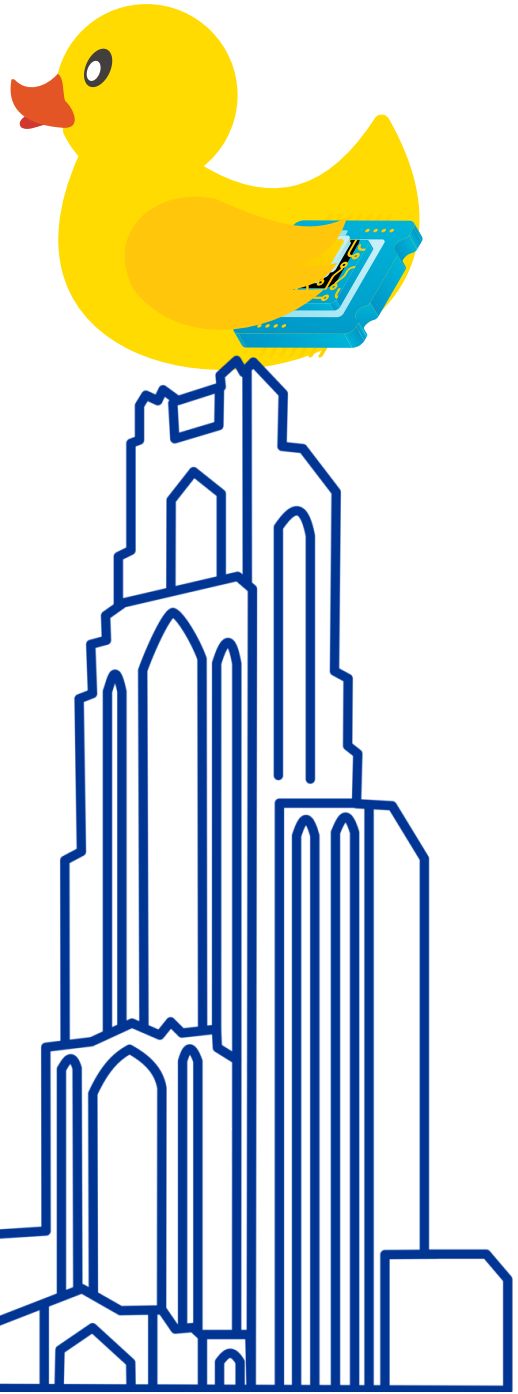


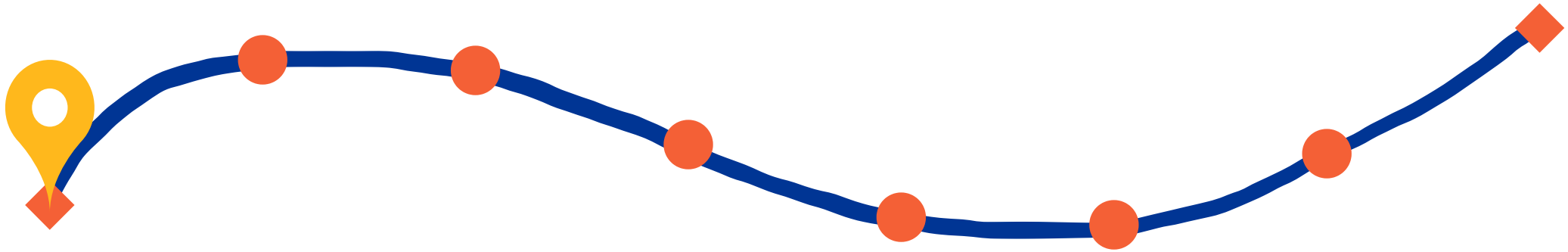
Computational Biology
(BIOSC 1540)

Lecture 01
Computational biology overview

Jan 9, 2025

University of Pittsburgh

After today, you should be able to



Identify your teaching team

Meet your instructor

Alex Maldonado, PhD

he/him

B.S.E in Chemical Engineering, 2018
Western Michigan University

Ph.D. in Chemical Engineering, 2023
University of Pittsburgh

Office hours: Tuesdays from 1:00 - 2:00 pm in L10 Clapp Hall

Email: alex.maldonado@pitt.edu

Position: Postdoctoral Associate

Acceptable ways to address me: **Alex** (preferred) Dr. Maldonado Dr. M

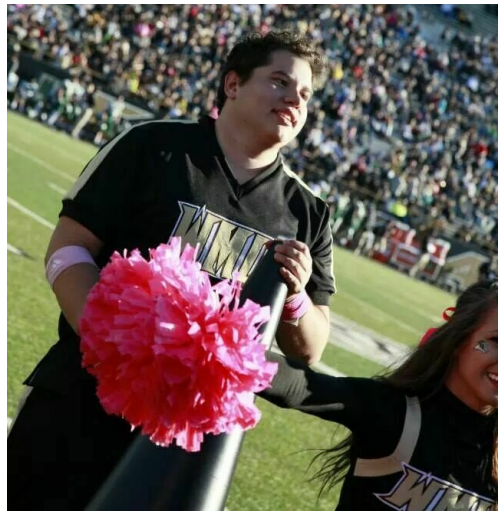
My website

aalexmmaldonado.com

https://aalexmmaldonado.com/

Alex's fun facts

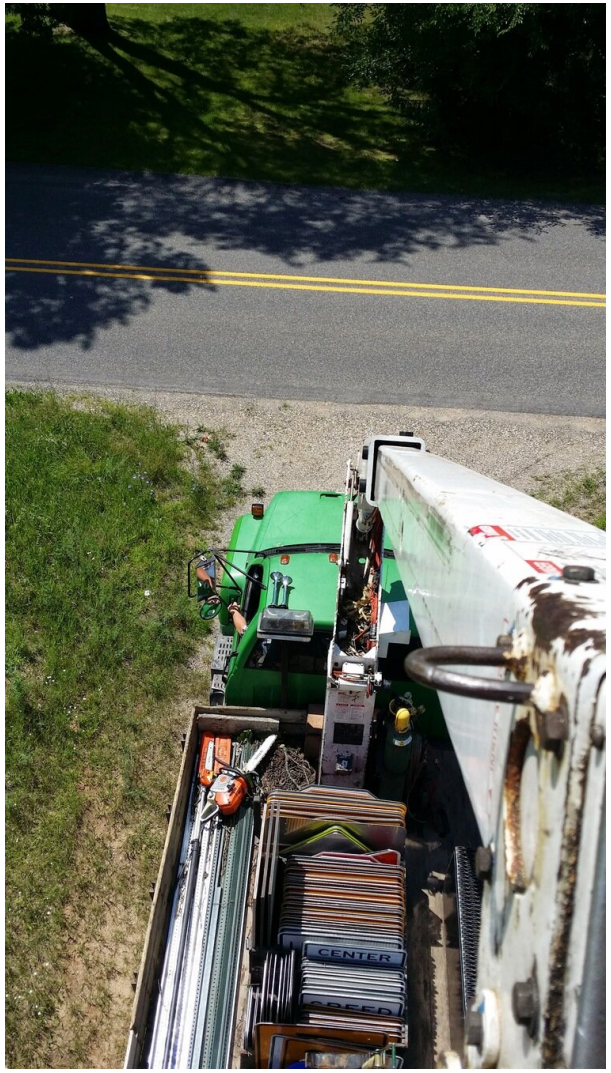
Every male in my (maternal) family played football —I rebelled



Alex's fun facts

Part-time jobs during college

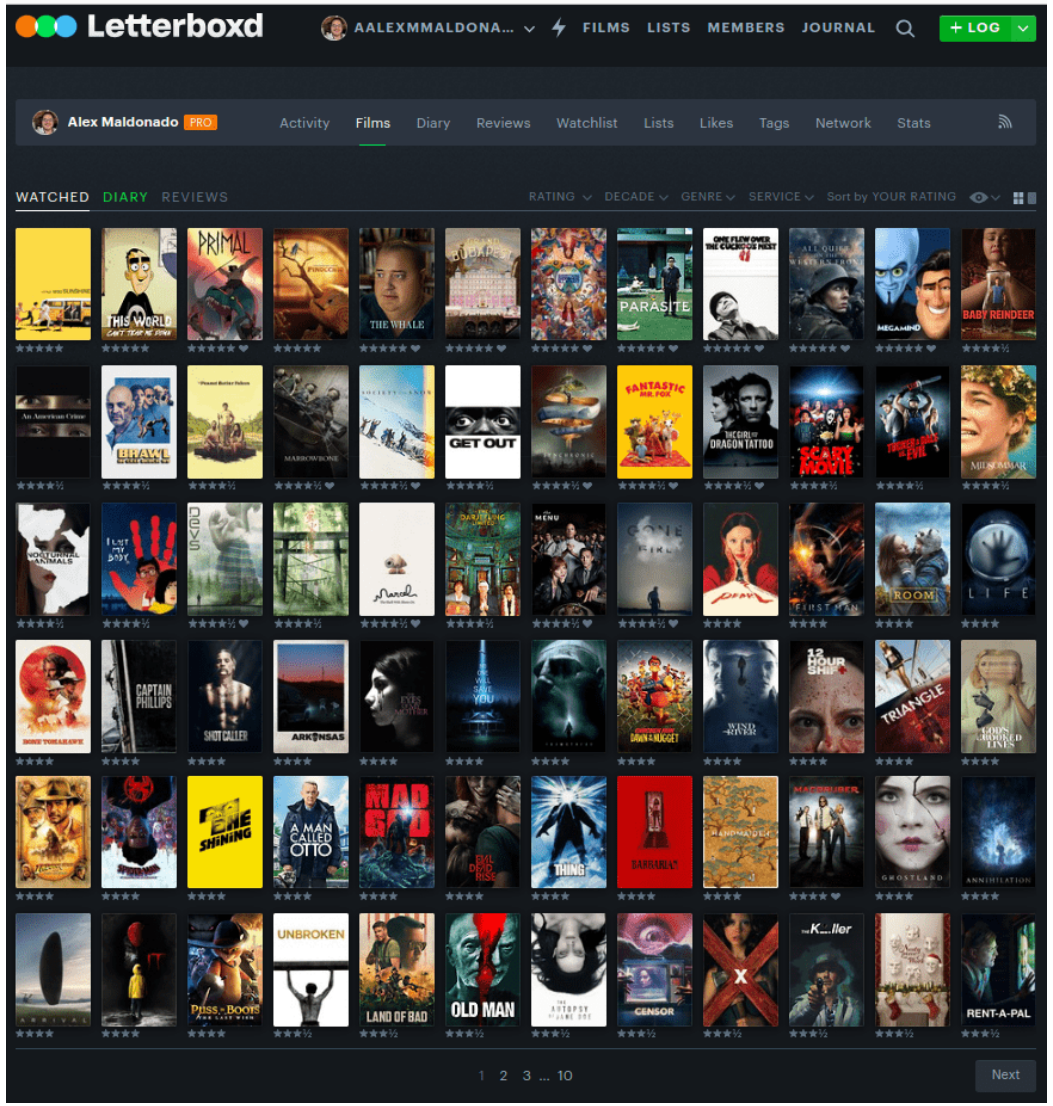
- Construction
- UPS package handler
- Kent County Traffic safety
- Jimmy John's delivery driver
- Wings West ice events



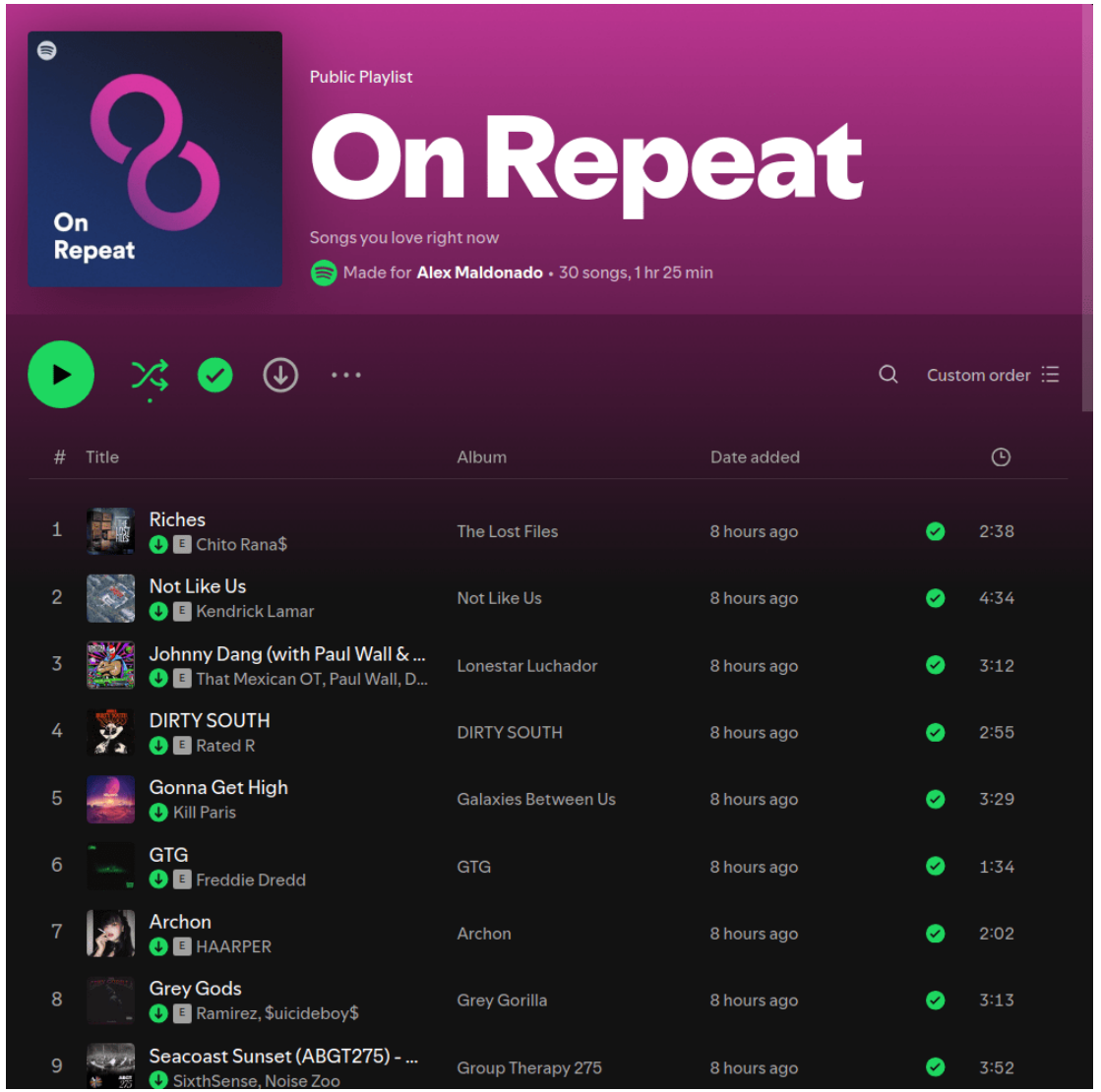
Tessa the Princessa

Get to know my ...

Movie taste



Music taste



Meet your teaching assistants

Mondays

11:00 am to 12:00 pm
in 102 Clapp

Justine Denby (she/her)

Rushali Patel (she/her)

Wednesdays

1:00 to 2:00 pm
in L1 Clapp

Priyam Chauhan (she/her)

Jay Grimsdall (he/they)

Thursdays

3:00 - 4:00 pm
in L10 Clapp

Caelyn Peppler (Any)

Mariska Goswami (she/her)

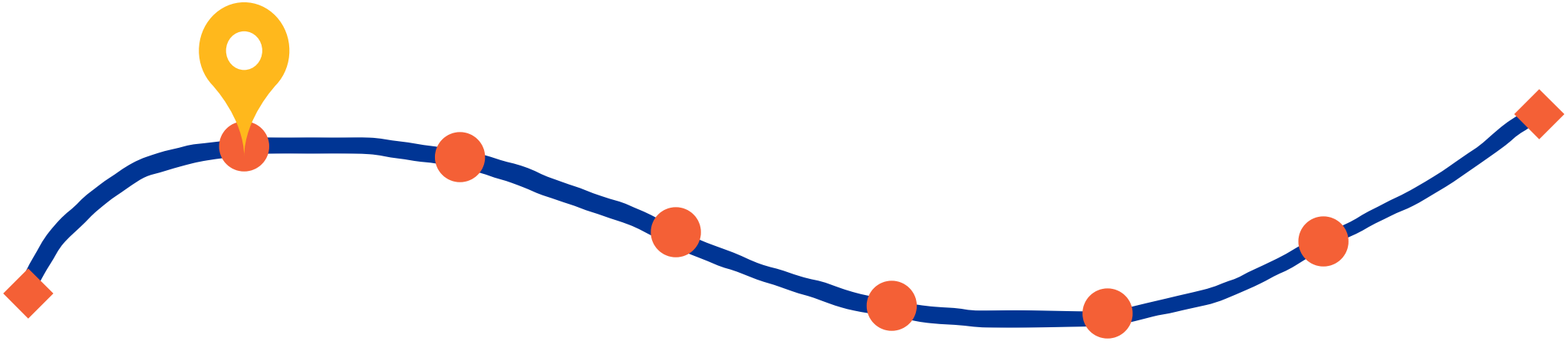
CByte UTA

César Guerra-Solano (he/him)

(More on this later)

More information

After today, you should be able to



Describe the course structure, expectations,

and available resources for success

Single source of truth

All course materials will be posted on this website: pitt-biosc1540-2025s.oasci.org/

Why? There are few comprehensive resources for this rapidly changing field

Things that contain student information will be only on Canvas to be FERPA compliant

Assignments will be submitted on Gradescope

https://pitt-biosc1540-2025s.oasci.org/

Assessments and grade distribution

We will have …

- Four 15-minute quizzes (28%)
- 13 project-based assignments (72%)

Rationale:

(1) Hands-on projects are key for mastering material.
(2) Quizzes prove your comprehension without outside help.

Minimum quiz average:

To pass the course with a C or higher, **your quiz average must be at least 73%**. If your quiz average is below 73%, your overall course grade will be capped at a C–, regardless of your project grades.

Rationale:

This reduces the impact of quizzes on your grade while still requiring that you understand the material

Late penalties

We have a forgiving late penalty for a few hours but it rapidly increases after 12 hours

$$\% \text{ Penalty} = 0.01(2.083 \times \text{hours late})^2$$

Hours late	Penalty
6	1.6%
12	6.2%
24	25.0%
36	56.2%
48	100.0%

Typically, it is in your best interests to take a few more hours to do your best work

Semester overview

Modules

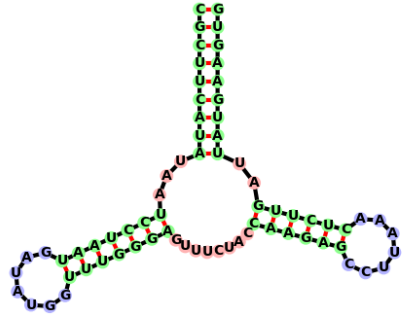
Bioinformatics

Computational Structural Biology

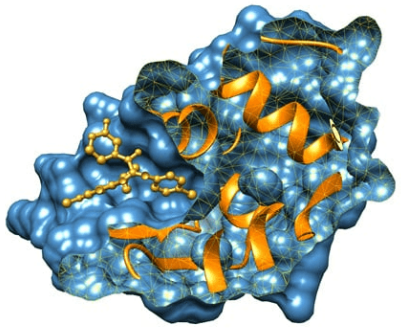
Where do we get our insight from?

1. Genomics

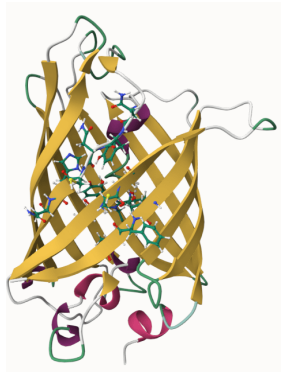
2. Transcriptomics



3. Computer-aided drug design



4. Molecular simulations



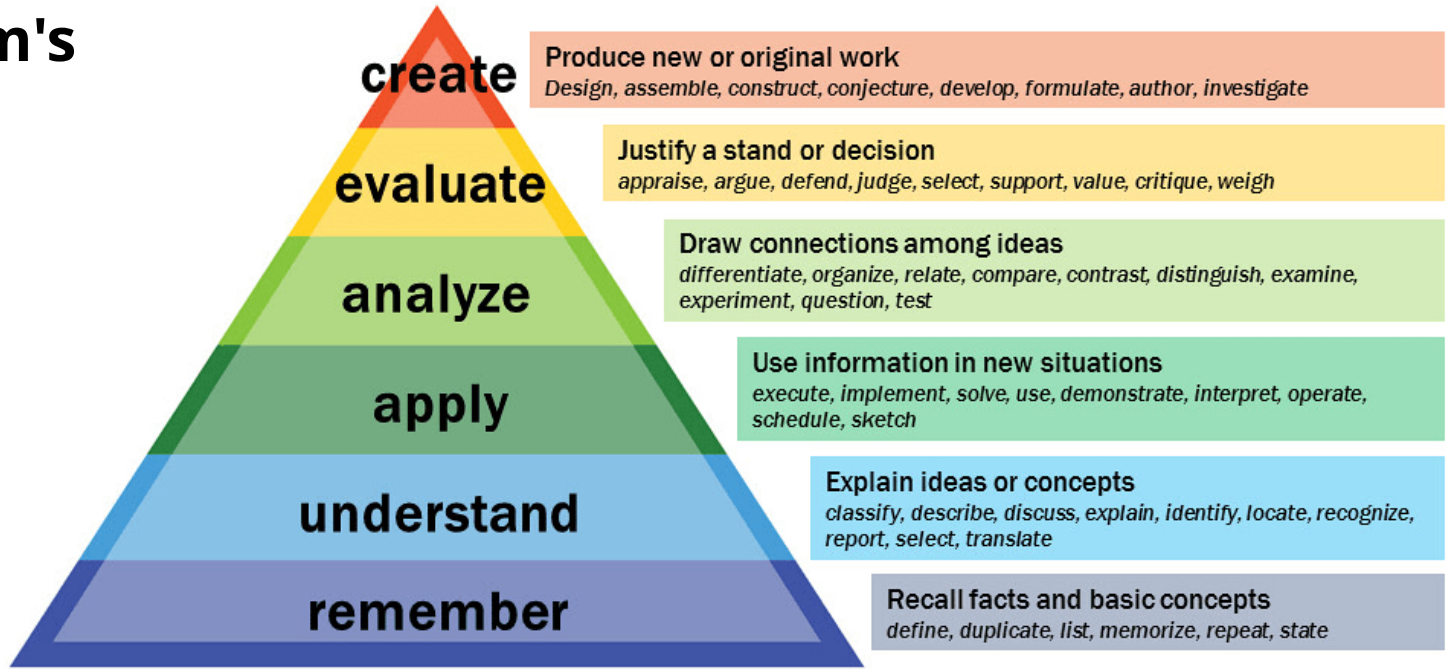
Critical thinking is paramount and happens outside your comfort zone

How does this influence my teaching?

I primarily **focus on the top of Bloom's taxonomy**, more akin to computer science and engineering courses

Challenging problems are worth fewer points to encourage creative problem solving

Few points



Many points

Python will be necessary to successfully complete this course

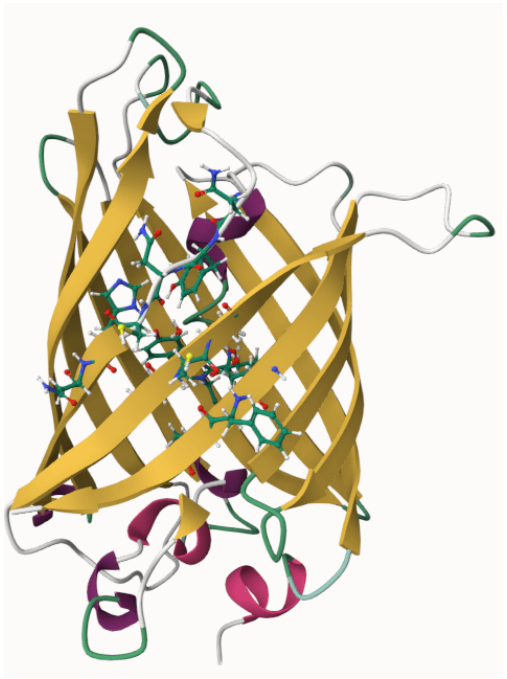
Student feedback

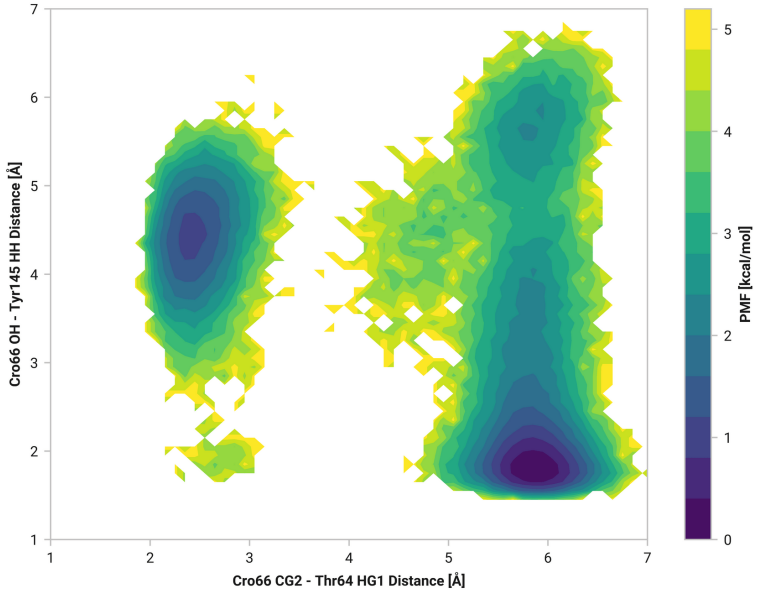
- Past students found programming challenging but rewarding, emphasizing its relevance to career goals and practical applications.
- Python will be introduced alongside simplified problem sets and examples to address varied skill levels, ensuring inclusivity and accessibility.
- I have also trimmed down the material we cover to make room

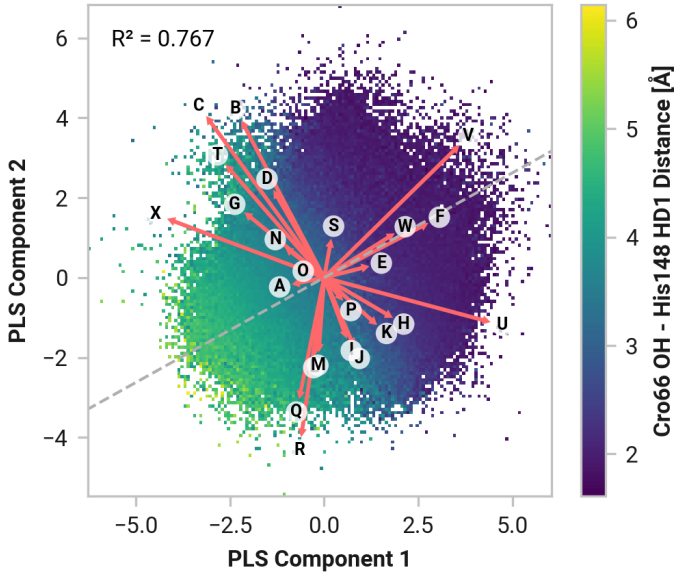
Why Python? Python underpins key methods like molecular simulations, gene expression analysis, and structural prediction. Practical programming skills are foundational for solving real-world problems in this field.

Python is absolutely necessary for a career in computational biology

Programming is how you manage and analyze data





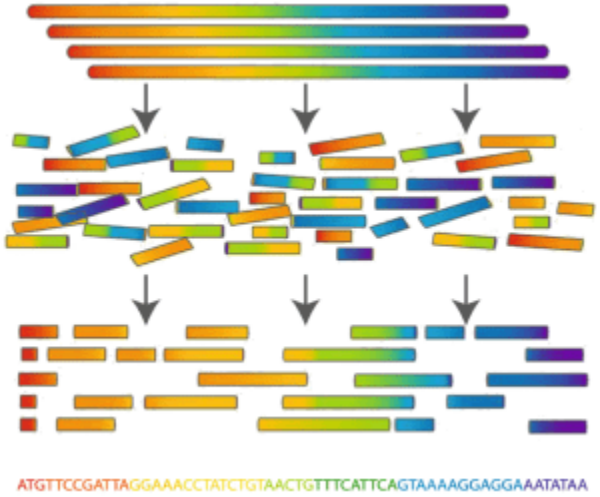


Data

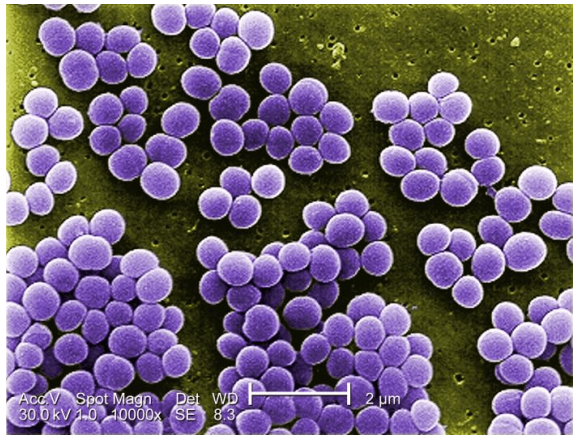
Results and insights

Projects: Antibiotic discovery

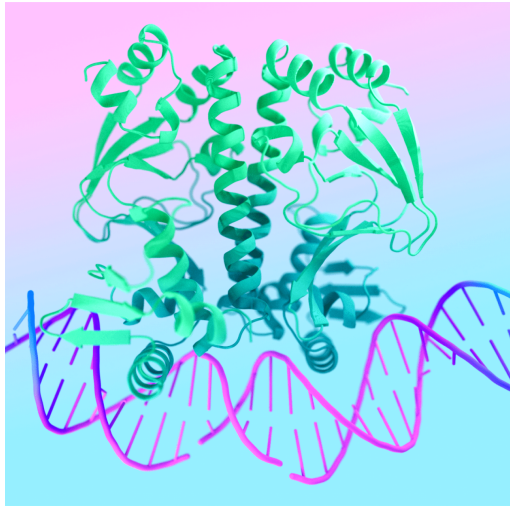
You will complete hands-on projects mirroring the steps to discover novel antibiotics



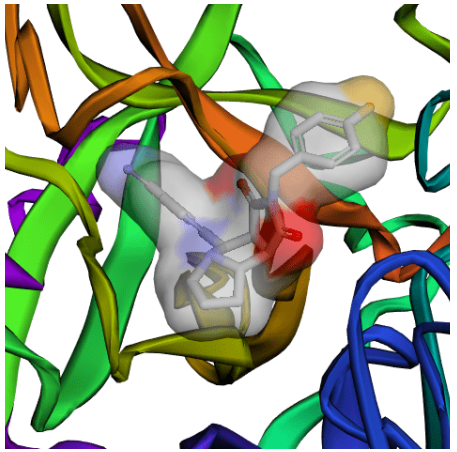
Identify mutations conferring antibiotic resistance



Gene expression and biofilm changes under microgravity



Elucidate protein structure and prediction differences



Virtually screen for potential novel antibiotics

Computational Bytes (CBytes) are optional, bite-sized programming challenges tailored to computational biology

César Guerra-Solano was awarded the Chancellor's Undergraduate Teaching Fellowship to develop these CBytes

Objective: Encourage you to interact more deeply with the course material without a direct impact on your grade

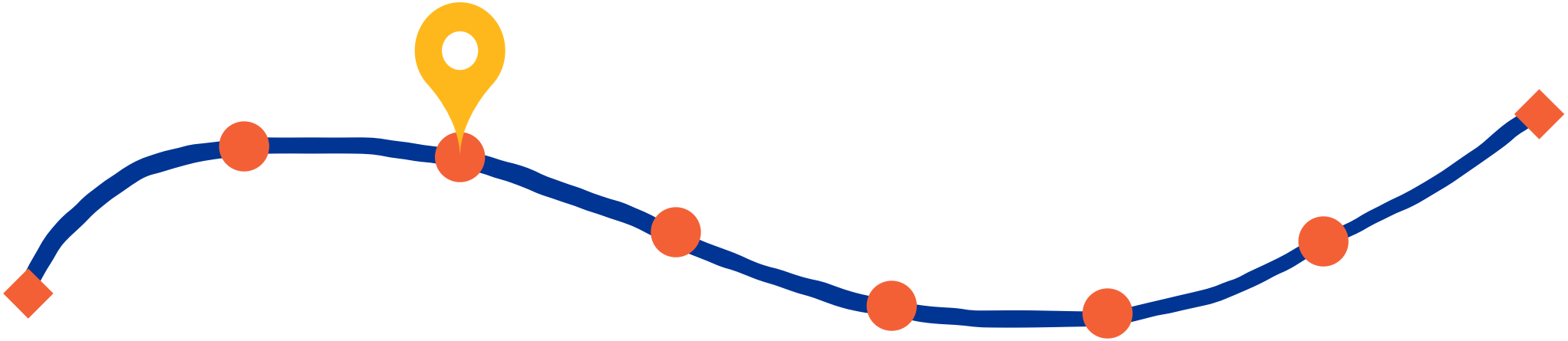
Gamification and incentives: Gradescope autograder will be used to award "Advanced Training Points" (ATP) to students who participate within two weeks of a CByte's release

Rewards: Cumulative ATP can be used to redeem class-wide rewards. For example, everyone can drop an assignment or quiz or extend a deadline.

First one will be released **Jan 17th**

(See Advent of Code.)

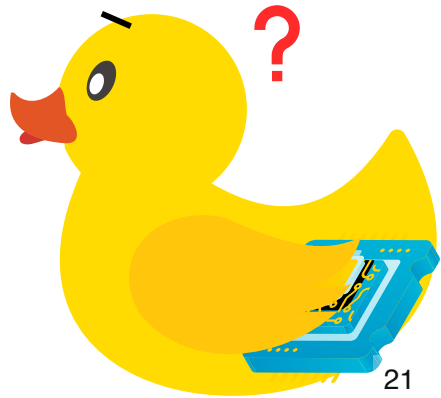
After today, you should be able to



Define computational biology and

explain its interdisciplinary nature

What is computational biology?



What is computational biology?

My definition . . .

Any application of computational methods to obtain insight into biological phenomena.

My main categories . . .

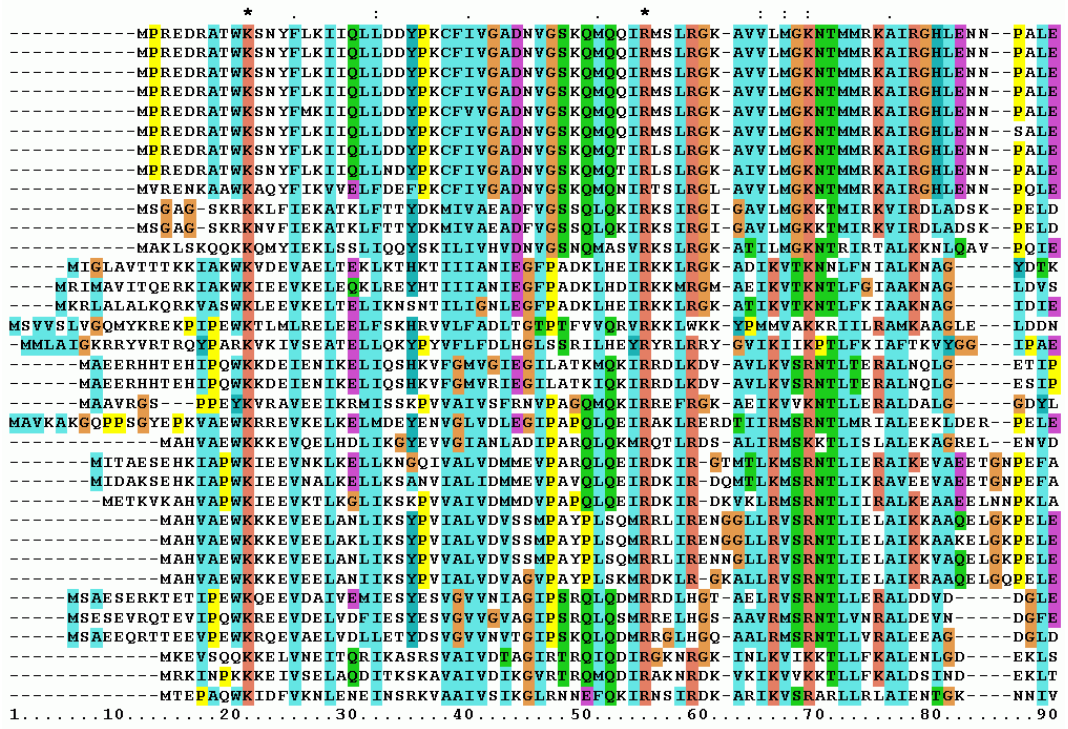
Bioinformatics

Computational structural biology

Bioinformatics deals with untangling big data for biological insights

Data

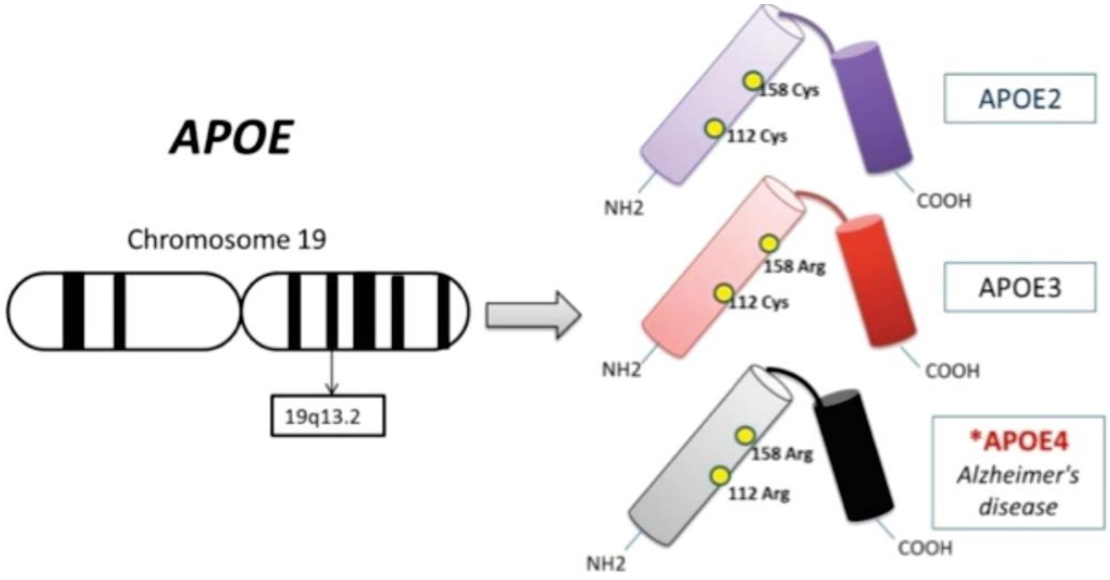
Genetic sequences of healthy and Alzheimer patients



Credit

Information

Find genetic risk factors

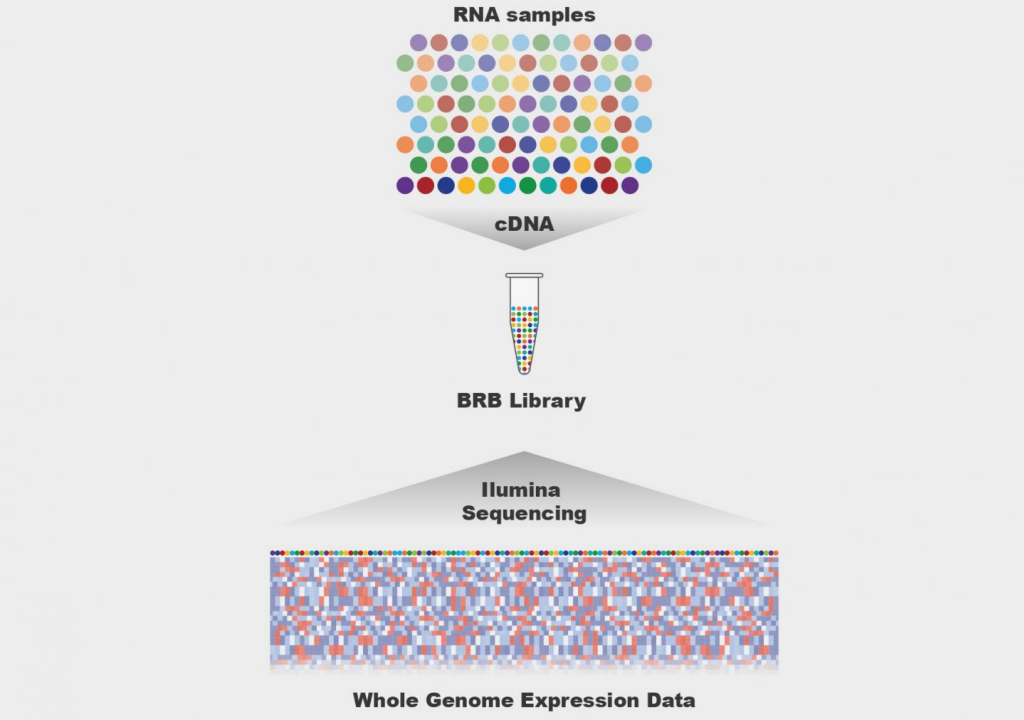


Credit

Bioinformatics deals with untangling big data for biological insights

Data

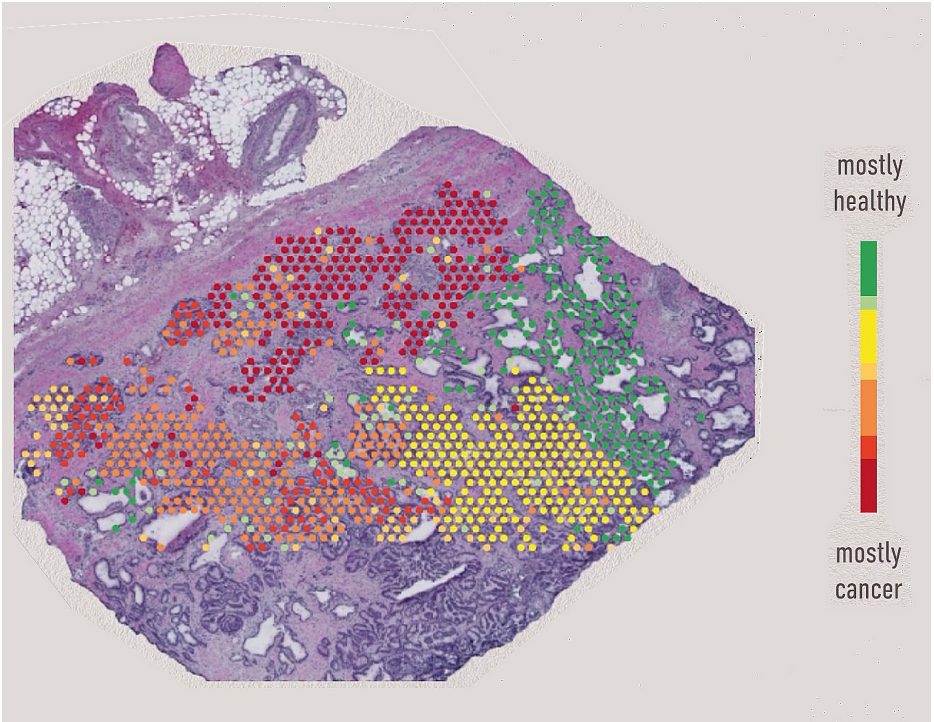
mRNA of cancer cells in a tumor



Credit

Information

Early detection of benign to cancerous cell transition

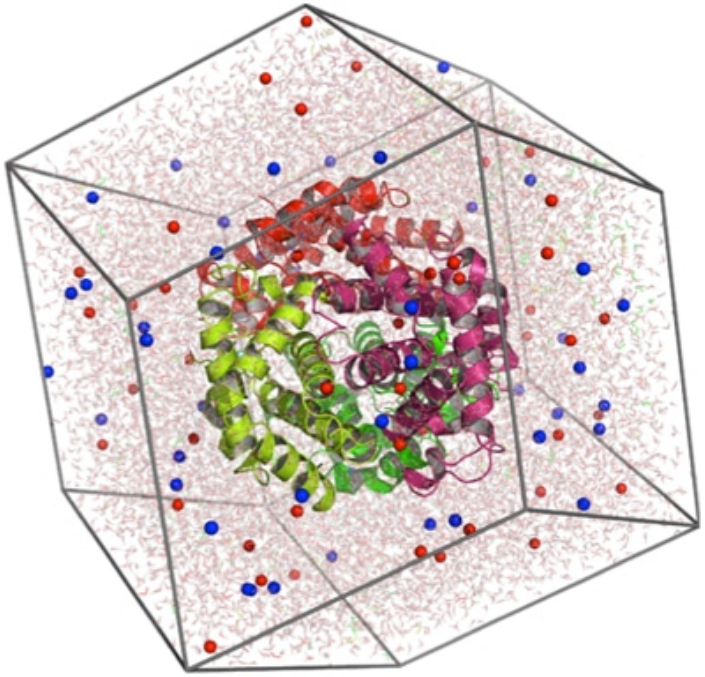


Credit

Modeling employs physical representations that mimic key biological phenomena

Phenomena

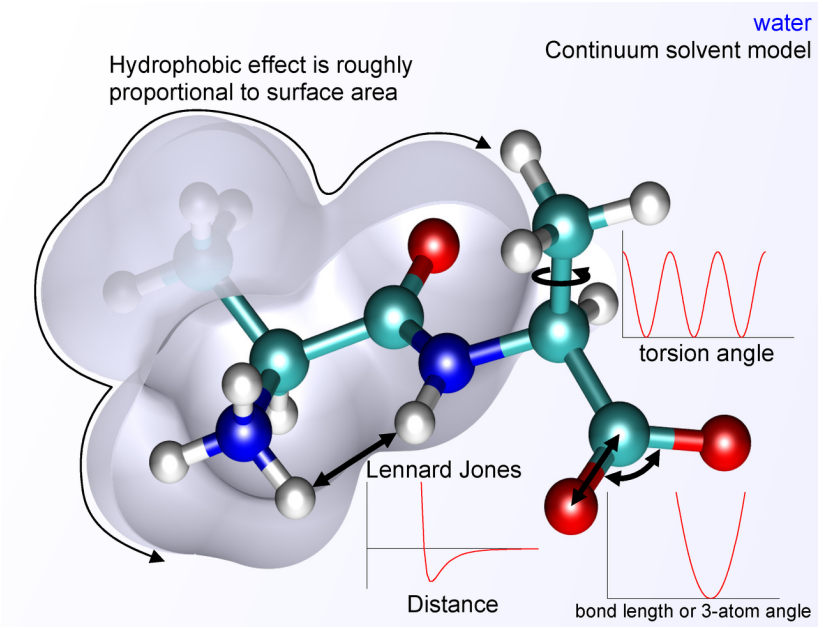
Protein-protein binding



Credit

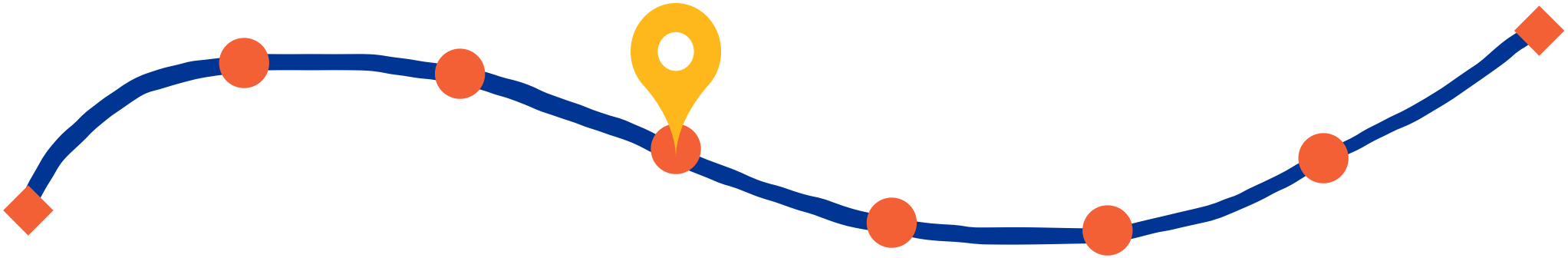
Representation

Classical force fields



Credit

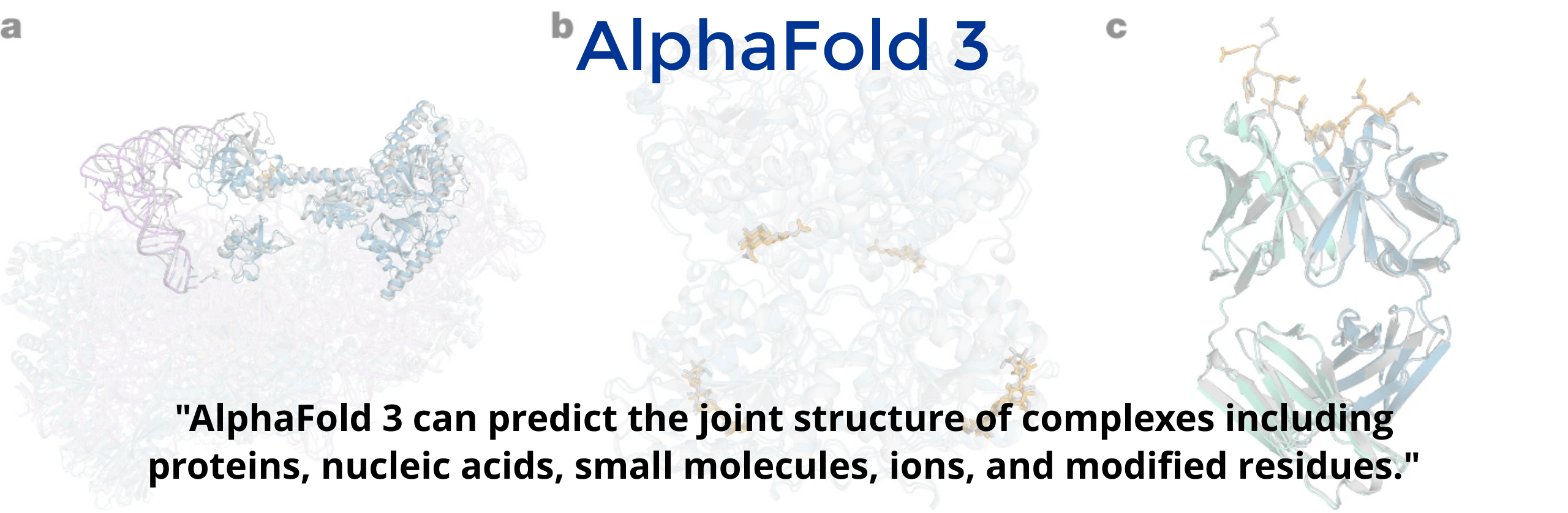
After today, you should be able to



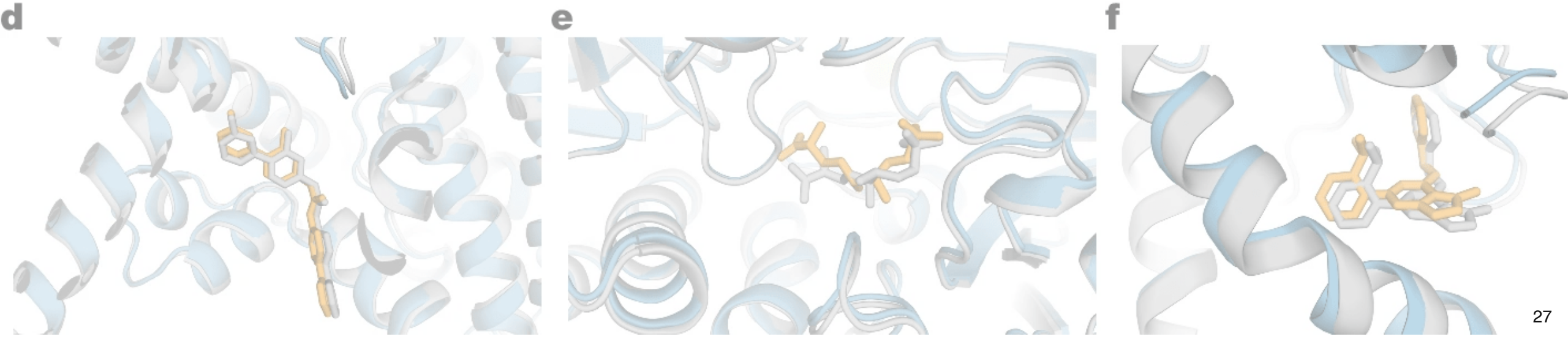
Identify key applications and

recent advancements

AlphaFold 3

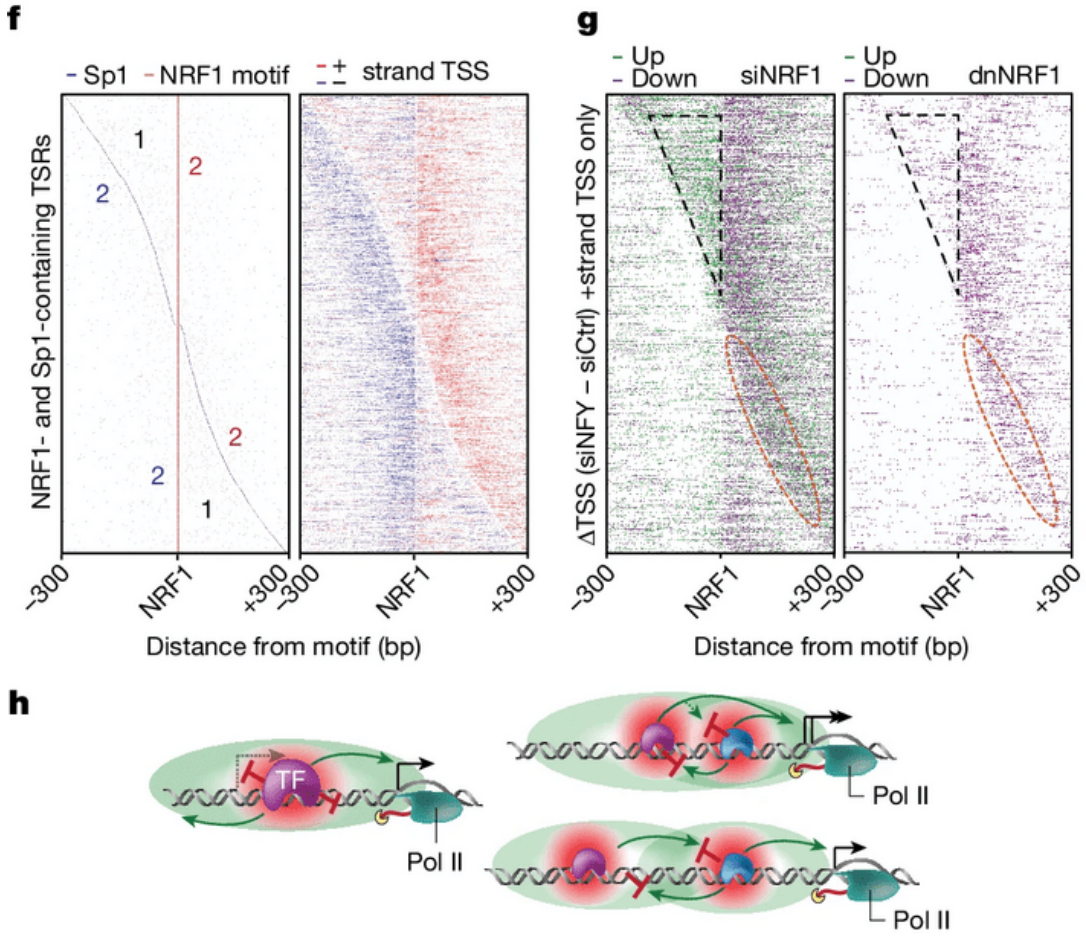


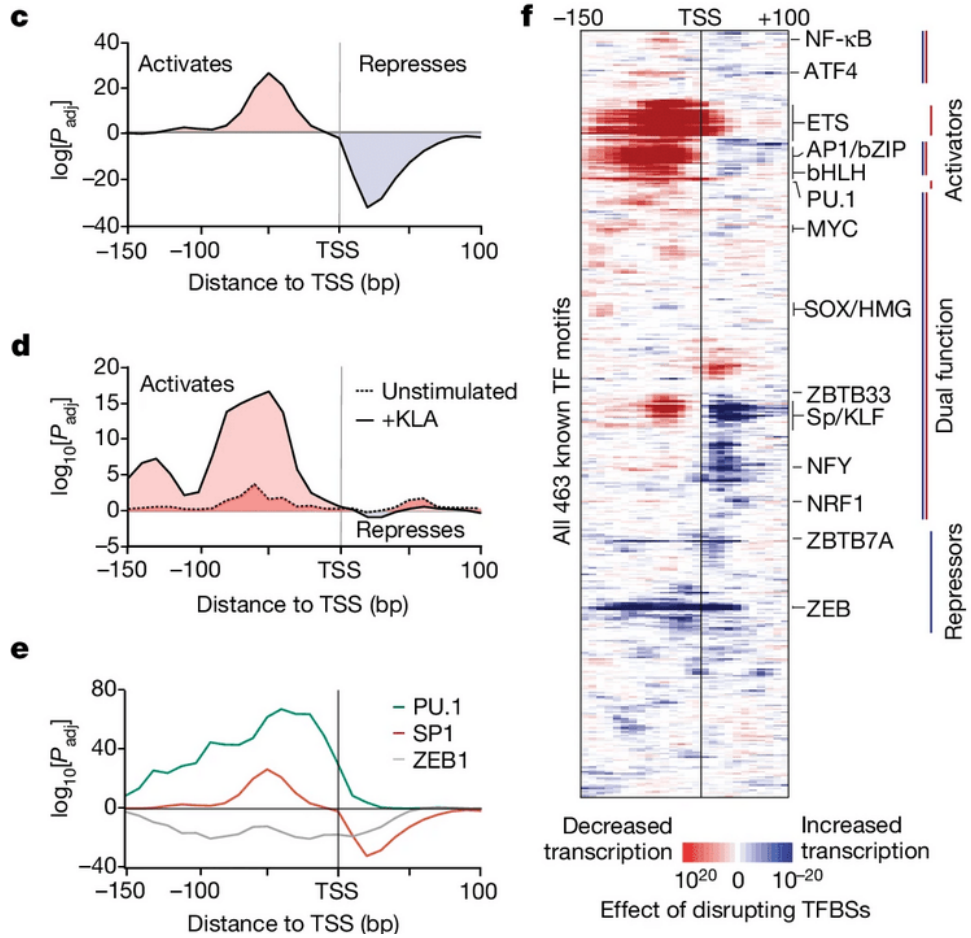
"AlphaFold 3 can predict the joint structure of complexes including proteins, nucleic acids, small molecules, ions, and modified residues."



HOMER2

"We show that the effect of transcription factor binding on transcription initiation is position dependent."





Miniprot: protein-genome aligner

"Miniprot [...] is tens of times faster than existing tools while achieving comparable accuracy on real data."

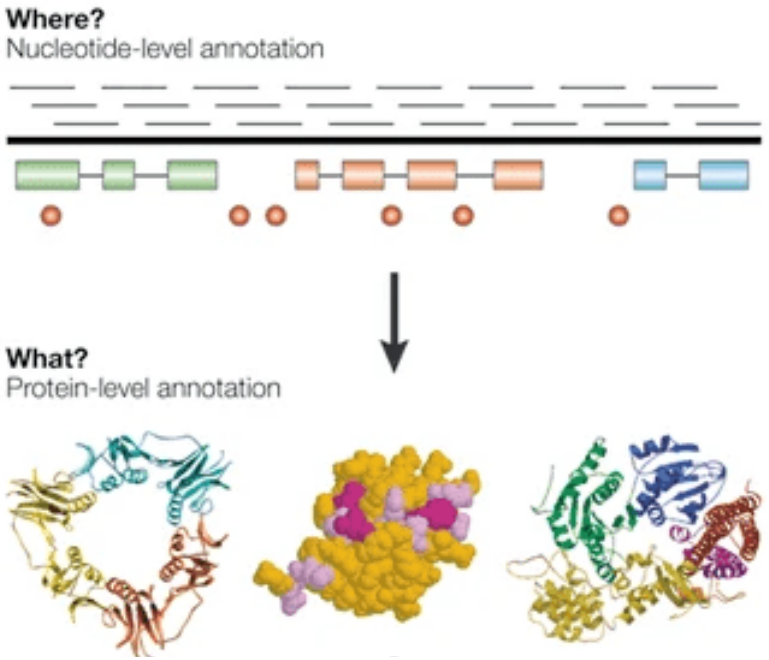


Table 1. Evaluating protein-to-genome alignment

Genome species	Human	Human	Human	Human	Human
Protein species	Zebrafish	Zebrafish	Zebrafish	Zebrafish	Zebrafish
Aligner	Miniprot	Miniprot	Spaln2	Spaln2	GeMoMa
Splice model	Human	General	Human	Default	N/A
Elapsed time (s)	267	257	10 708	11 097	8718
Peak RAM (GB)	21.8	22.5	9.3	8.9	146.9
No. of protein	25 007	25 007	25 007	25 007	25 007
No. of multi-exon	16 866	17 104	13 643	13 854	23 109
No. of predicted junc.	157 918	161 295	151 388	209 312	204 764
No. of non-ovlp. junc.	482	802	1206	15 658	5712
No. of confirmed junc.	145 545	144 734	136 916	129 645	153 781
% confirmed junc.	92.16	89.73	90.44	61.94	75.10
% base SN	63.11	63.16	57.16	55.74	67.02
% base SP	95.43	94.91	95.11	86.75	88.70

Why would we use protein-genome instead of genome-genome mapping?

A. Protein-genome mapping is more sensitive for detecting distant homologs

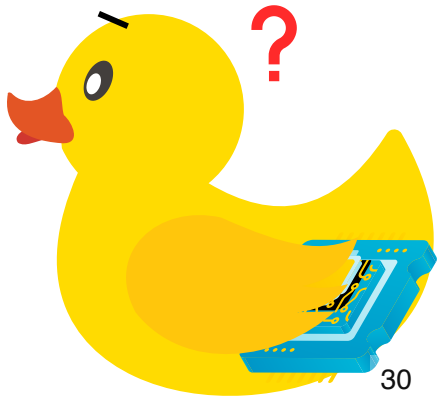
B. Genome-genome mapping is too slow for large-scale comparisons

C. Protein–genome mapping can detect all forms of RNA editing events automatically

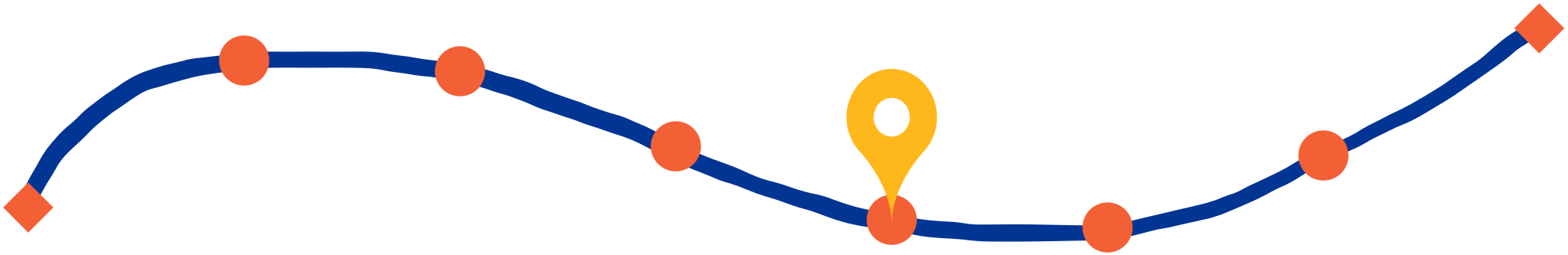
D. Genome-genome mapping cannot handle intron-exon structures

TopHat: 983999

(Not for points)

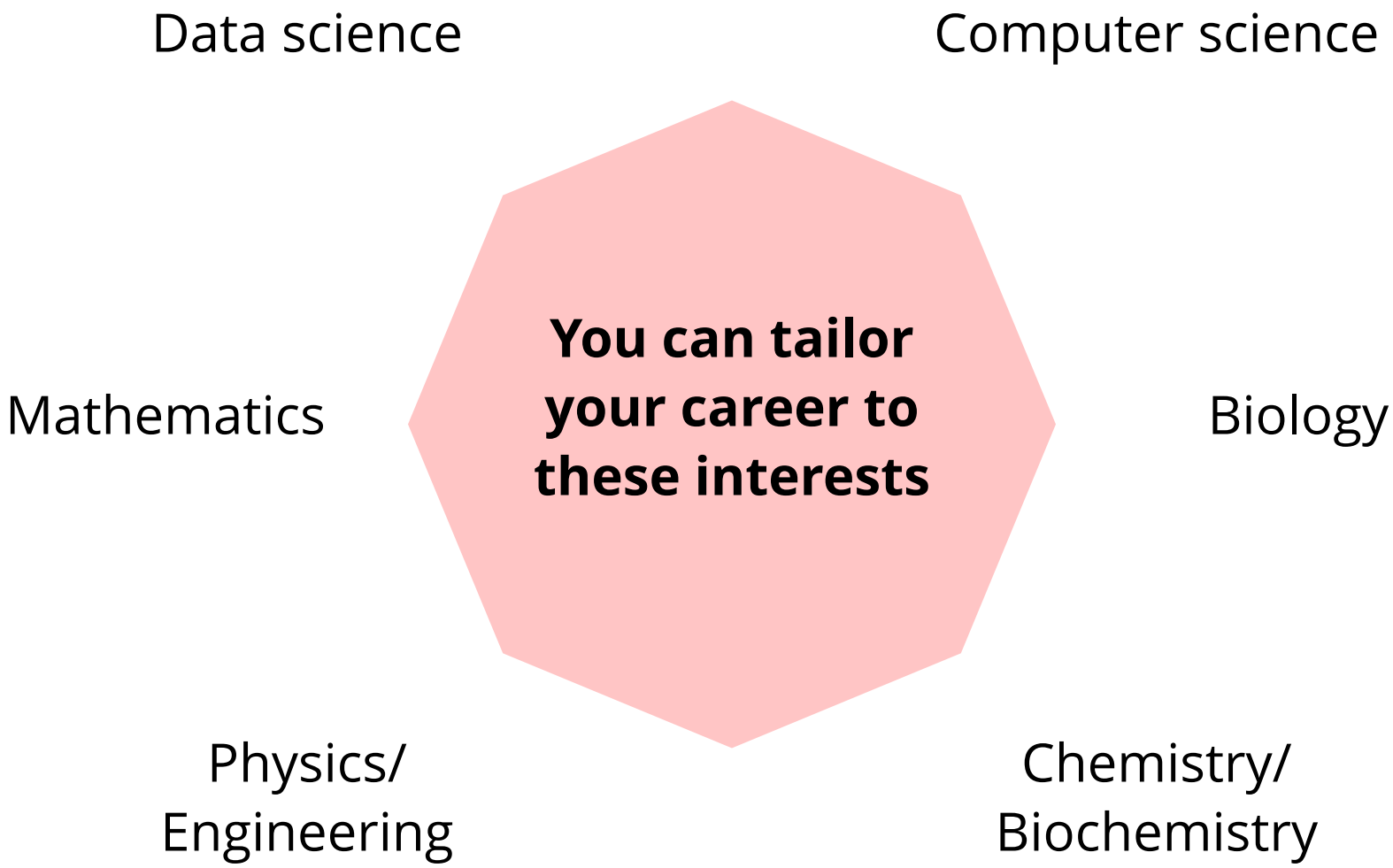


After today, you should be able to



Understand the balance between applications and development

Computational Biology is broad



We will touch on all of these topics in this course

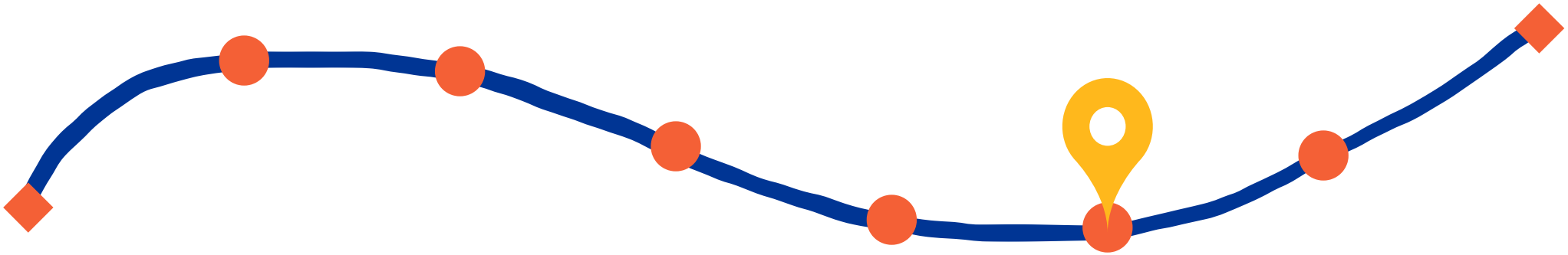
Method development or applying tools?

Typically, it is harder to pick up after the fact (a different way of thinking)

Many, many, many specalities

Both separately are pretty saturated

After today, you should be able to



Identify potential career paths and educational opportunities

Bioinformatics Scientist

Description: Develops software tools and approaches for analyzing biological data, particularly genomic and proteomic data.

Qualifications:

- PhD in Bioinformatics, Computational Biology, or related field
- Strong programming skills (Python, R, C++)

Example companies: UPMC, Illumina, 23andMe, Genentech, Regeneron Pharmaceuticals, Broad Institute

Expected Salary: $80,000 - $130,000

Computational Biologist

Description: Applies computational methods to study biological systems, often focusing on modeling complex biological processes.

Qualifications:

- PhD in Computational Biology, Systems Biology
- Expertise in mathematical modeling and simulation
- Strong programming and data analysis skills

Example companies: Moderna, Vertex Pharmaceuticals, Biogen, Allen Institute for Brain Science, Flatiron Health

Expected Salary: $75,000 - $135,000

Biostatistician

Description: Applies statistical methods to analyze biological and health-related data, often in clinical trials or epidemiological studies.

Qualifications:

- Master's or PhD in Biostatistics or related field
- Strong background in statistics and mathematical modeling
- Proficiency in statistical software (R, SAS, STATA)

Example companies: Pfizer, Merck, Johnson & Johnson, IQVIA, Fred Hutchinson Cancer Research Center

Expected Salary: $72,000 - $119,000

Molecular Modeler

Description: Uses computational methods to model and simulate molecular structures and interactions, often in drug discovery.

Qualifications:

- PhD in Computational Chemistry, Biophysics, or related field
- Experience with molecular dynamics simulations
- Knowledge of drug design principles

Example companies: Schrödinger, Novartis, GlaxoSmithKline (GSK), Atomwise, Dassault Systèmes BIOVIA

Expected Salary: $85,000 - $140,000

If these careers sound interesting, a PhD should be on your radar

Note: There tend to be more jobs in bioinformatics than simulation and modeling

Okay, but what about just a Bachelor's degree?

Challenging for computational biology jobs, but other options are available

Focus on one-half of your major

Biology

I'm unfamiliar with options here (your advisors are well-versed)

Computer Science

Software engineer, data science, machine learning, web development

What will help you prepare for _____?

Everyone applying for the same positions has a college degree

Distinguish yourself with extracirriculars

Employers and graduate schools do not care about the classes you took, they care about **what you can do**

How to do this? **Show and tell**

Your marketable skills are learned outside the classroom

Classes give foundational knowledge to learn hands-on skills in research and internships

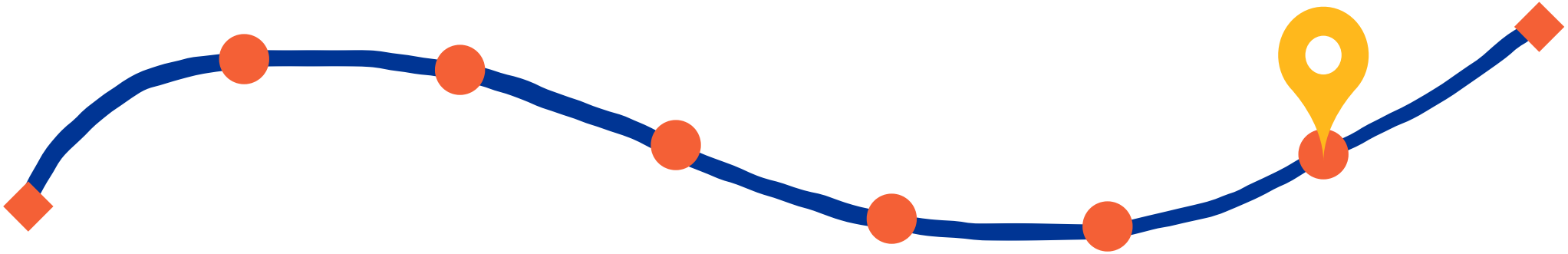
Computational biology: You will get a small taste of this in classes; you need some research or project experience

Computer science: Python, GitHub, machine learning, Rust

Graphic design: Illustrator/Inkscape, Photoshop/Gimp, Blender

Communication: Writing and presenting

After today, you should be able to



Navigate Google Colab and get acquainted with Python

(If time permits)

Python worksheet
(time permitting)

Before the next class, you should

Lecture 01:
Computational biology overview

Lecture 02A:
DNA sequencing - Foundations

Today

Tuesday

- P01A will be released tomorrow and is due Jan 17th
  - You should start Intro to Programming on Kaggle
- Check that you are subscribed to Canvas notifications