by Adam Drake

Introduction

As I was browsing the web and catching up on some sites I visit periodically, I found a cool article from Tom Hayden about using Amazon Elastic Map Reduce (EMR) and mrjob in order to compute some statistics on win/loss ratios for chess games he downloaded from the millionbase archive, and generally have fun with EMR. Since the data volume was only about 1.75GB containing around 2 million chess games, I was skeptical of using Hadoop for the task, but I can understand his goal of learning and having fun with mrjob and EMR. Since the problem is basically just to look at the result lines of each file and aggregate the different results, it seems ideally suited to stream processing with shell commands. I tried this out, and for the same amount of data I was able to use my laptop to get the results in about 12 seconds (processing speed of about 270MB/sec), while the Hadoop processing took about 26 minutes (processing speed of about 1.14MB/sec).

After reporting that the time required to process the data with 7 c1.medium machine in the cluster took 26 minutes, Tom remarks

> "This is probably better than it would take to run serially on my machine but probably not as good as if I did some kind of clever multi-threaded application locally."

This is absolutely correct, although even serial processing may beat 26 minutes. Although Tom was doing the project for fun, often people use Hadoop and other so-called *Big Data (tm)* tools for real-world processing and analysis

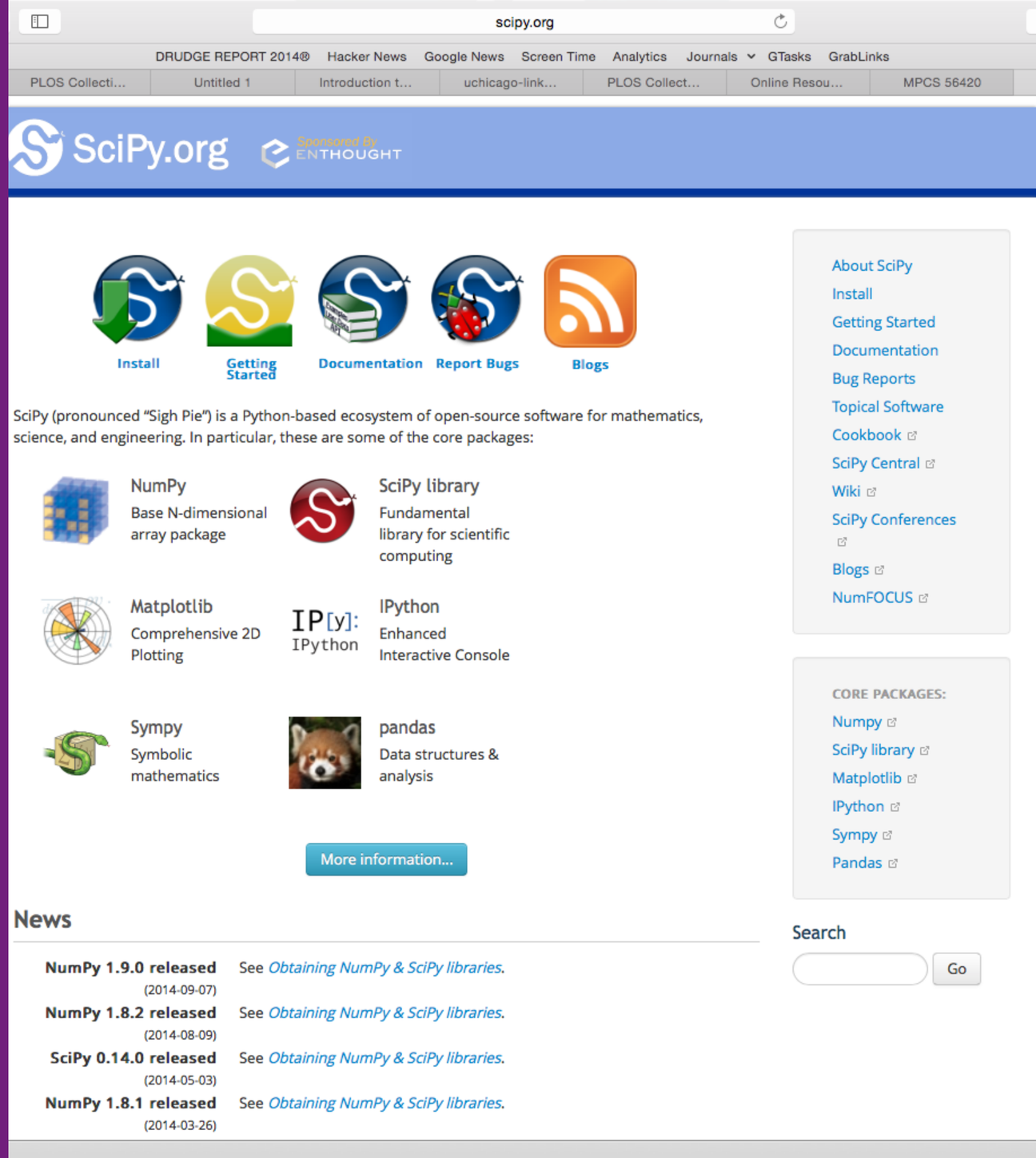
COURSE TECHNOLOGIES

CONSIDERATIONS FOR BIOINFORMATICS PROGRAMMING

- Portability - Will it run on my desktop and a legacy SGI machine?
- Scalability - Will it run on a cluster?
- Development speed - How long will it take to write
- Longevity - Is this a one time script or a full fledged application?
- Deliverability - Will this eventually be a web app or in the App Store?
- Target Audience - Who (besides me) may be running this? What is their background?

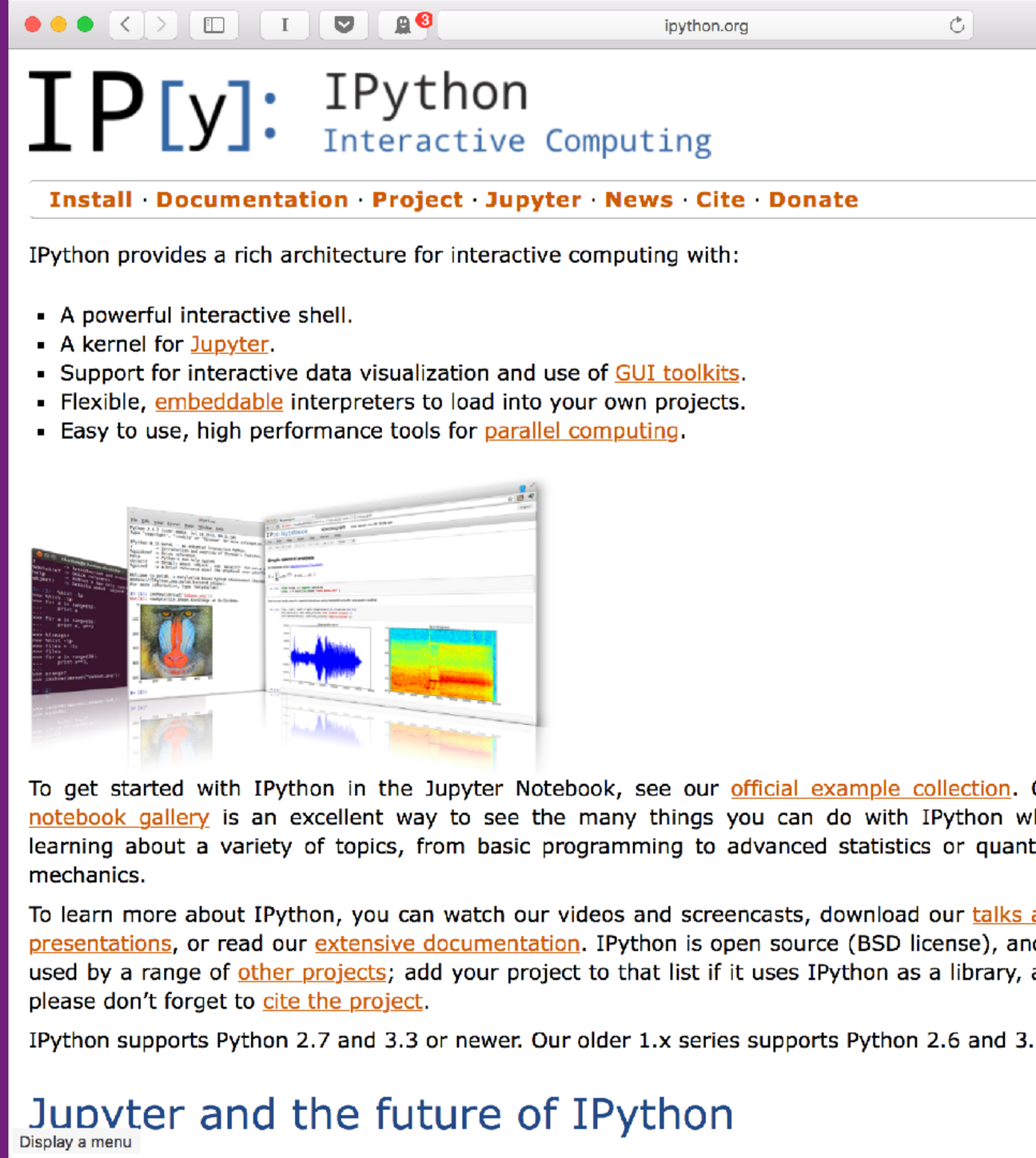
COURSE TECHNOLOGIES

- Programming languages for this course
  - Lectures and demonstrations will be conducted mostly in Python
  - Historical language of bioinformatics is Perl
- Why Python?
  - Great online resources and support
  - Support for SciPy scientific programming
  - Extensibility for C modules
    - Optimize when you need it
  - Native web application language



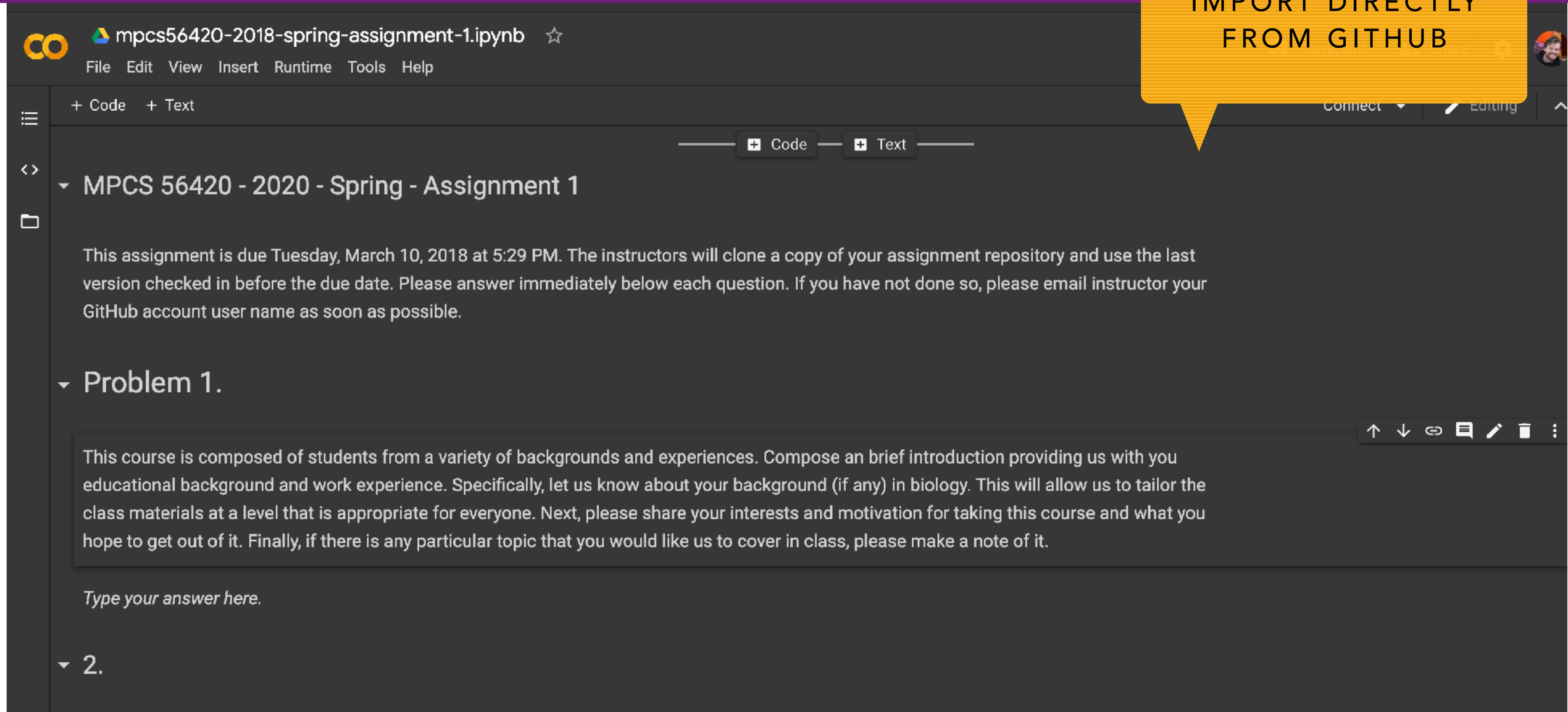
COURSE TECHNOLOGIES

- Jupyter notebooks
  - Support for reproducible workflows
  - Excellent support for SciPy
  - Consistent environment across platforms
  - Quickly becoming a “industry” standard
    - Publishing reproducible results



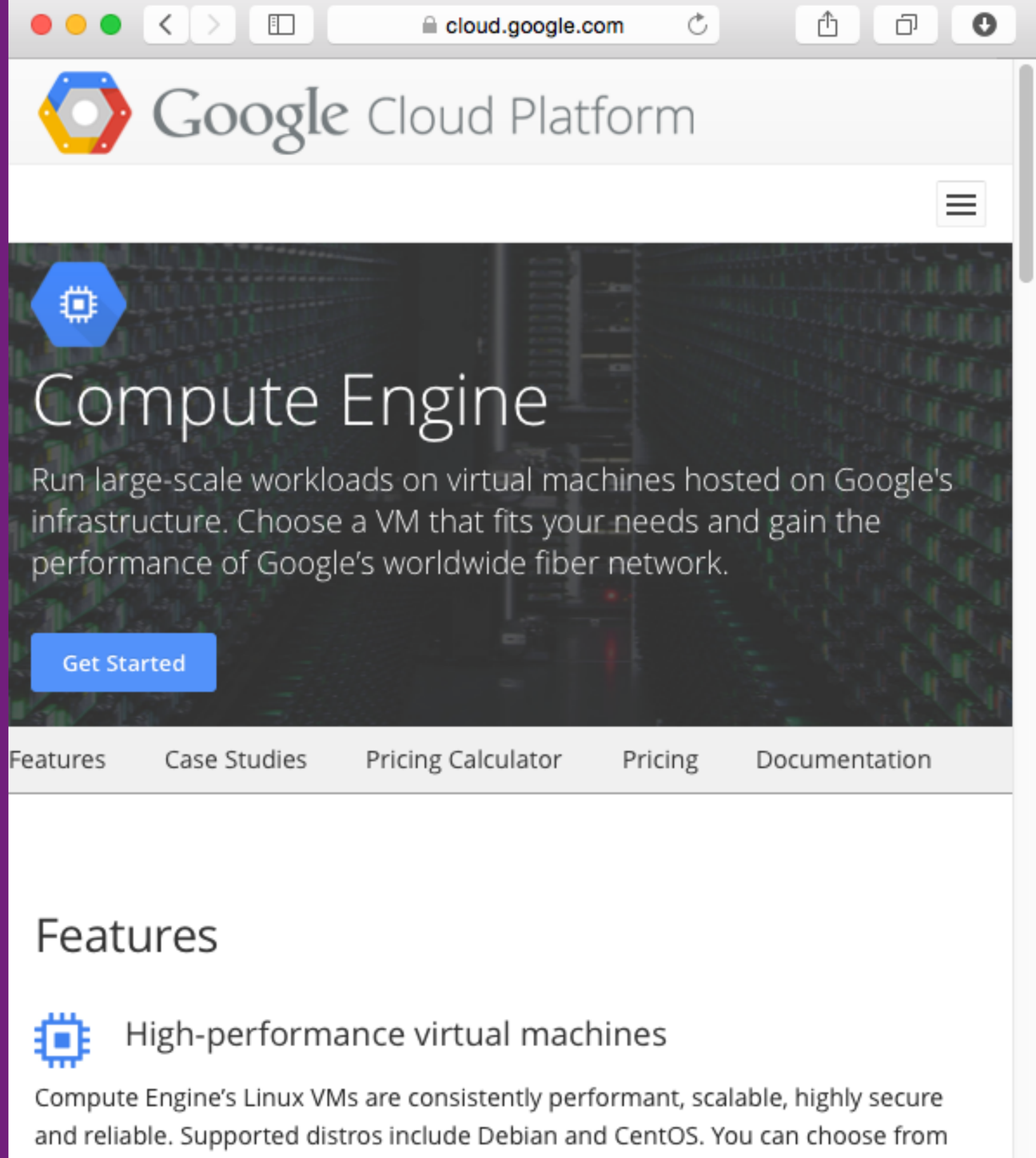
COURSE TECHNOLOGIES

GOOGLE COLLABORATORY CAN IMPORT DIRECTLY FROM GITHUB



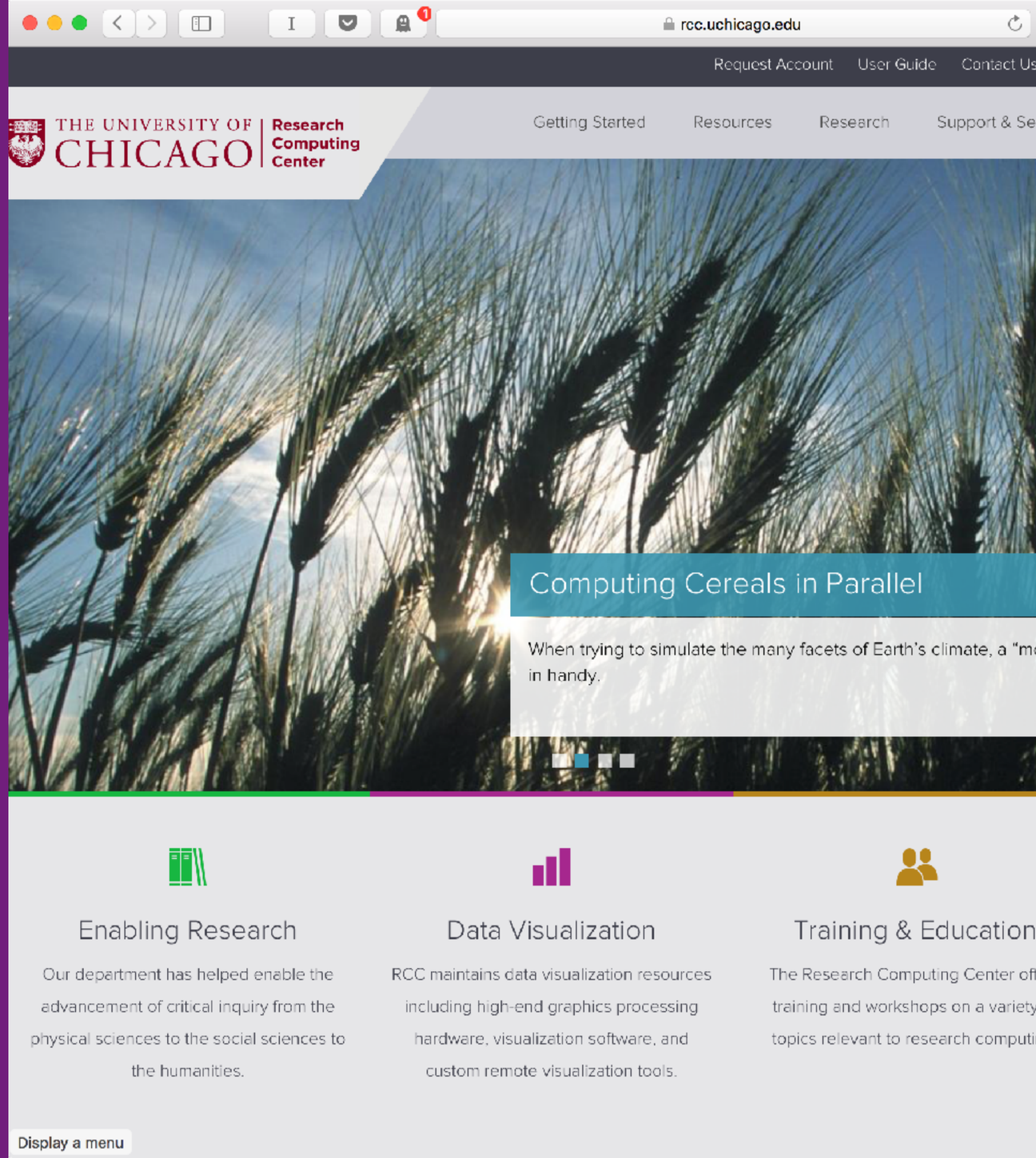
COURSE TECHNOLOGIES

- Google Could Platform
  - App Engine, Compute Engine, Storage, Genomics
  - Python, Java, PHP and Go
  - NumPy and Matplotlib are available on Google App Engine
  - Support for “big data” workflows



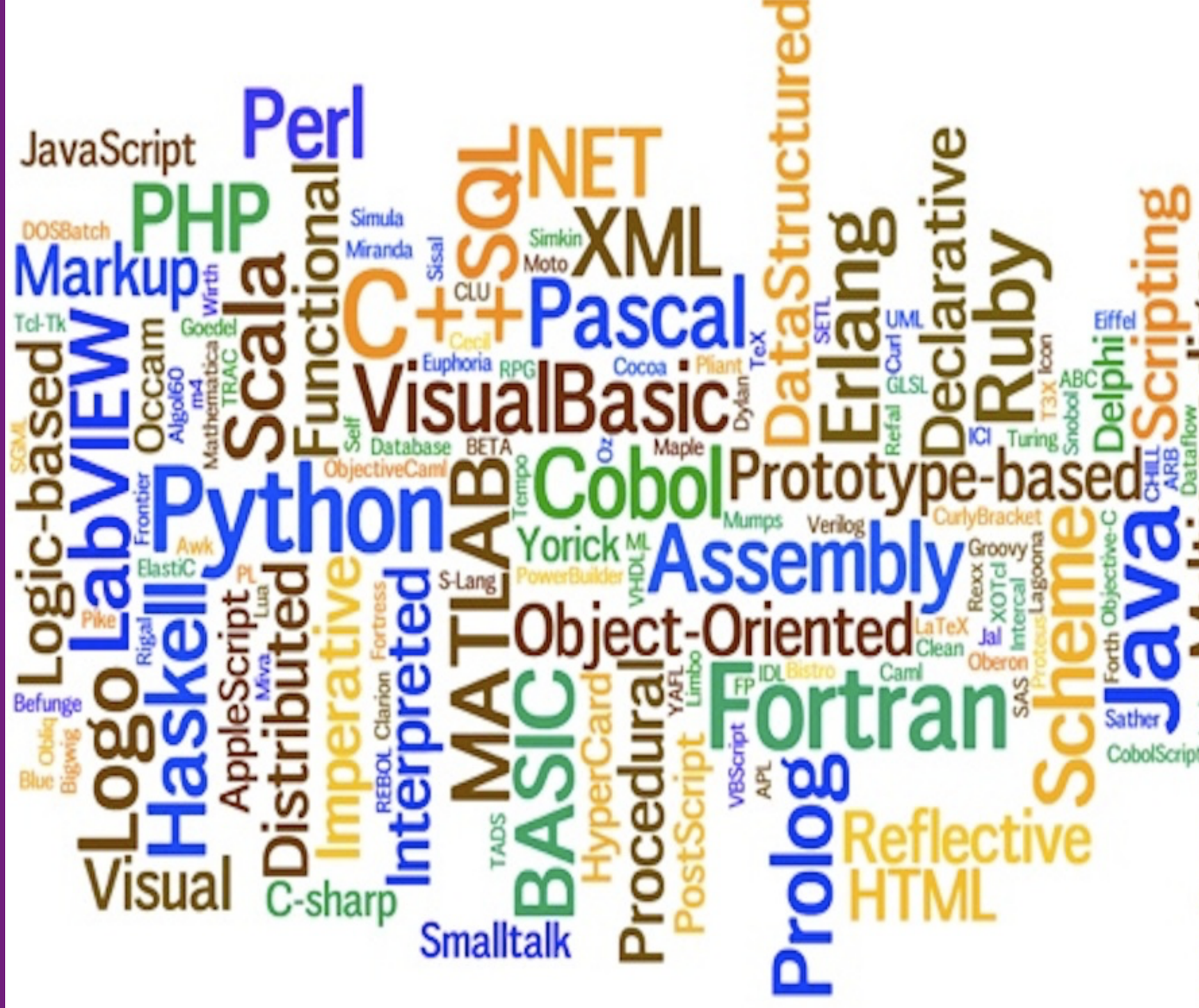
COURSE TECHNOLOGIES

- Research Computing Center
  - High-performance computing and visualization resources
  - High-capacity storage and backup
  - Software
  - High-speed networking
  - Hosted data sets



COURSE TECHNOLOGIES

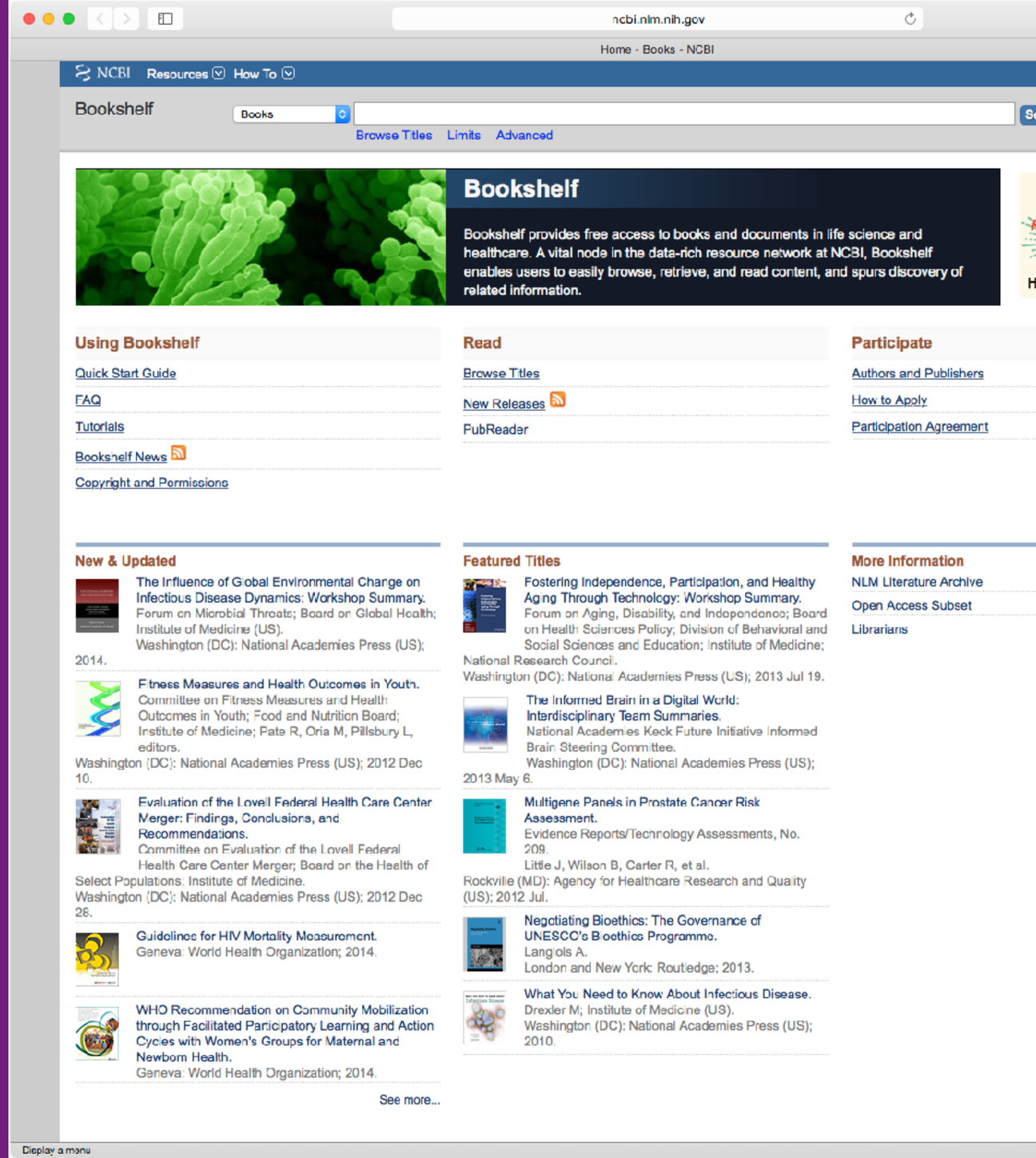
- You can use any other languages and technologies for your final projects
- Limited support from instructors



Course Resources

COURSE RESOURCES

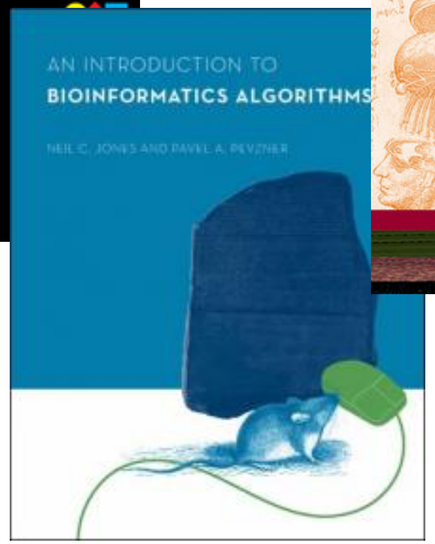
- Text Books
  - No required text books
  - Books become outdated quickly
  - Publications (scientific journals) have latest information
  - Quality online references from reputable sources
    - NCBI Handbook (http://www.ncbi.nlm.nih.gov/books/NBK143764/?term=handbook)

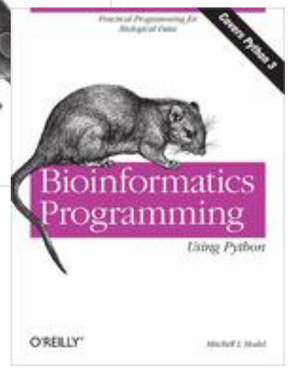


COURSE RESOURCES

TEXT BOOKS

FREE DIGITAL COPIES AVAILABLE THROUGH UCHICAGO LIBRARY





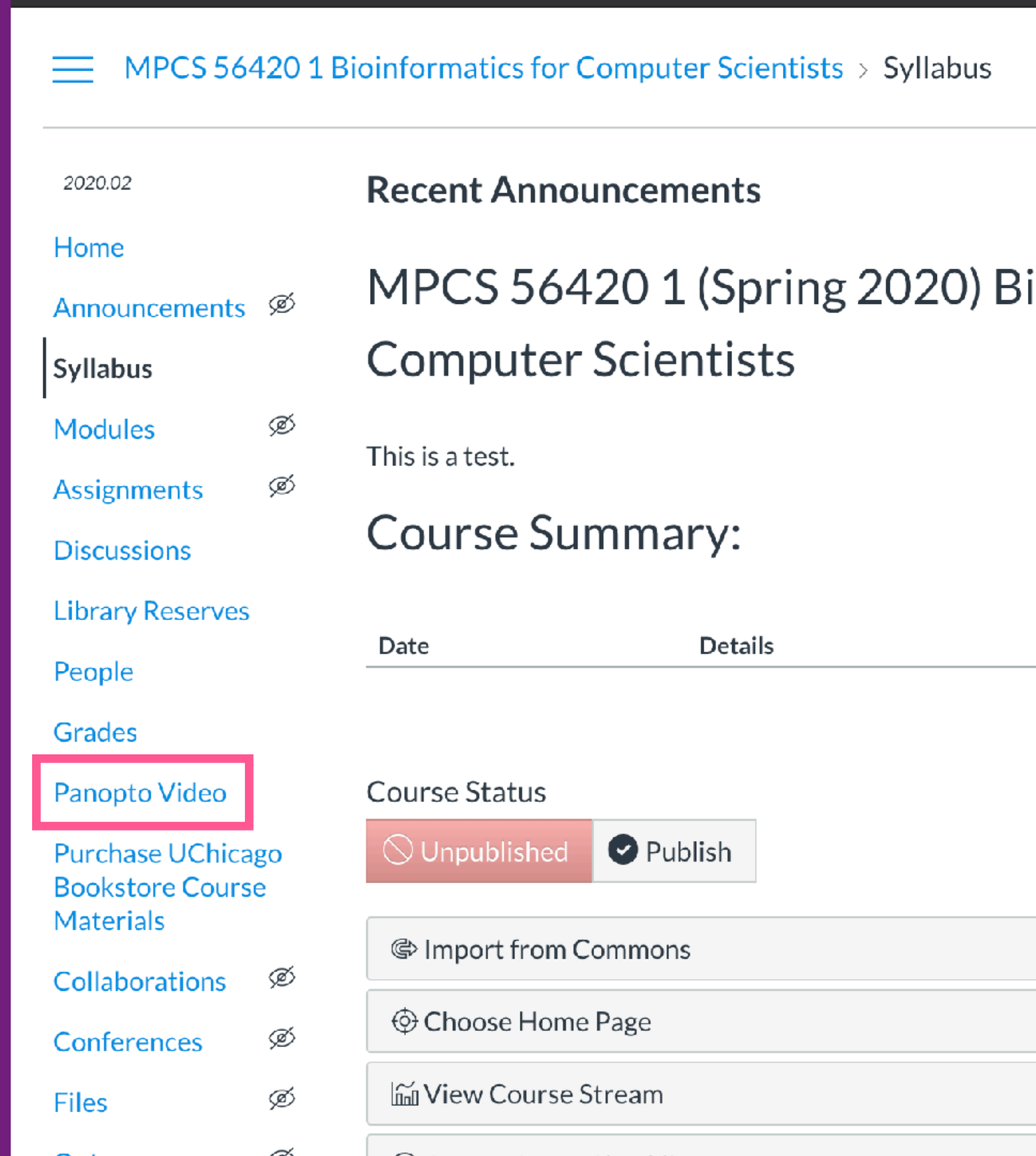
RECOMMENDED

COURSE RESOURCES

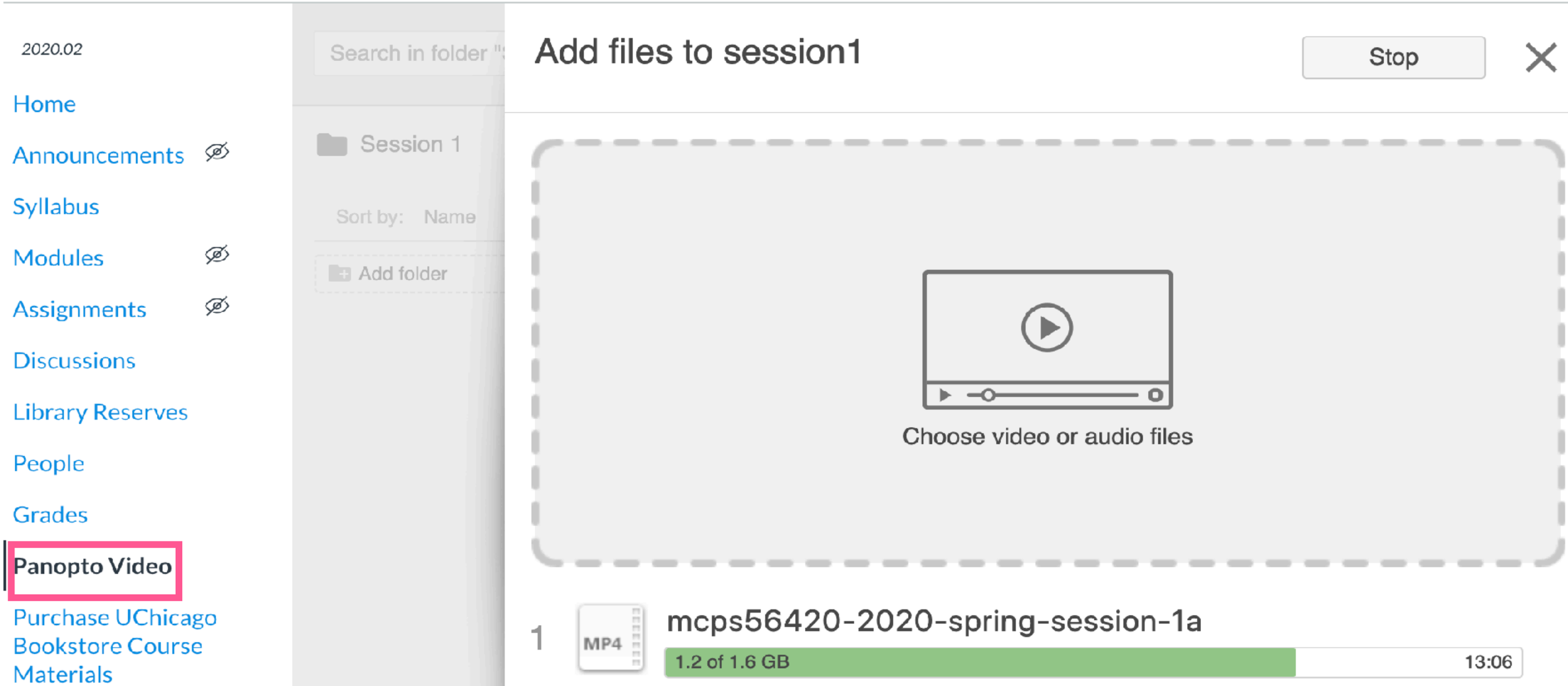
- Biopython
  - Open sources framework for bioinformatics
  - Great resources and place to get started for algorithms and data structures
  - Unless specified, you should not use it

COURSE RESOURCES

- Canvas "Lite"
- Links to materials
  - Github
  - Panopto

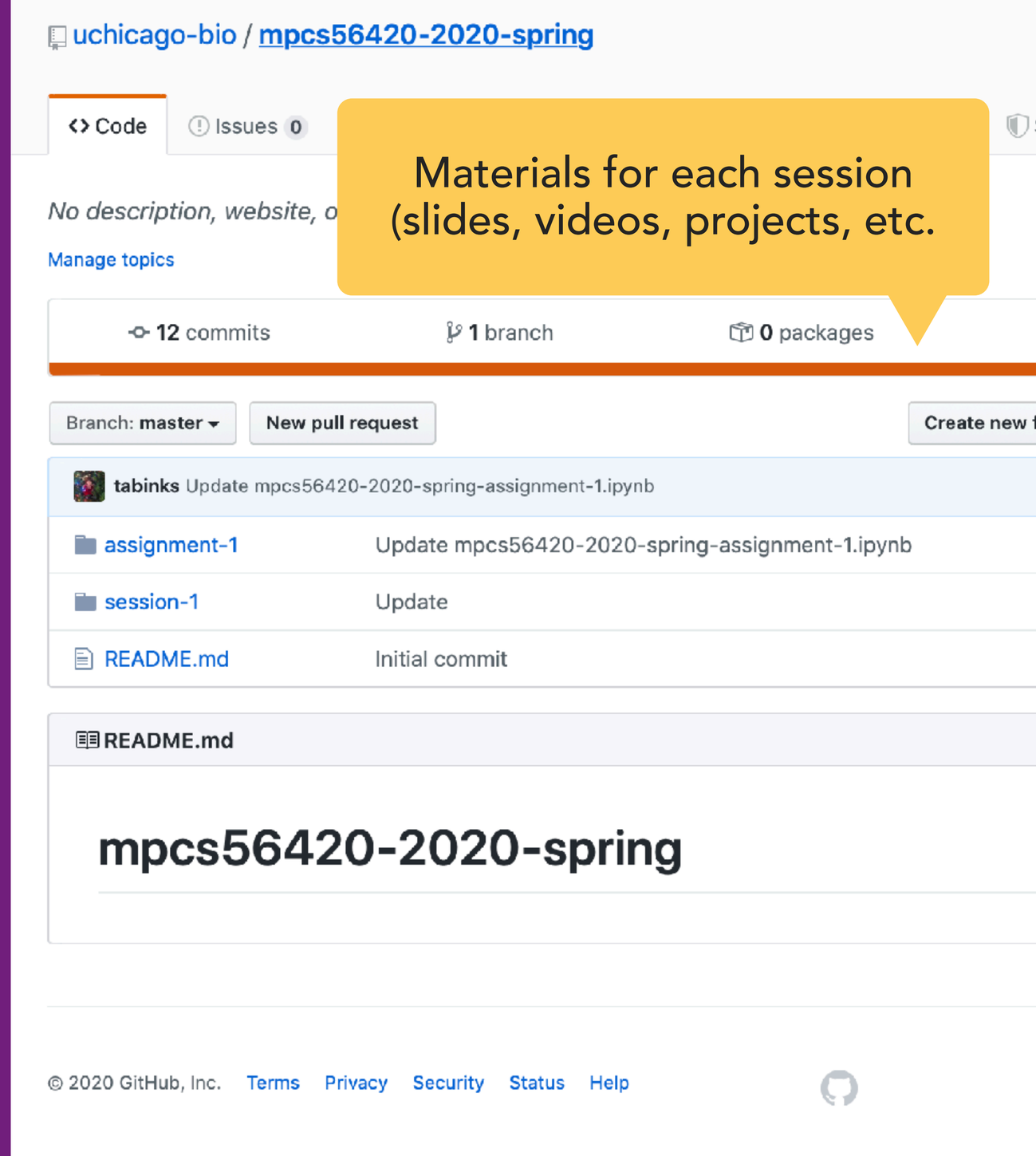


COURSE RESOURCES



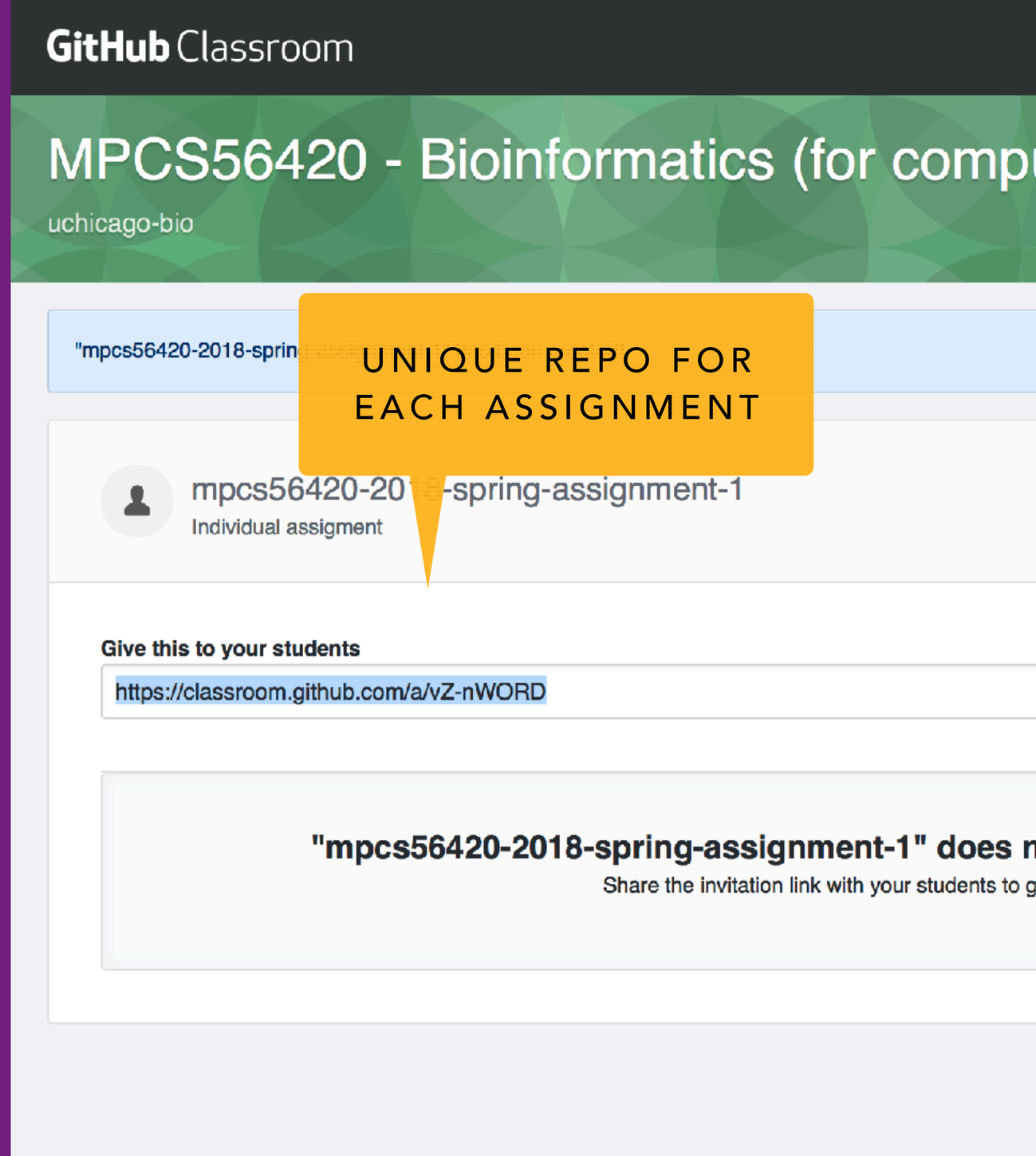
COURSE RESOURCES

- Github Repository for all course materials (no website 😢)
- https://github.com/uchicago-mobi/mpcs51032-2020-spring



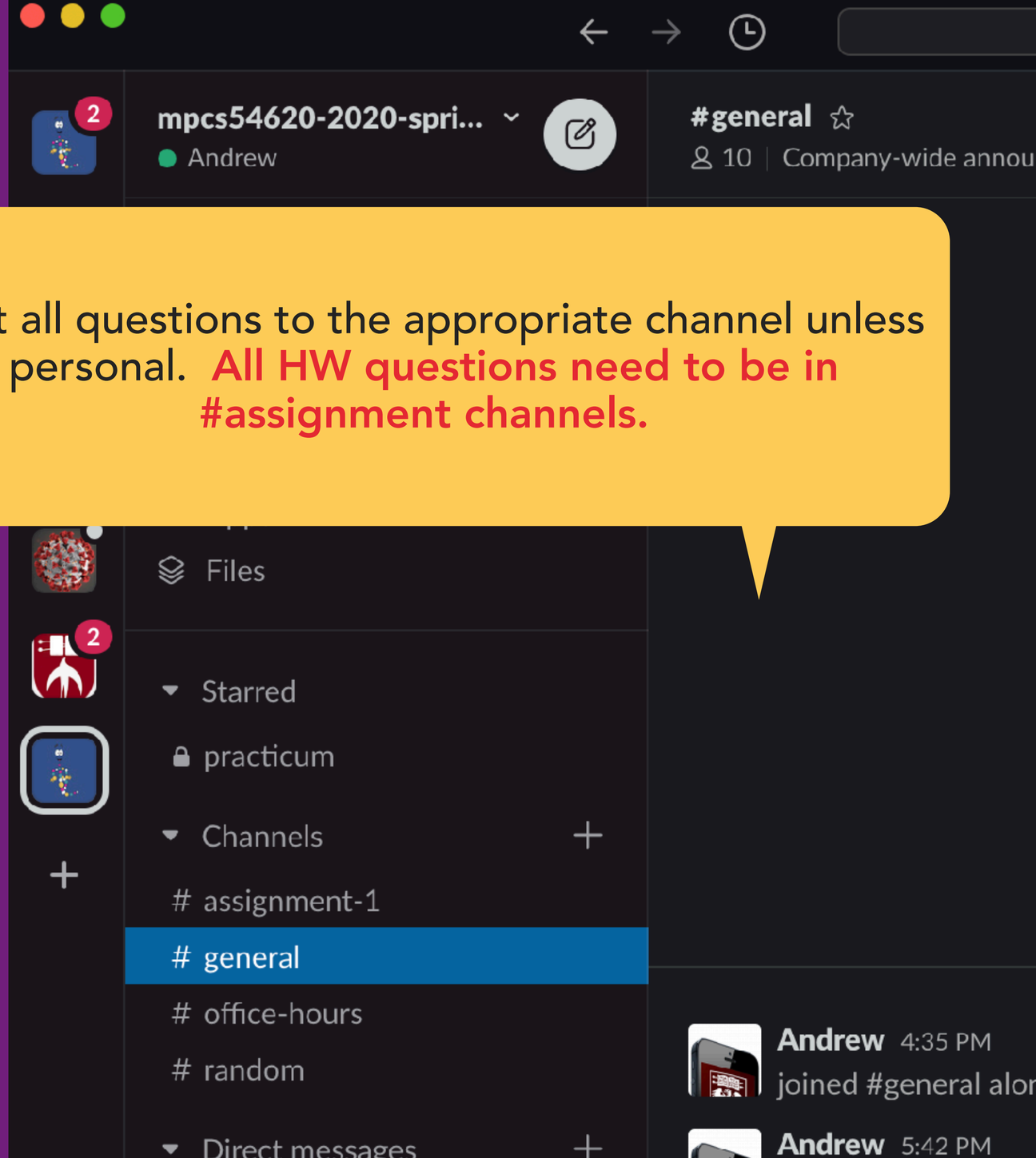
COURSE RESOURCES

- Github for coursework
  - Commit your repo to “turn in” assignments
  - Private repositories
  - Link on each assignment



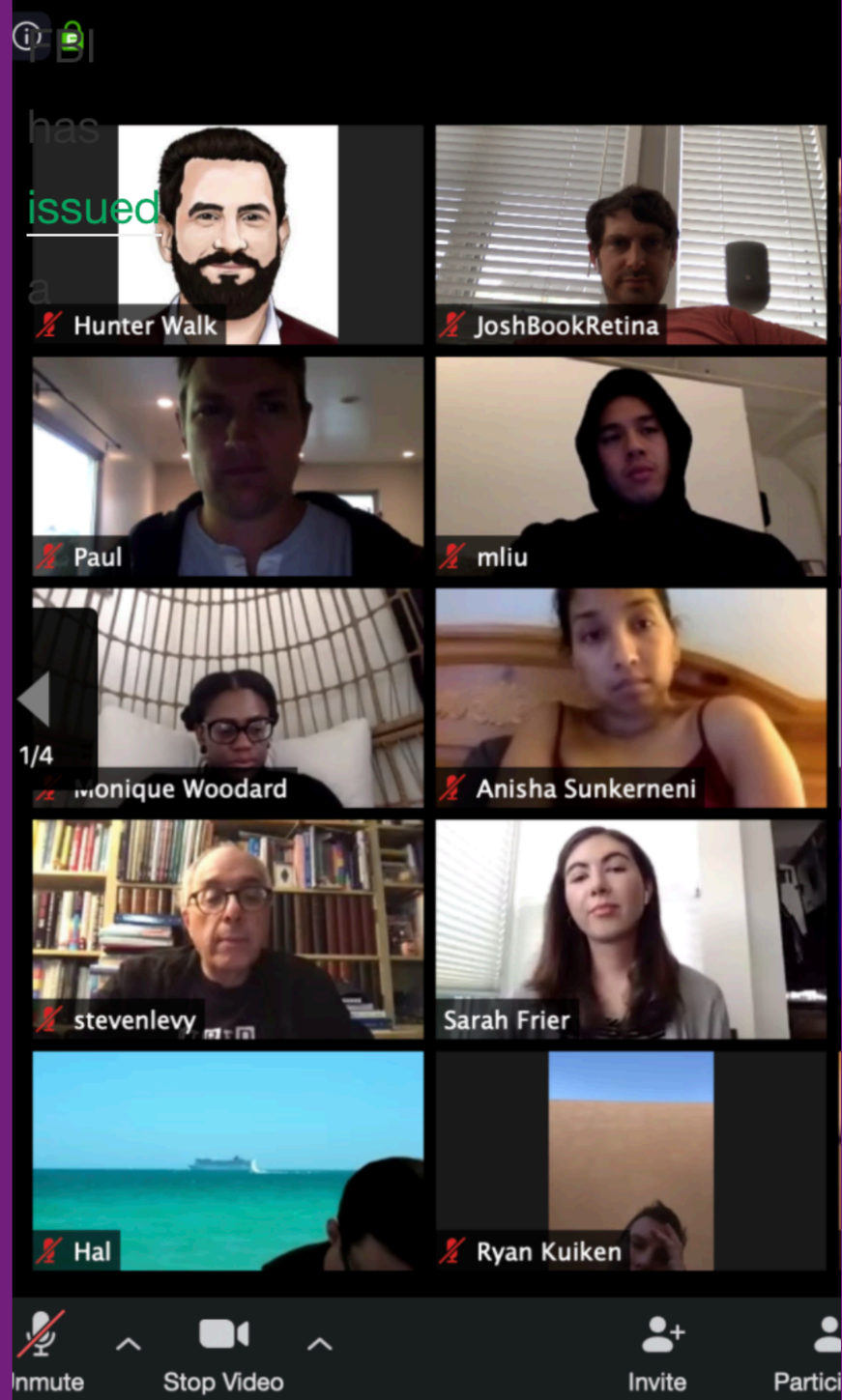
COURSE LOGISTICS

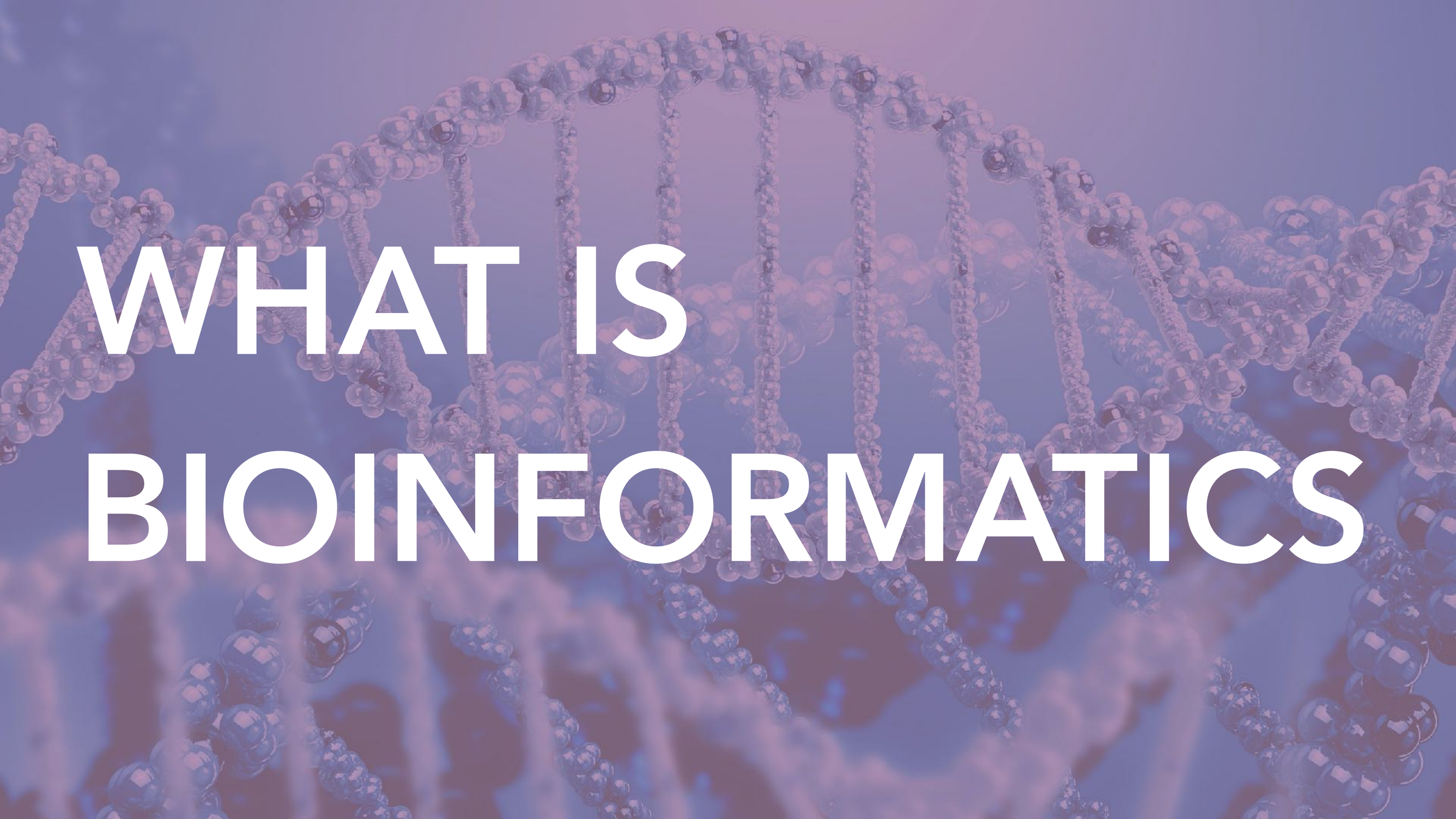
- Discussion forum rules
  - Anyone can answer a question
  - Post sources (if available) when answering questions



COURSE RESOURCES

- Please make sure you are muted if you are not talking
- Questions
  - "Raise Hand" if you have a question
  - Type it in the chat
  - Post in slack
- All sessions will be recorded
  - You can opt-out by turning camera and microphone off
  - Will not be distributed outside of class




WHAT IS
BIOINFORMATICS

WHAT IS BIOINFORMATICS

- "Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline”
  - Official definition from the National Center for Biotechnology Information (NCBI)

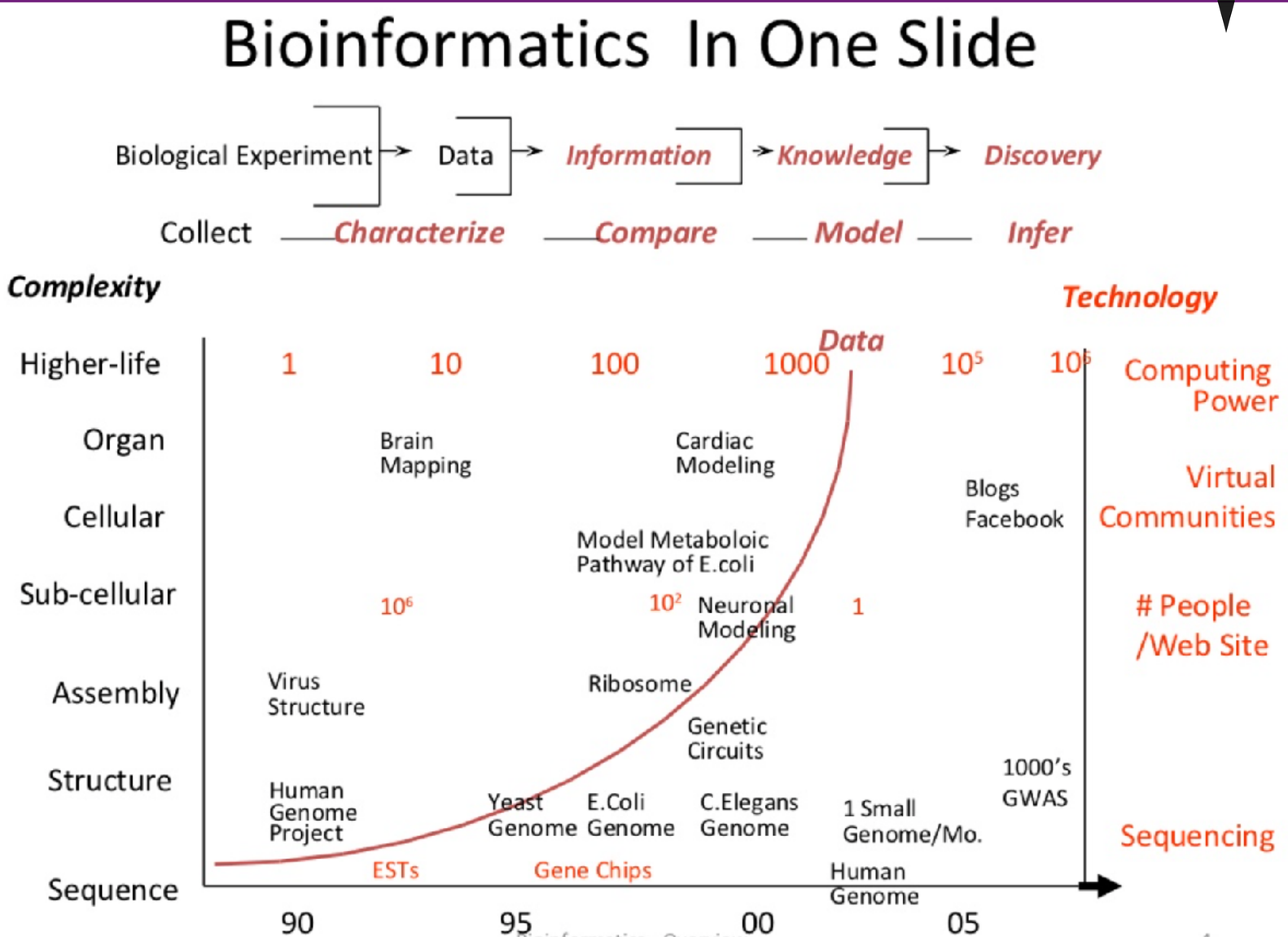
WHAT IS BIOINFORMATICS

THREE IMPORTANT SUB-DISCIPLINES

- Development of new algorithms and statistics with which to assess relationships among members of large data sets
- Analysis and interpretation of various types of data
  - Nucleotide and amino acid sequences, proteins
- Development and implementation of tools that enable efficient access and management of different types of information

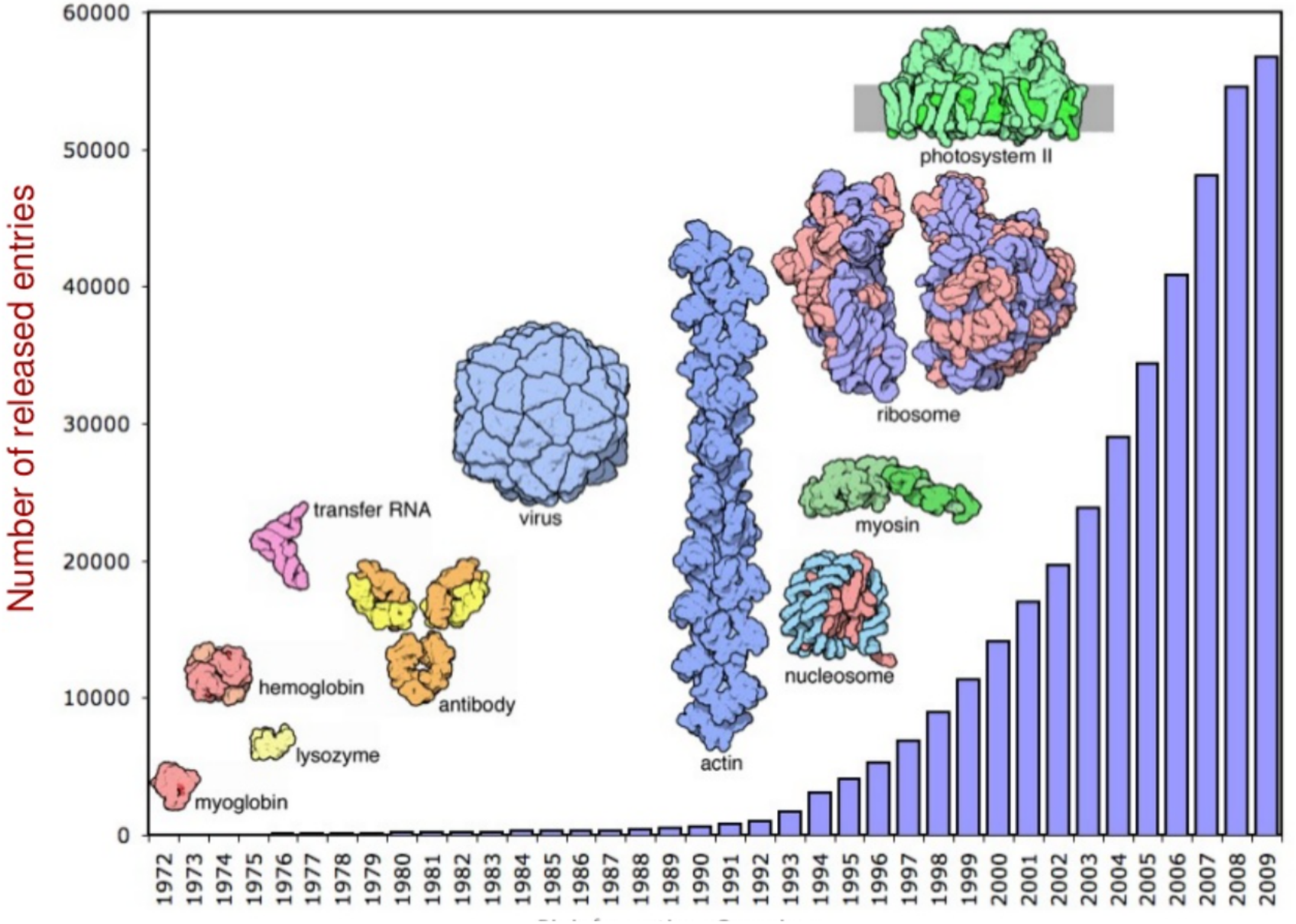
WHAT IS BIOINFORMATICS?

PHILIP BOURNE, UCSD



WHAT IS BIOINFORMATICS

- Growth of data
  - Protein Data Bank (PDB)



WHAT IS BIOINFORMATICS

- Using computers to answer biological questions
  - Including management and use of biological information
- Not LIMS (laboratory information management systems)
  - Collect data from experiments
  - Organize lab notebooks

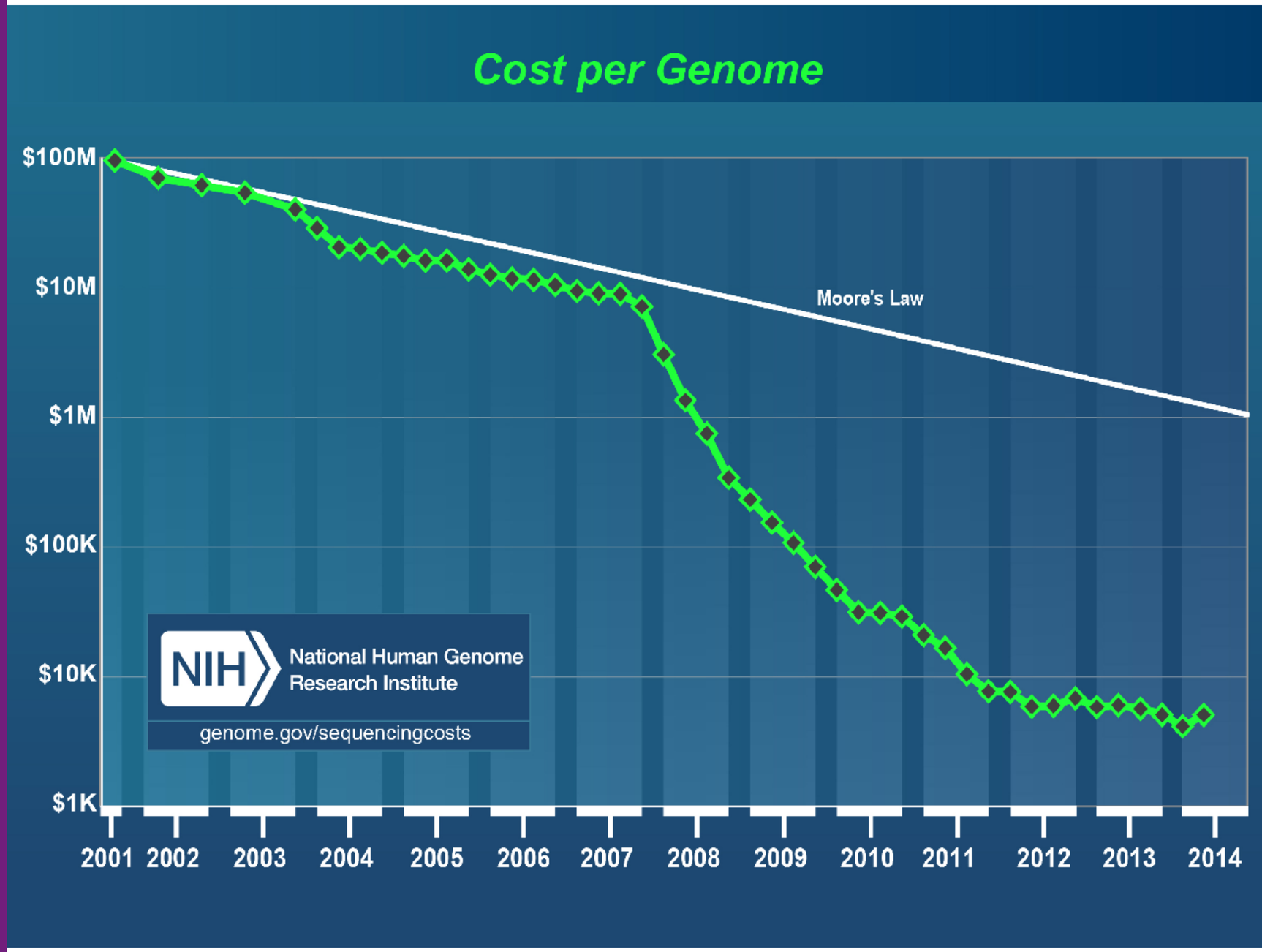
WHAT IS BIOINFORMATICS

BANKING USES COMPUTERS BUT ITS NOT CALLED BANKFORMATICS

- My answer: Its the next step in biology and biological research

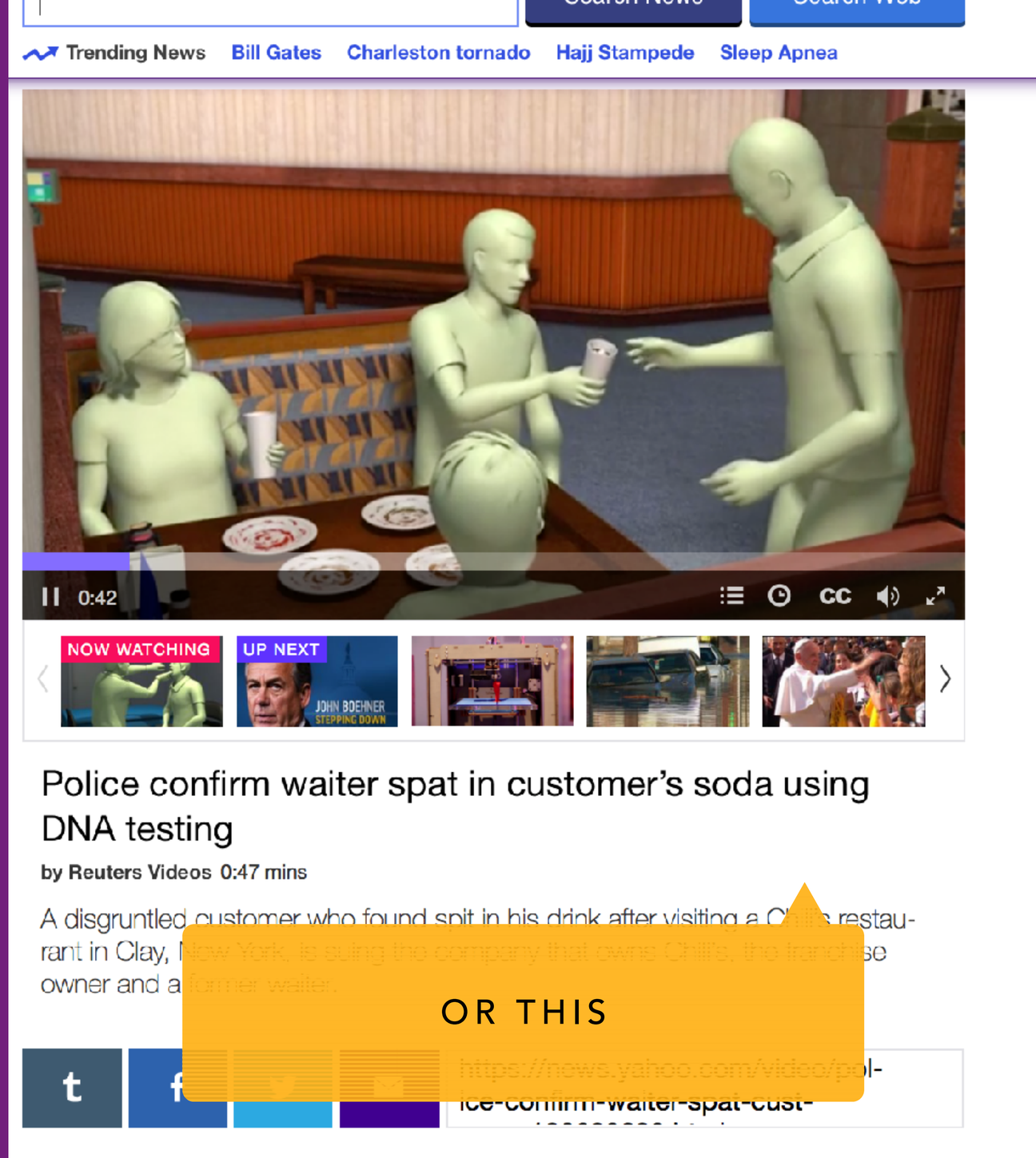
WHAT IS BIOINFORMATICS

- The drive for the next-new-new revolution in bioinformatics
  - Genome sequencing costs
  - Do things that were literally not possible years ago



WHAT IS BIOINFORMATICS

- What can you do with bioinformatics?
  - Really important scientific research that contributes to a better humanity



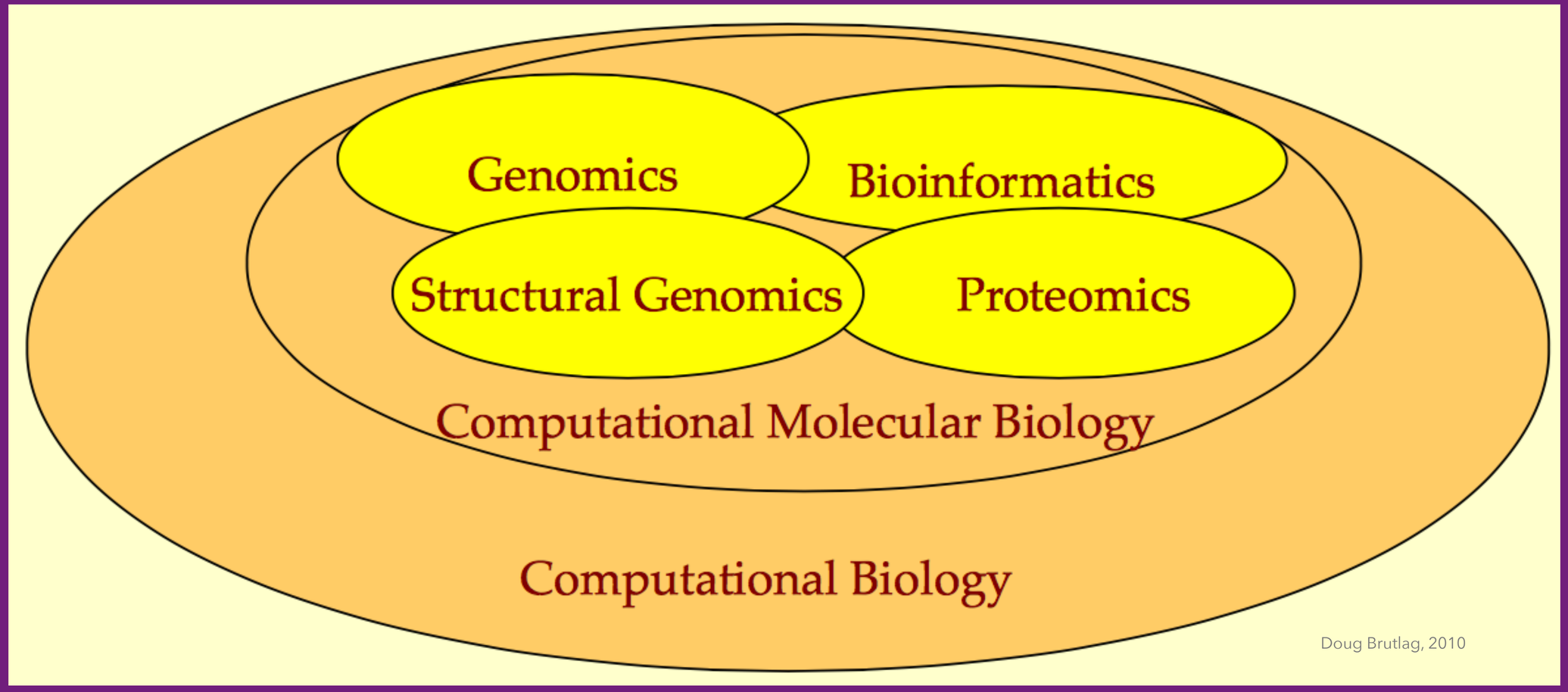
WHAT IS BIOINFORMATICS

WHAT ABOUT "COMPUTATIONAL BIOLOGY"?

INCLUDES BIOINFORMATICS

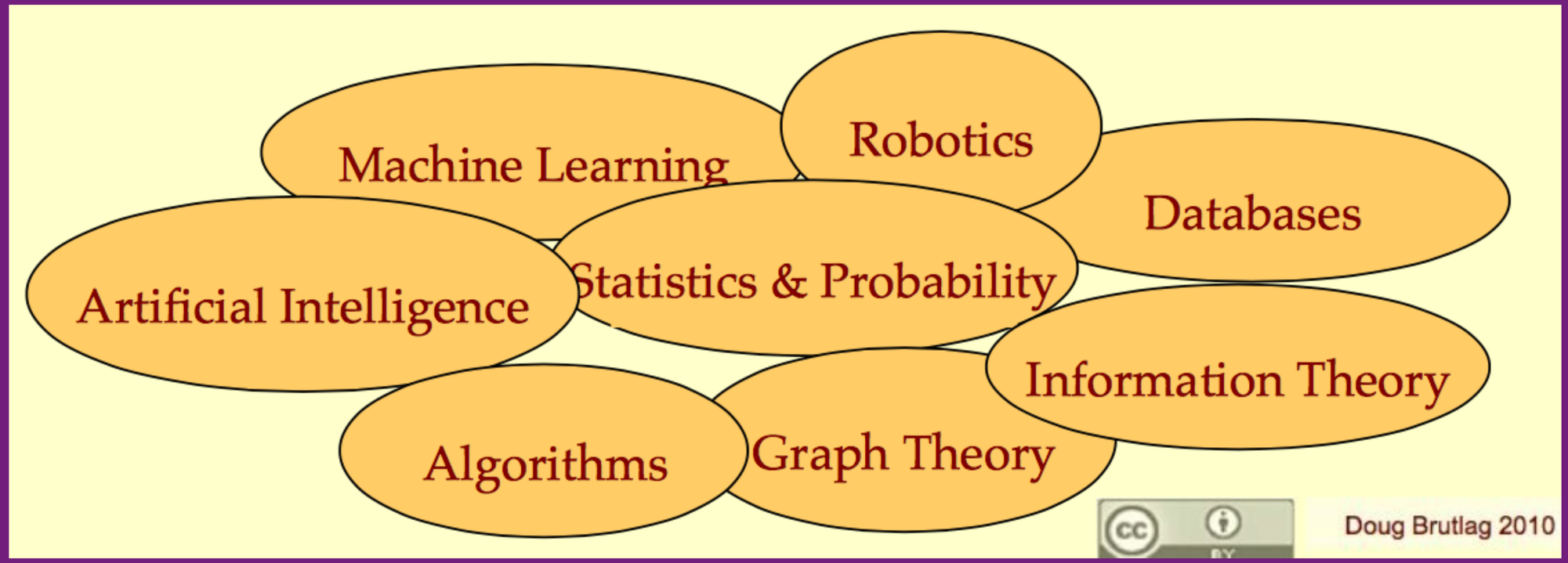
- Computational biology is an interdisciplinary field that applies the techniques to address biological problems
  - Computer science
  - Applied mathematics
  - Statistics
- A broader term (Can be used interchangeably, in my opinion)

WHAT IS BIOINFORMATICS



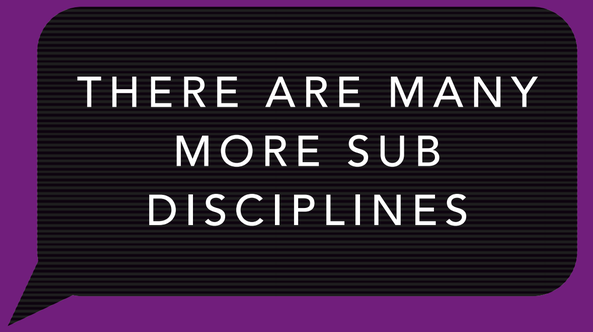
- Using computers to answer biological question

WHAT IS BIOINFORMATICS?



- Computational biology pulls from many subjects for ideas
  - Unconventional ways (mosquito robot, docking game)

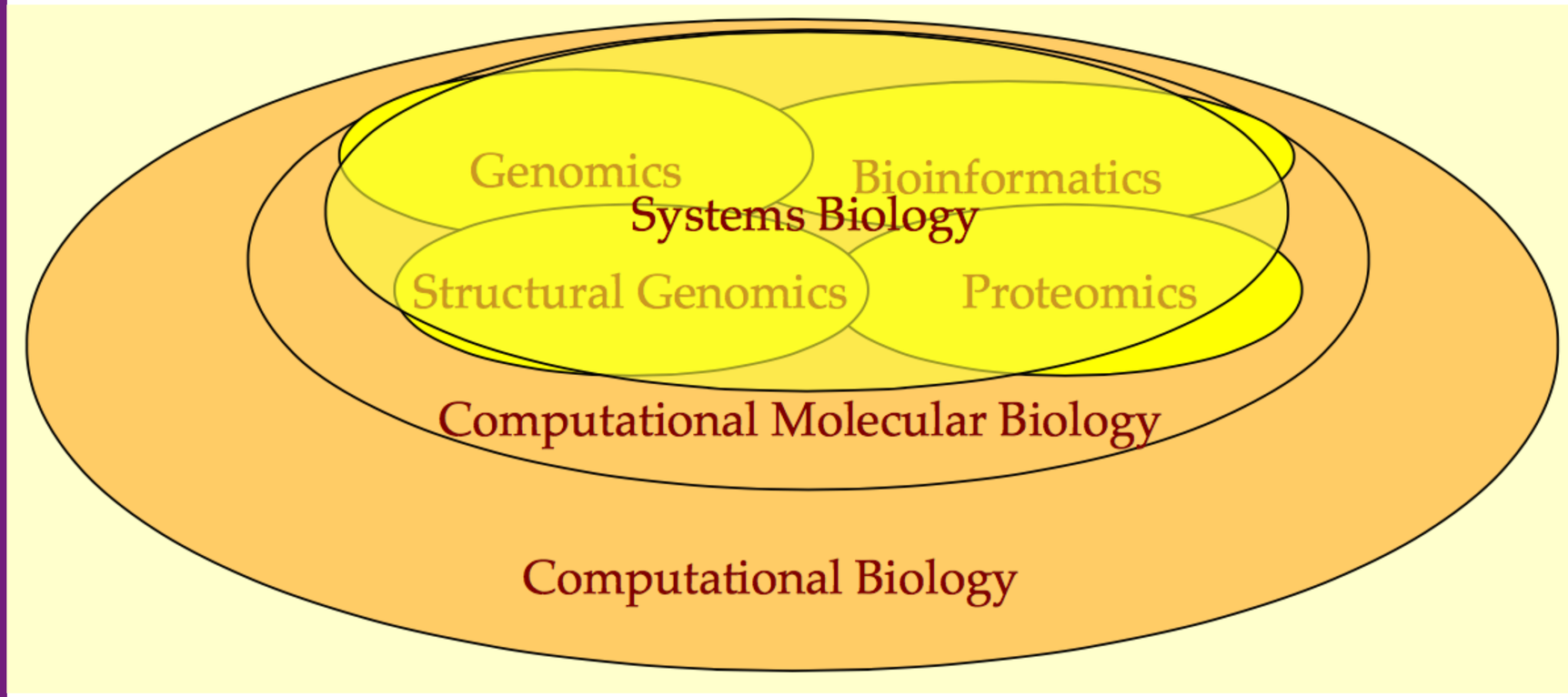
WHAT IS BIOINFORMATICS?



- What about all the “-omics”?
  - Genomics - Discovery, arrangement, expression of genes
  - Proteomics - Study the proteins in a system; mass spec analysis; markers
  - Structural Genomics - Study structural representatives of all proteins
  - Bioinformatics - Studying biological information

WHAT IS BIOINFORMATICS

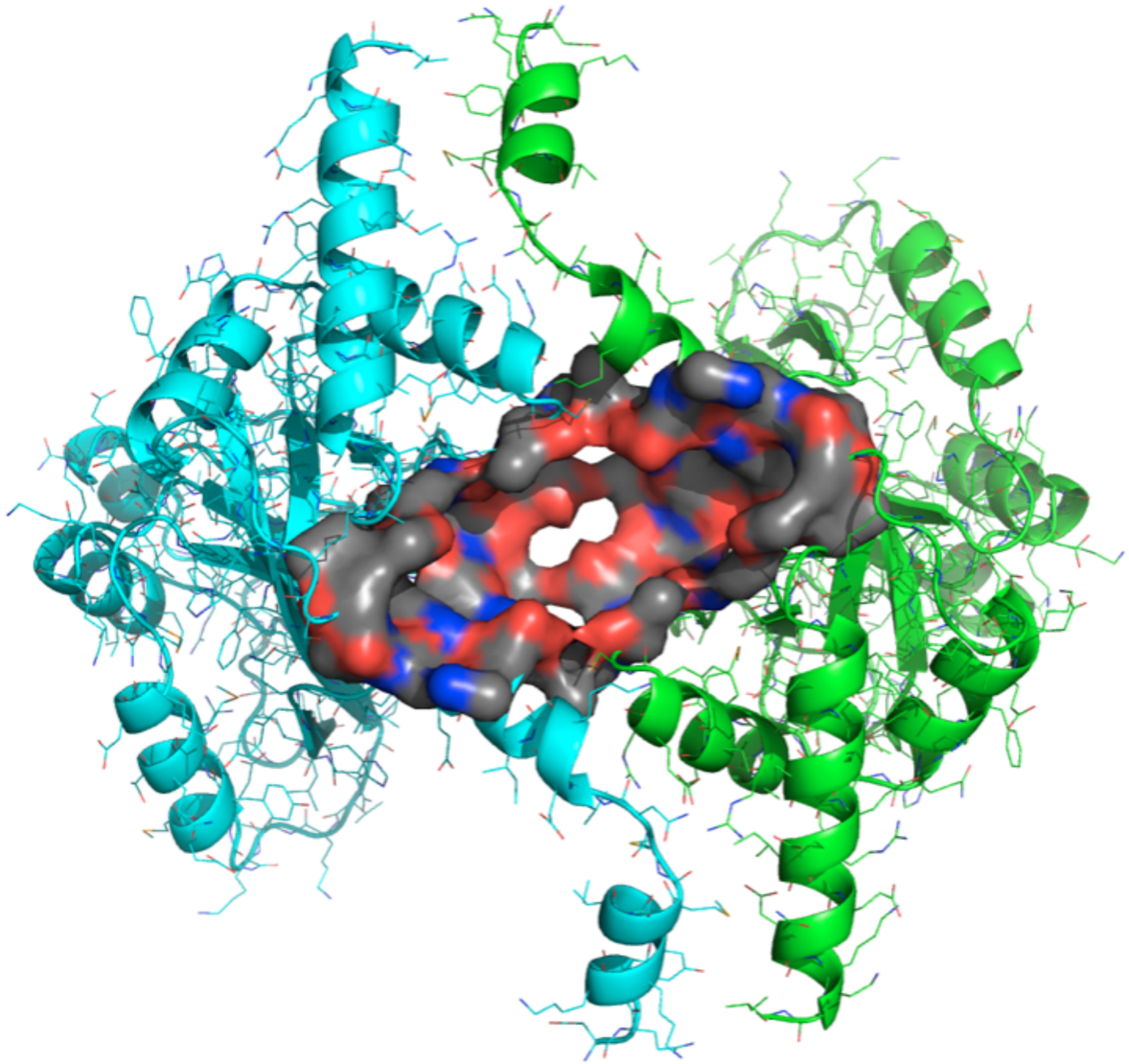
- Systems biology
  - New(er) field uses information to understand how systems work together
  - Enabled by the other fields



Doug Brutlag, 2010

WHAT IS BIOINFORMATICS

- Means different things to different people
  - Population modeling, disease modeling, molecular modeling, etc.
- Spend a career in one small subset
  - Protein surface analysis
- We will try to cover a little bit of everything



Goals of Bioinformatics

GOALS OF BIOINFORMATICS

- Discover
- Predict
- Infer
- Organize
- Integrate
- Simulate
- Engineer

GOALS OF BIOINFORMATICS

- Analysis of genes and proteins
  - Homolog detection
  - Alignment (the residual-level mapping among homologous genes/proteins)
  - Application of the alignments
    - Detect the conserved residues, Functional sites, Prediction of protein structures, Motif finding (cis-elements)
  - Phylogeny
  - Engineer

03.30.18

CRISPR-altered plants are not going to be regulated (for now)

[Photo: dimitrisvetsikas1969/Pixabay]

Good news for people who like genetically altered tomatoes and other plants. The U.S. Department of Agriculture announced it will no longer regulate them.

The USDA not only rolled back Obama-era rules regulating genetically edited plants, but now it claims that plants whose genomes have been altered using gene-editing technology

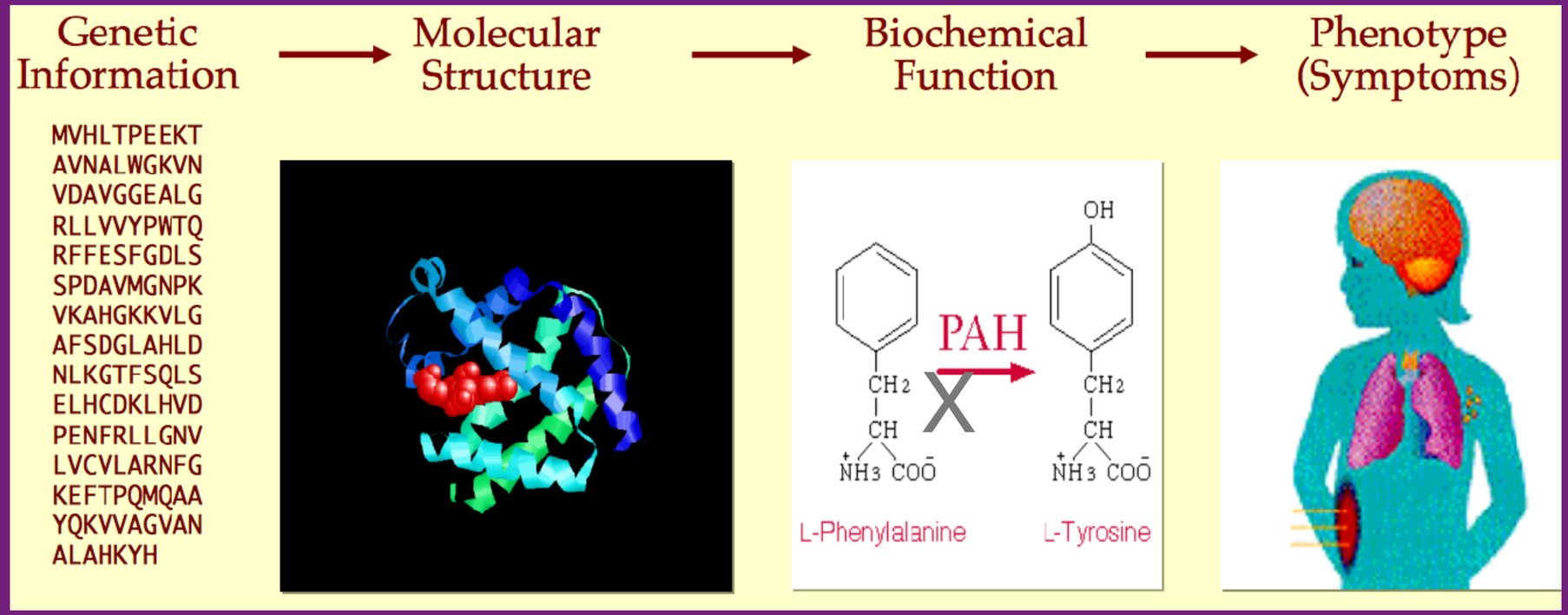
GOALS OF BIOINFORMATICS

LEARN BY COMPARISON AND INFERENCE

- Is Protein A similar to Protein B?
  - Sequence similarity (alignment!)
  - Structure similarity (structural comparison)
  - Co-expression (Microarray data analysis)
  - Any types of correlation (operon-structure, etc)

GOALS OF BIOINFORMATICS

CENTRAL PARADIGM OF BIOINFORMATICS: PREDICT/INFER



- Phenyalaline hydroxylase (PAH) gene adds a hydroxyl group to create tyrosine
  - Tyrosine import to formation of hormones and neurotransmitters
- Mutation in PAH allows toxic buildup of Phenynlalaine
- PKU (Phenylketonuria) - intellectual disability and seizures; low birth weight in babies

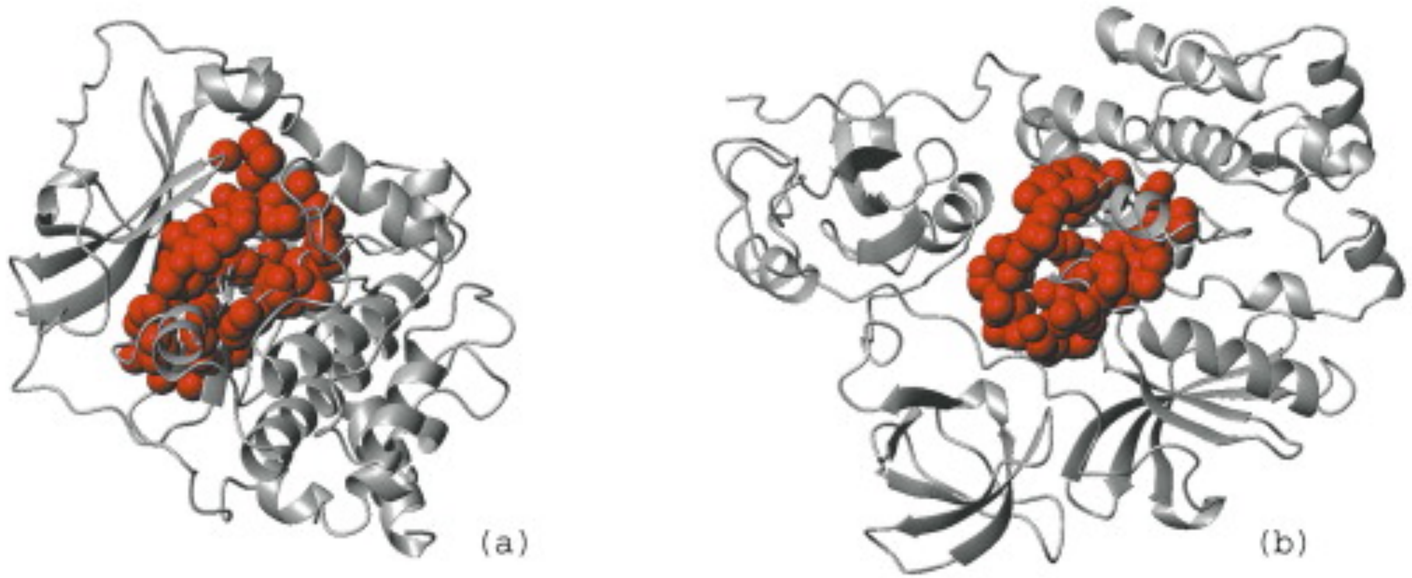
GOALS OF BIOINFORMATICS

- Challenges in bioinformatics (molecular biology)
  - Genetic information is redundant
  - Structural information is redundant
  - Genes and proteins aren’t stable
    - Flexibility important for function
  - Genes have multiple functions
  - Translate 1 dimensional to 3 dimensional data

GOALS OF BIOINFORMATICS

- Challenges: 1D to 3D

IMPORTANT RESIDUES ARE NEIGHBORS IN 3D SPACE, NOT IN 2D



(c)

>1cdk_A

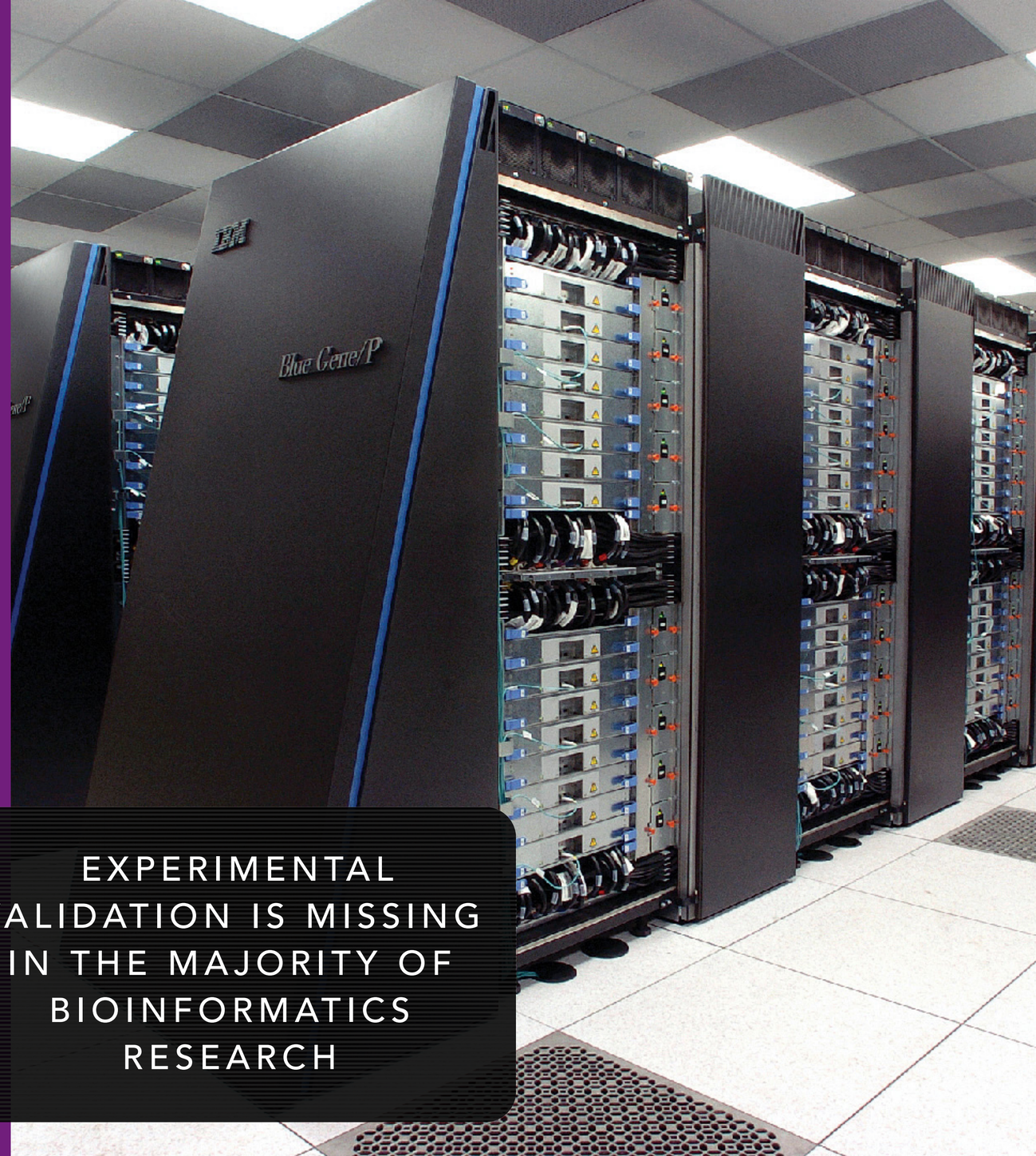
GNAAAAKKGSEQESVKEFLAKAKEDFLKKWENPAQNTAHLDQFERIKTLGTGSFGRVMLVKHKETGNHFAMKILD
KQKVVKLKQIEHTLNEKRILQAVNFPFLVKLEYSFKDNSNLYMVMEYVPGGEMFSHLRRIGRFSEPHARFYAAQI
VLTFEYLHSLDLIYRDLKPENLLIDQQGYIQVTDFGFAKRVKGRTWTLCGTPEYLAPEIILSKGYNKAVDWWALG
VLIYEMAAGYPPFFADQPIQIYEKIVSGKVRFPSHFSSDLKDLLRNLLQVDLTKRFGNLKDGVNDIKNHKWFATT
DWIAIYQRKVEAPFIPKFKGPGDTSNFDDYEEEEIRVSINEKCGKEFSEF

>2src_

MVTTFVALYDYESRTETDLSFKKGERLQIVNNTEGDWWLAHSLSTGQTGYIPSNYVAPSDSIQAEEWYFGKITRR
ESERLLLNAENPRGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFNSLQQLVAYYS
KHADGLCHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTRVAIKTLKPGTMSPEA
FLQEAQVMKKLRHEKLVQLYAVVSEEPIYIVTEYMSKGSLLDFLKGETGKYLRLPQLVDMAAQIASGMAYVERMN
YVHRDLRAANILVGENLVCKVADFGLARLIEDNEYTARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELT
TKGRVPYPGMVNREVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPTFEYLQAFLEDYFTSTEPQXQPGE
NL

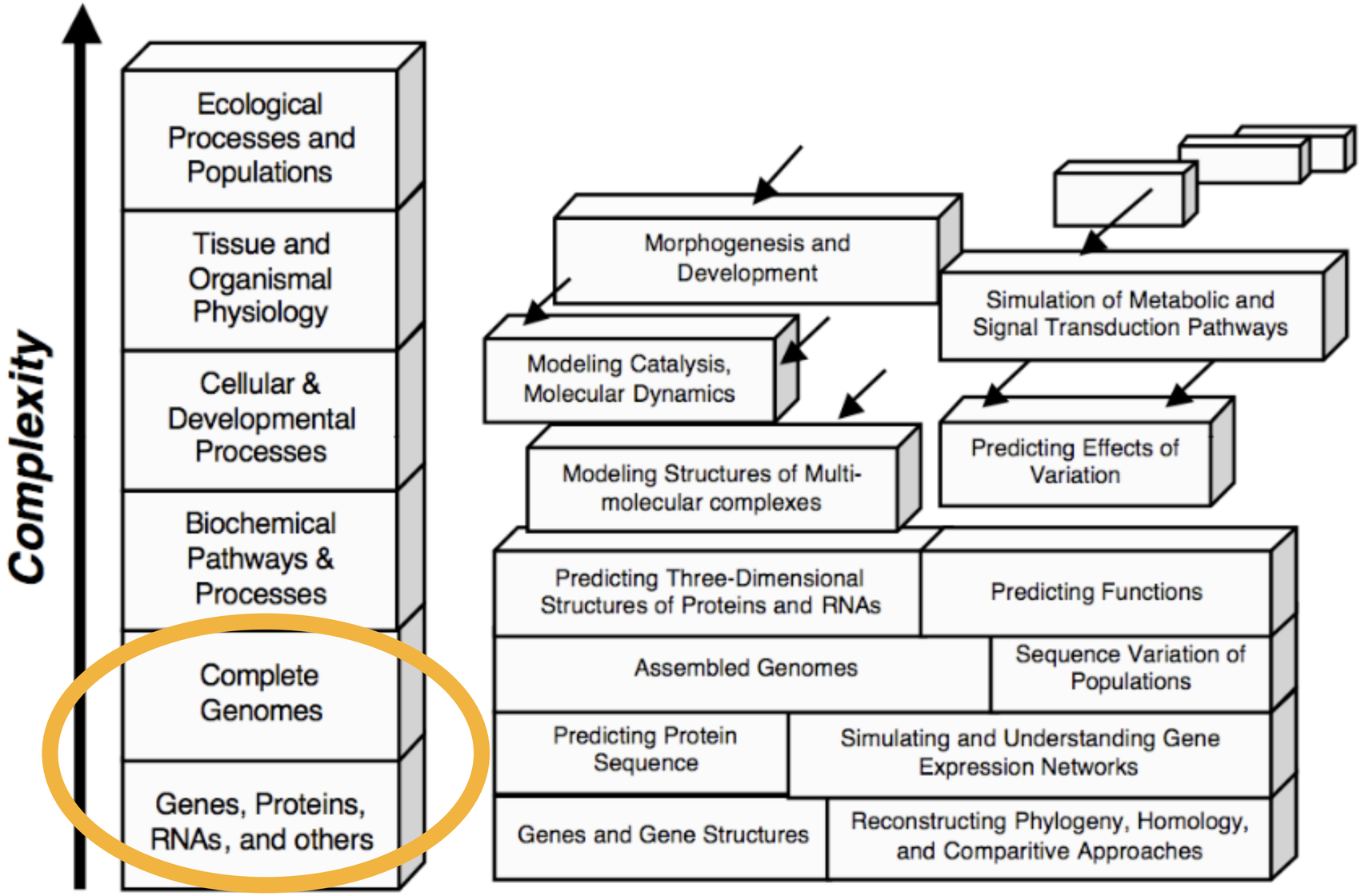
GOALS OF BIOINFORMATICS

- Challenges in bioinformatics (computing)
  - Growth, complexity and volume of biological data
  - Propagation of errors
    - Problem of inference
  - Economics of research
    - Computation is cheap (and doesn't go on vacation)
    - People are expensive
    - Wet lab work is expensive



EXPERIMENTAL VALIDATION IS MISSING IN THE MAJORITY OF BIOINFORMATICS RESEARCH

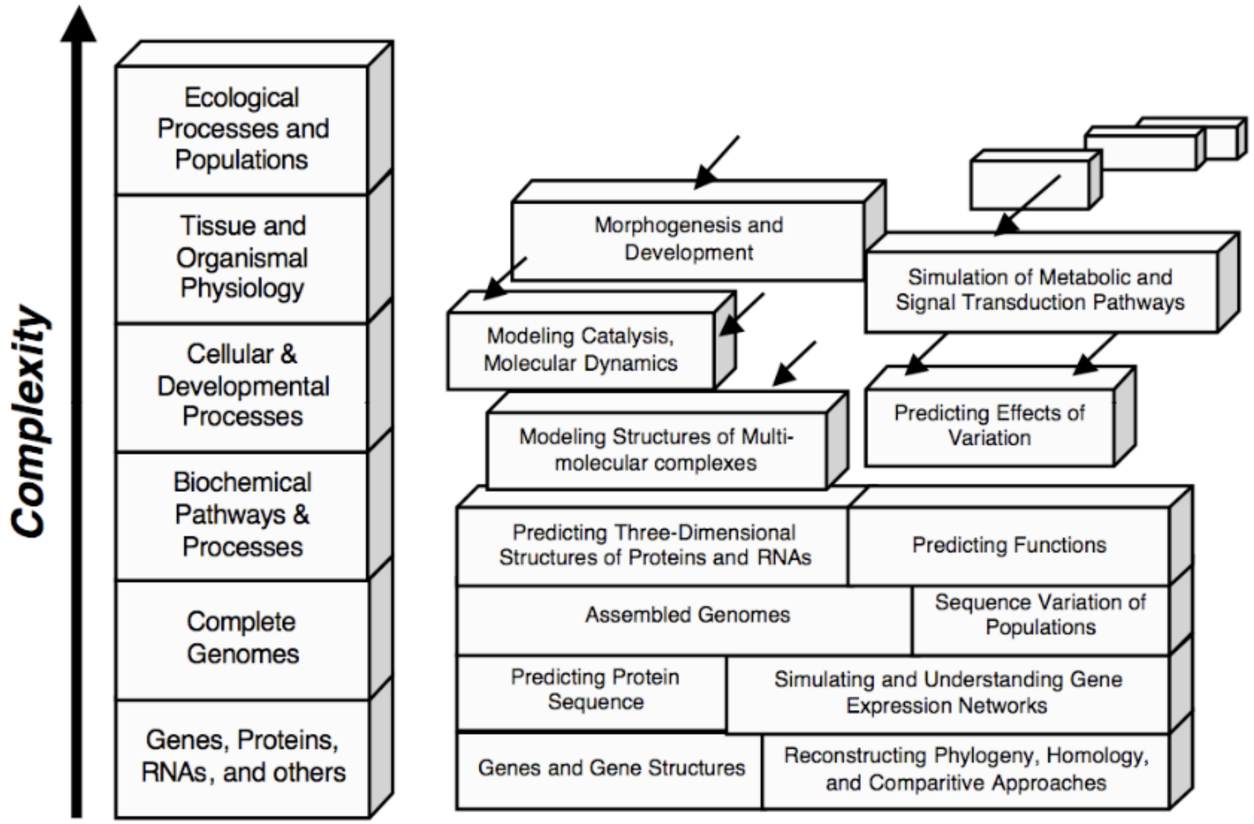
GOALS OF BIOINFORMATICS



Impact of Bioinformatics

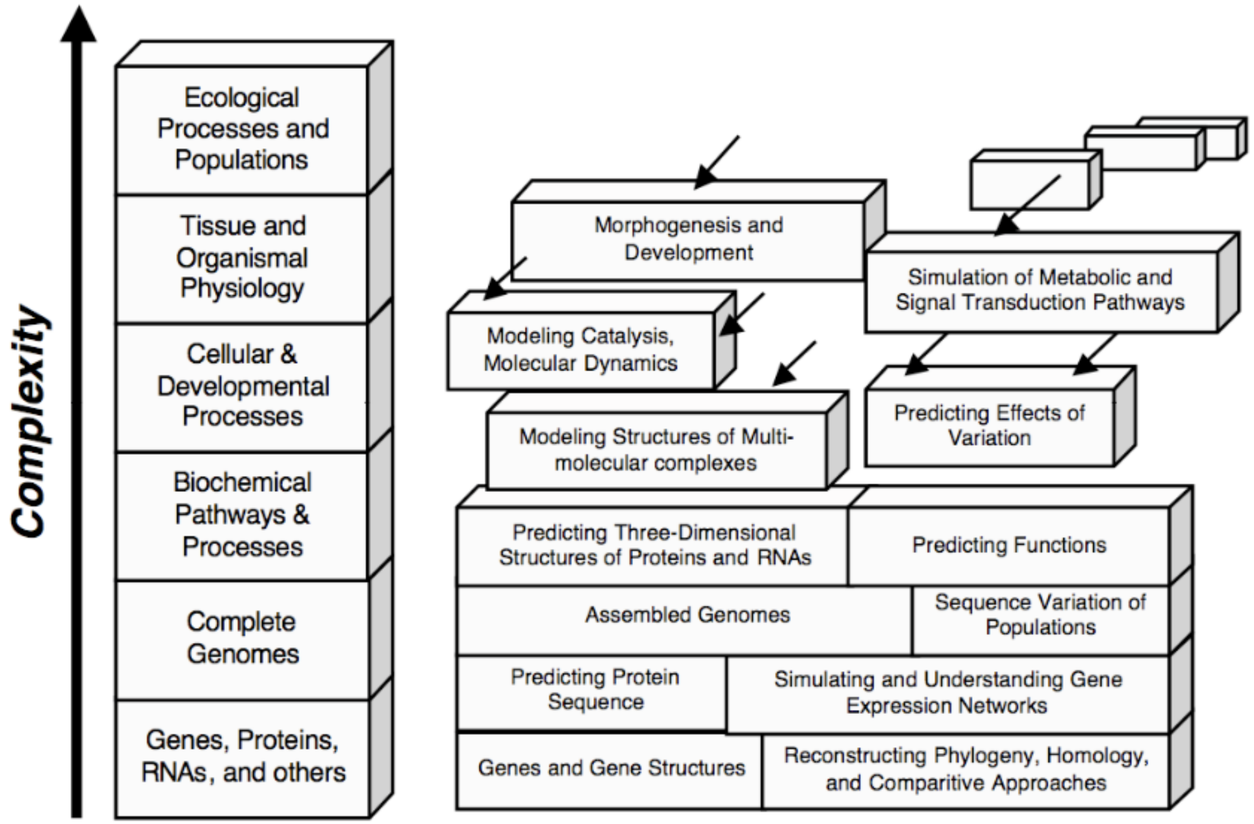
IMPACT OF BIOINFORMATICS

- On Biological sciences
  - Large scale experimental techniques
    - Analyze an entire genome
  - Information growth
  - Integration of data between fields



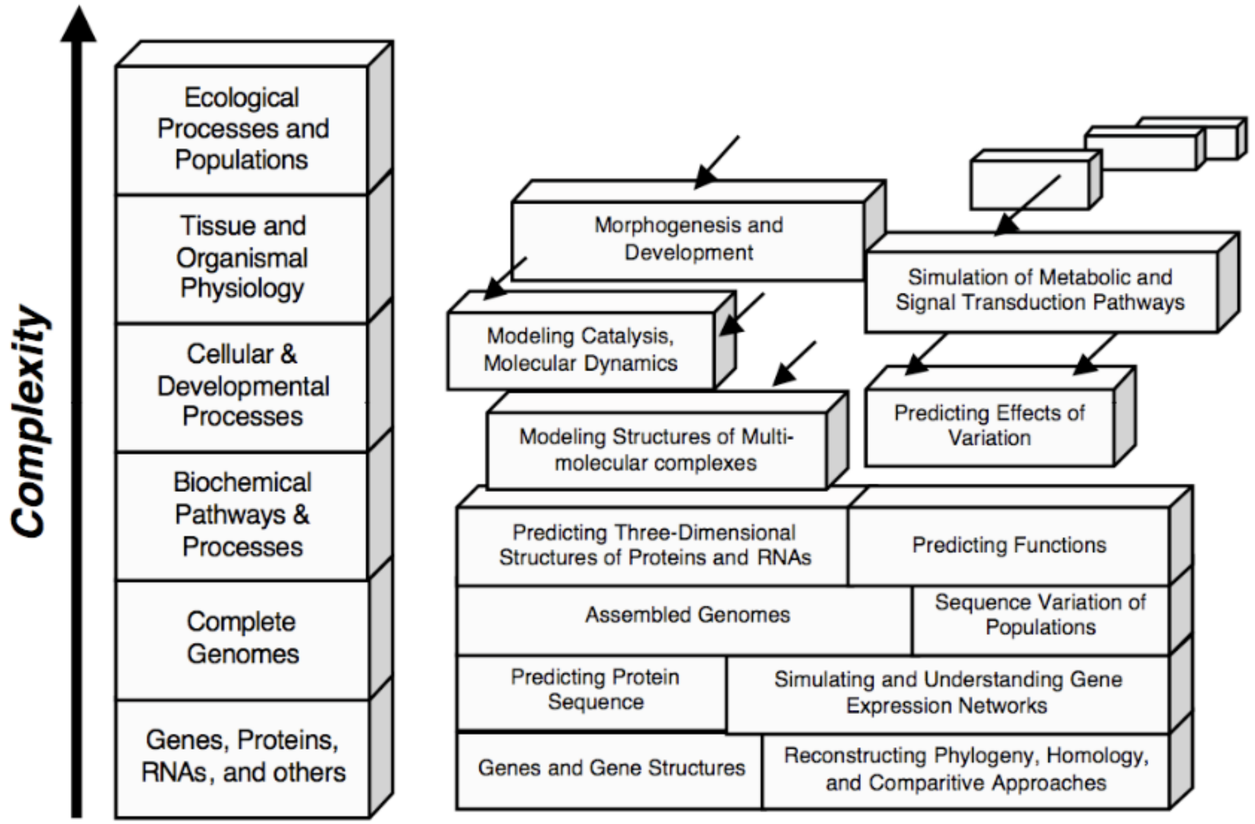
IMPACT OF BIOINFORMATICS

- On computational sciences
  - New algorithmic and statistical problems
  - Big data



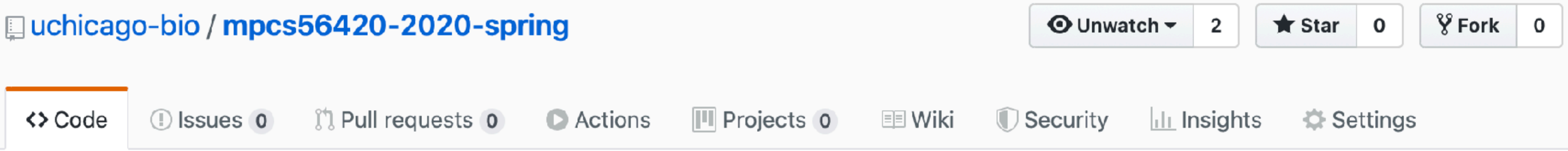
IMPACT OF BIOINFORMATICS

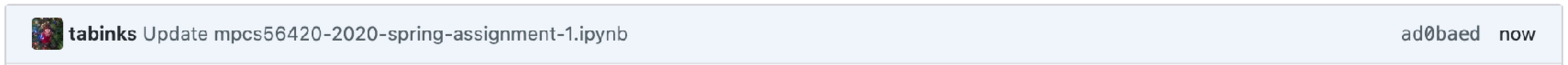
- On Medical sciences
  - Translational research
  - Personalized treatment

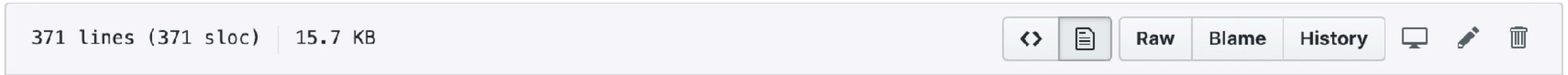



ASSIGNMENT

ASSIGNMENT 1







MPCS 56420 - 2020 - Spring - Assignment 1

This assignment is due Thursday, April 16, 2020 at 5:29 PM. The instructors will clone a copy of your assignment repository and use the last version checked in before the due date. Please answer immediately below each question.

Github Classroom Repo: https://classroom.github.com/a/ro8lKKH2 *If you have not done so, please DM the instructors your GitHub account user name as soon as possible.*

HTTPS://GITHUB.COM/UCHICAGO-BIO/MPCS56420-2020-SPRING/BLOB/MASTER/ASSIGNMENT-1/MPCS56420-2020-SPRING-ASSIGNMENT-1.IPYNB