3000788 Intro to Comp Molec Biol

Fall 2025

Sira Sriswasdi, PhD

- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

Please introduce yourself

- Name (how should I call you)
- What do you want to learn from this course?
- Undergraduate background
- Graduate program and research interest
- Thesis advisor & topic (if already picked)

3000788 Intro to Comp Molec Biol

Lecture 1: Computational thinking in biology

Fall 2025

Sira Sriswasdi, PhD

- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

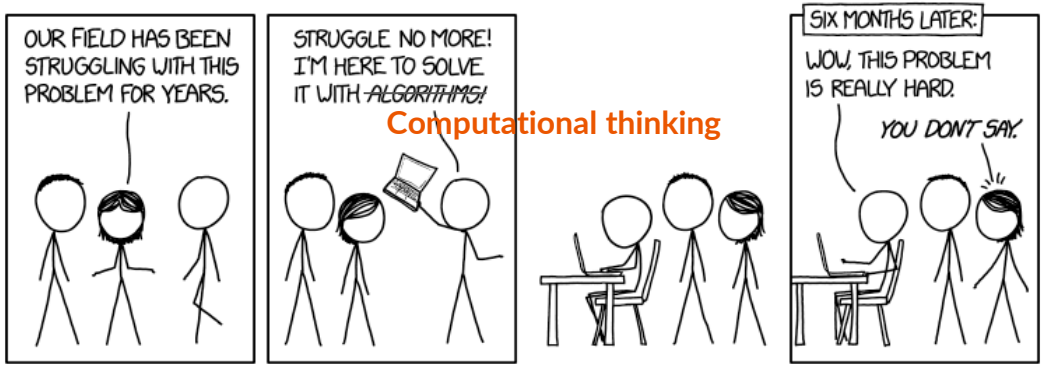
About this course

The Instructor

- **Sira Sriswasdi (สิระ ศรีสวัสดิ์)**
- BSc in Mathematics
- PhD in Computational Biology

- Apply computational thinking and tools to answer biological and biomedical problems

- Machine learning and AI in medicine



Credit: "Here to Help" from xkcd comic, reprinted under Creative Commons License

Course Objectives

- Intended for new students
- Cover broad topics
- **Provide biological motivation (why) but focus on computational aspect (how)**
- Hands-on assignments with instruction

Structure

- 28 sessions
- 6 assignments
  - 15% each
- Can work in group but write your own
- Don’t abuse AI, and credit AI if used
- Can consult me

Week	Session	Date	Day	Time	Description	Activity
1	1	11-Aug-25	Mon		Holiday	
	2	13-Aug-25	Wed	10:00-11:30	Course introduction + history of computation in biology	
2	3	18-Aug-25	Mon	10:00-11:30	Sequence alignment	
	4	20-Aug-25	Wed	10:00-11:30	DNA sequencing techniques and applications	**Assignment 1**
3	5	25-Aug-25	Mon	10:00-11:30	Sequencing data processing workflows	
	6	27-Aug-25	Wed	10:00-11:30	Genome assembly and annotation	
4	7	1-Sep-25	Mon	10:00-11:30	Variant calling and genome-wide association study	
	8	3-Sep-25	Wed	10:00-11:30	Phylogenetics and evolutionary models	**Assignment 2**
5	9	8-Sep-25	Mon	10:00-11:30	Metagenomics	
	10	10-Sep-25	Wed	10:00-11:30	ChIP-seq and DNA motif detection	
6	11	15-Sep-25	Mon	10:00-11:30	Transcriptomics techniques	
	12	17-Sep-25	Wed	10:00-11:30	Differential expression analysis	**Assignment 3**
7	13	22-Sep-25	Mon	10:00-11:30	Functional enrichment analysis	
	14	24-Sep-25	Wed	10:00-11:30	Single-cell and spatial techniques	
8	15	29-Sep-25	Mon	10:00-11:30	Single-cell data processing	
	16	1-Oct-25	Wed	10:00-11:30	Proteomics and mass spectrometry	
9	17	6-Oct-25	Mon	10:00-11:30	Introduction to systems biology and dynamics	**Assignment 4**
	18	8-Oct-25	Wed	10:00-11:30	Multi-omics integration	
10	19	13-Oct-25	Mon		Holiday	
	20	15-Oct-25	Wed	10:00-11:30	Biological networks	
11	21	20-Oct-25	Mon	10:00-11:30	Chromatin conformation capture	**Assignment 5**
	22	22-Oct-25	Wed	10:00-11:30	RNA and protein structure models	
12	23	27-Oct-25	Mon	10:00-11:30	Online tools and resources	
	24	29-Oct-25	Wed	10:00-11:30	Microscopy data analysis	
13	25	3-Nov-25	Mon	10:00-11:30	Essential statistics for computational biologist	**Assignment 6**
	26	5-Nov-25	Wed	10:00-11:30	Omics data visualization	
14	27	10-Nov-25	Mon	10:00-11:30	Introduction to machine learning and AI	
	28	12-Nov-25	Wed	10:00-11:30	Machine learning applications	
15	29	17-Nov-25	Mon	10:00-11:30	Foundational AI models in biology	
	30	19-Nov-25	Wed	10:00-11:30	Biomarker discovery	

Self-study resources

Companion courses from MIT

FOUNDATIONS OF COMPUTATIONAL AND SYSTEMS BIOLOGY

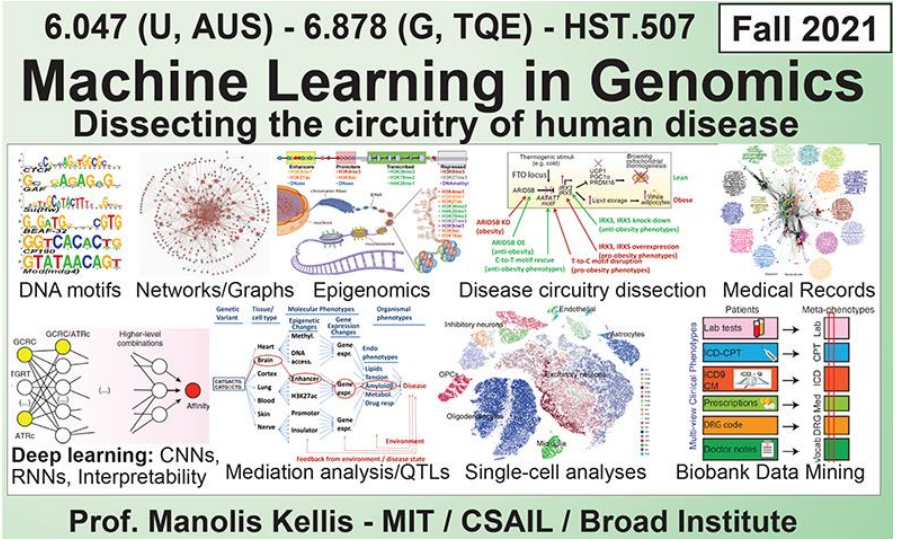
Lecture 1: Introduction to Computational and Systems Biology

Lecture 2: Local Alignment (BLAST) and Statistics

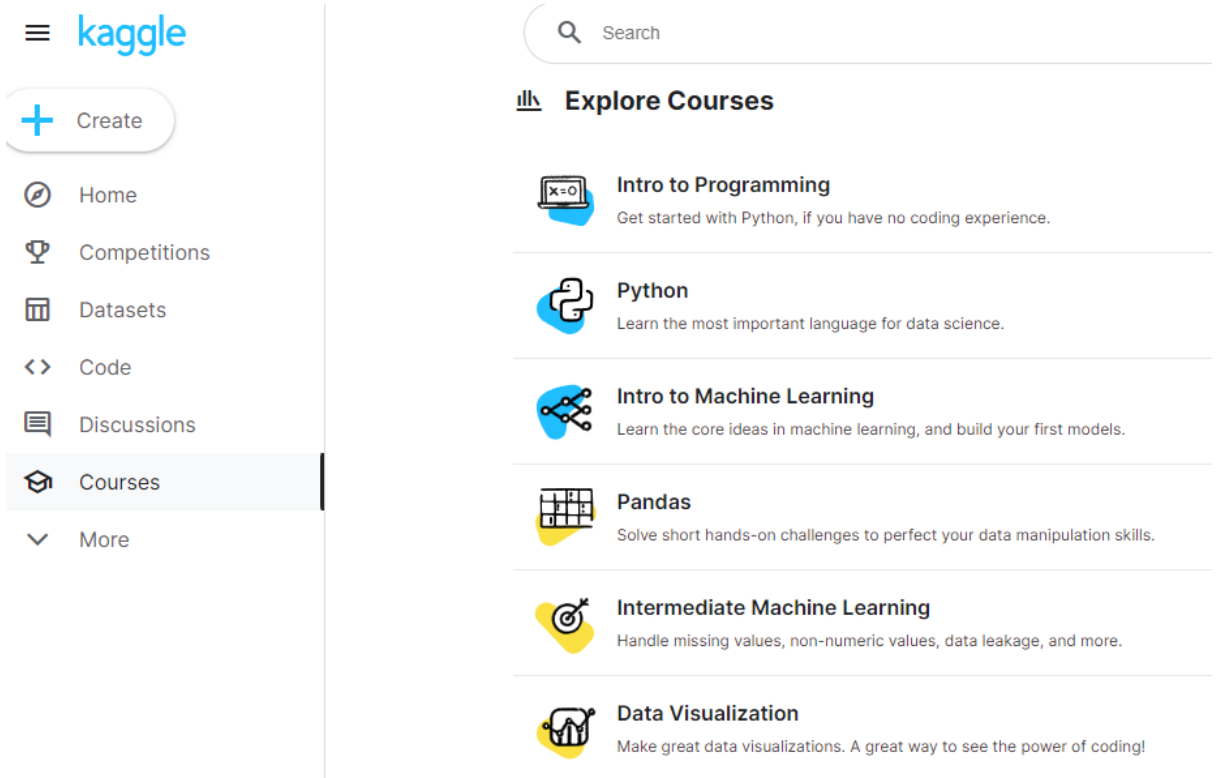
Lecture 3: Global Alignment of Protein Sequences (NW, SW, PAM, BLOSUM)

Lecture 4: Comparative Genomic Analysis of Gene Regulation

Visit https://web.mit.edu/manoli/



Kaggle's programming courses



MIT 6.0001/6.0002

Introduction To Computer Science And Programming In Python

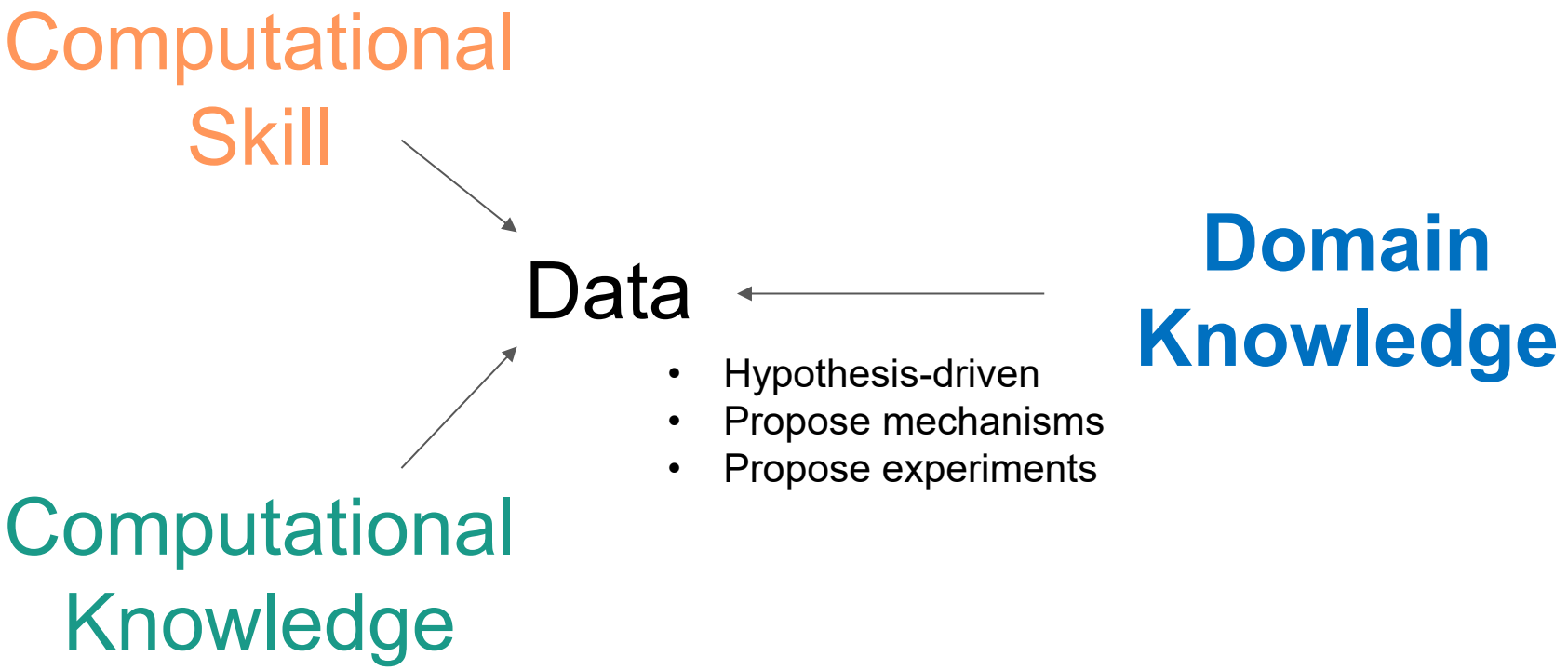
6.0001 Introduction to Computer Science and Programming in Python is intended for students with little or no programming experience. It aims to provide students with an understanding of the role computation can play in solving problems and to help students, regardless of their major, feel justifiably confident of their …Show more

Introduction To Computational Thinking And Data Science

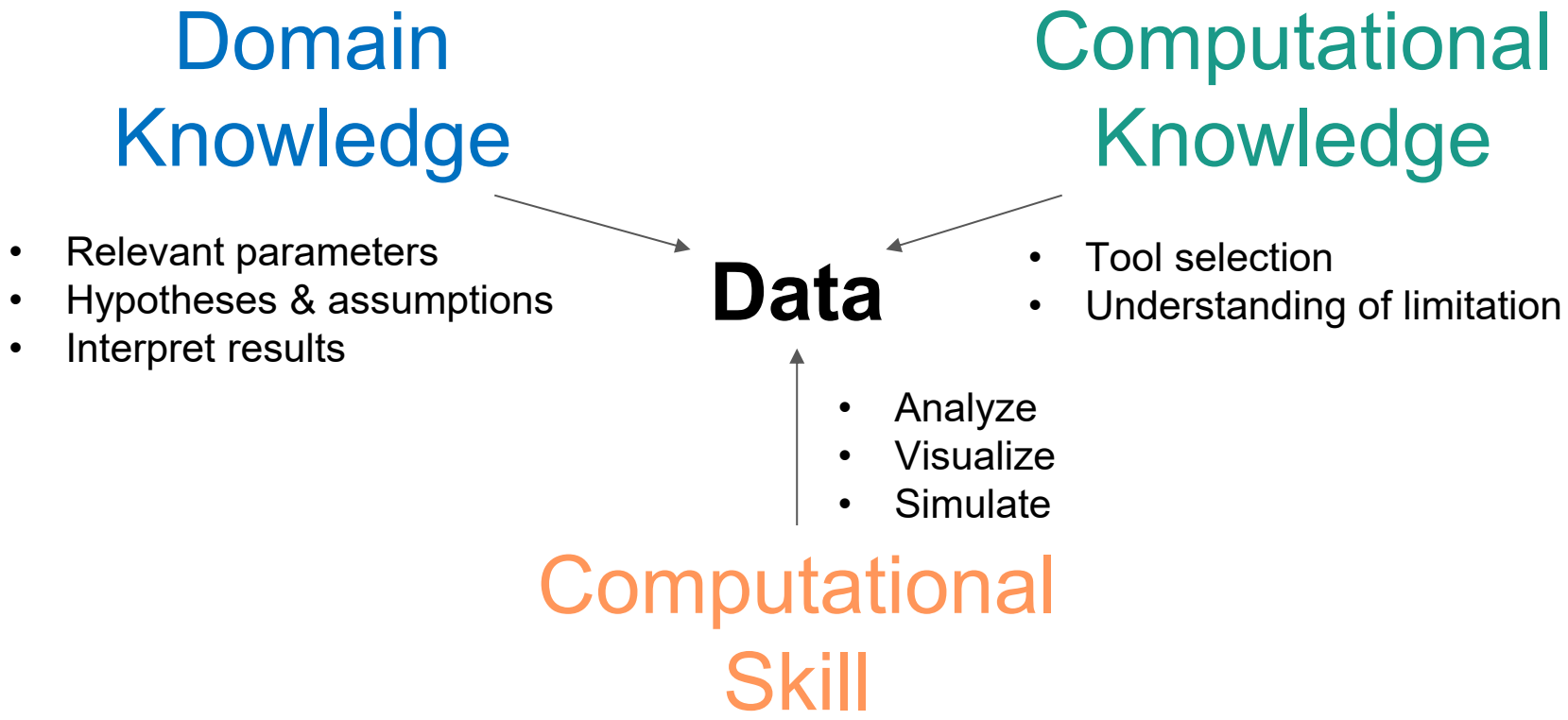
6.0002 is the continuation of *6.0001 Introduction to Computer Science and Programming in Python* and is intended for students with little or no programming experience. It aims to provide students with an understanding of the role computation can play in solving problems and to help students, regardless of their major, …

Computational thinking

A knowledge-centric view



A data-centric view



Many tools for a task

- **Computational knowledge is helpful for assessing tools and techniques**
- Choose the right tools
- Know the limitation and assumption
- Also communicate with peers

List of sequence alignment software

3 languages

Article Talk | Read Edit View history Tools

From Wikipedia, the free encyclopedia

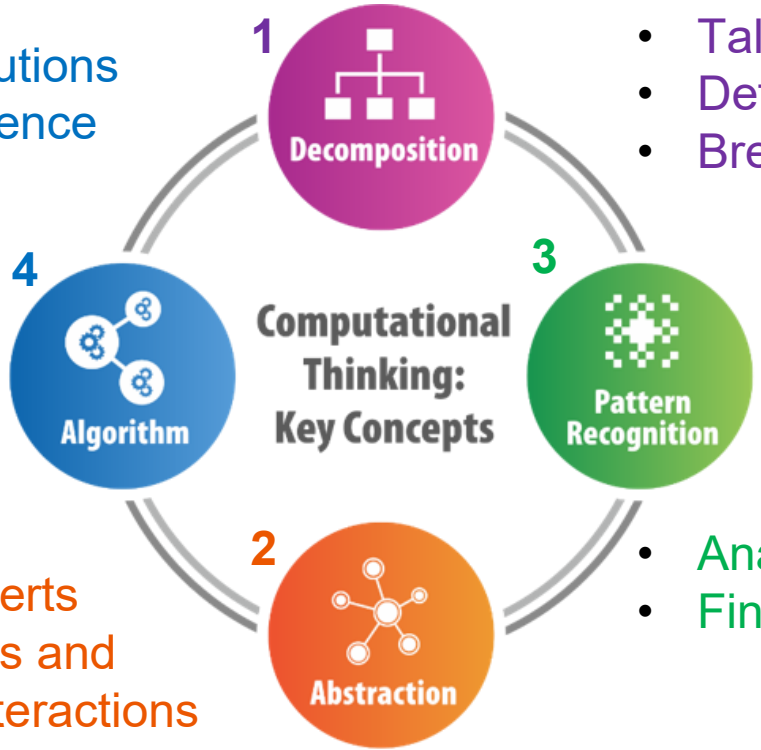
This **list of sequence alignment software** is a compilation of software tools and web portals used in pairwise sequence alignment and multiple sequence alignment. See structural alignment software for structural alignment of proteins.

Database search only [edit]

Name	Description	Sequence type*	Authors	Year
BLAST	Local search with fast k-tuple heuristic (Basic Local Alignment Search Tool)	Both	Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ[1]	1990
HPC-BLAST	NCBI compliant multinode and multicore BLAST wrapper. Distributed with the latest version of BLAST, this wrapper facilitates parallelization of the algorithm on modern hybrid architectures with many nodes and many cores within each node.[2]	Protein	Burdyshaw CE, Sawyer S, Horton MD, Brook RG, Rekapalli B	2017
CS-BLAST	Sequence-context specific BLAST, more sensitive than BLAST, FASTA, and SSEARCH. Position-specific iterative version CSI-BLAST more sensitive than PSI-BLAST	Protein	Angermueller C, Biegert A, Soeding J[3]	2013
CUDASW++	GPU accelerated Smith Waterman algorithm for multiple shared-host GPUs	Protein	Liu Y, Maskell DL and Schmidt B	2009/2010
DIAMOND	BLASTX and BLASTP aligner based on double indexing	Protein	Buchfink B, Xie C, Huson DH, Reuter K, Drost HG [4][5]	2015/2021
FASTA	Local search with fast *k*-tuple heuristic, slower but more sensitive than BLAST	Both		
GGSEARCH, GLSEARCH	Global:Global (GG), Global:Local (GL) alignment with statistics	Protein		
Genome Magician	Software for ultra fast local DNA sequence motif search and pairwise alignment for NGS data (FASTA, FASTQ).	DNA	Hepperle D (www.sequentix.de)	2020
	Genoogle uses indexing and parallel processing techniques for			

Key procedures in computational thinking

- Develop robust solutions
- Design user experience

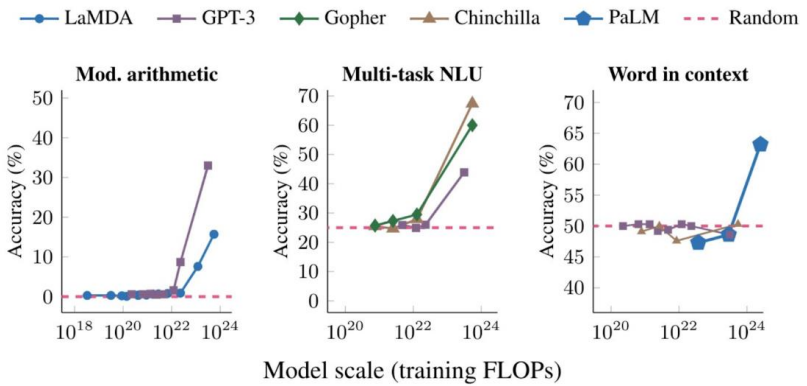


- Talk to stakeholder
- Define problems
- Break into individual tasks

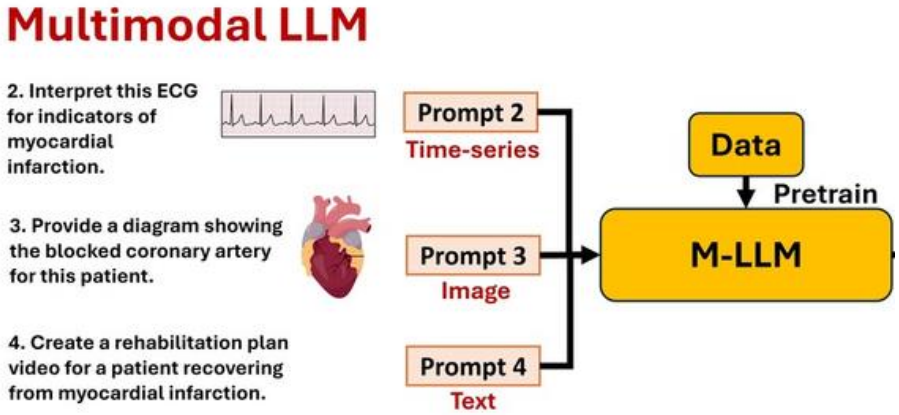
- Talk to domain experts
- Map out parameters and expected causal interactions

- Analyze data and models
- Find recurring patterns

Will large language model solve everything?



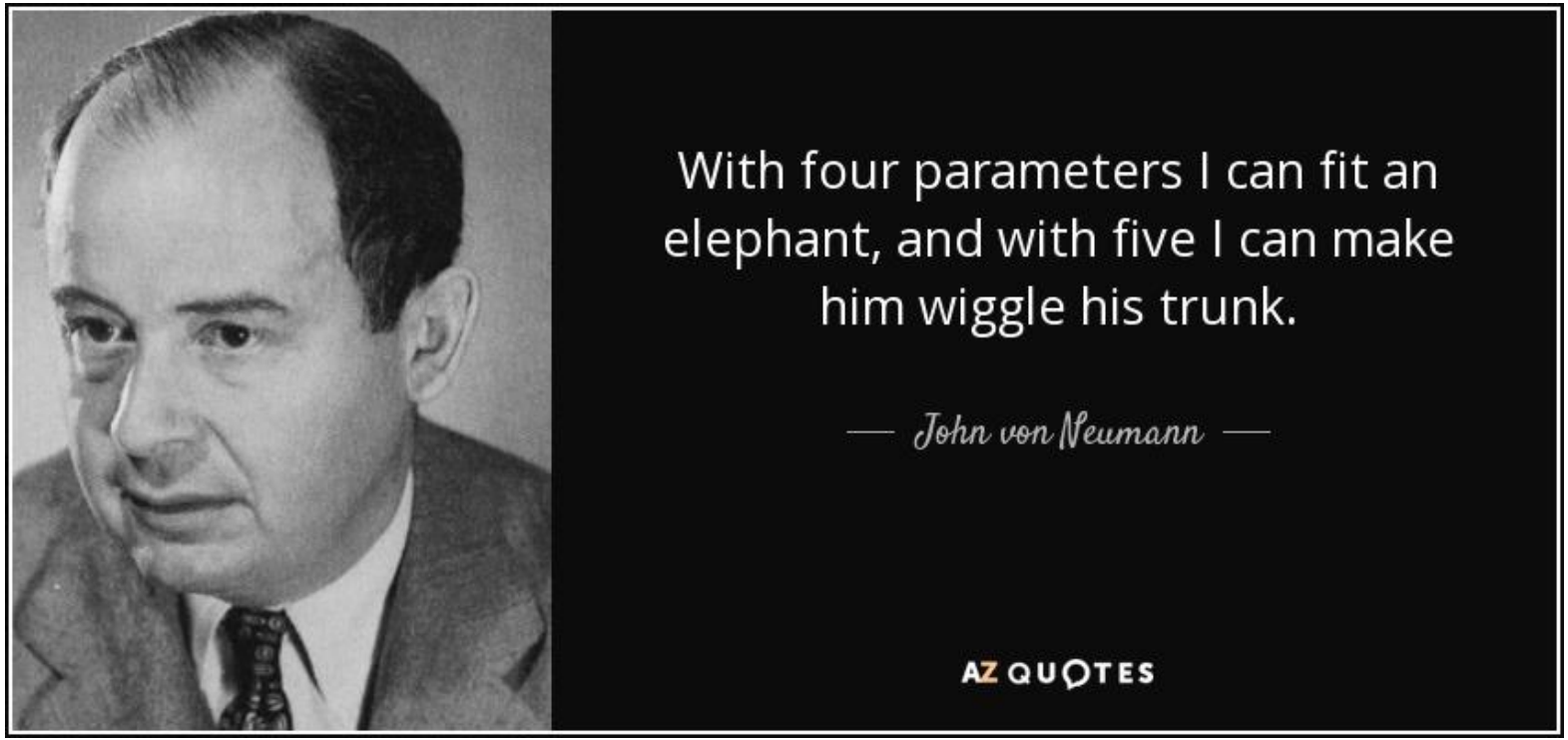
https://hai.stanford.edu/news/examining-emergent-abilities-large-language-models



Alsaad, R. et al. J of Medical Internet Research 2024

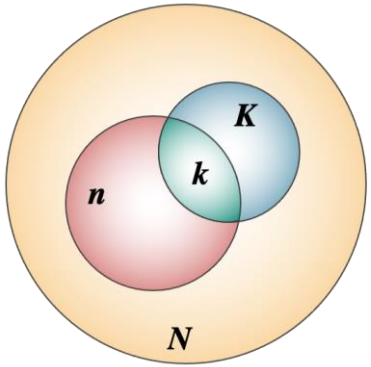
- LLM learns (mostly) correlation, not causation
- Require knowledge to verify
- Require mechanistic explanation to trust

Danger of complexity



Statistical modeling

Statistics are grounded in probability models



$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

N = Background *(e.g. Transcriptome ~ 40.000 genes)*

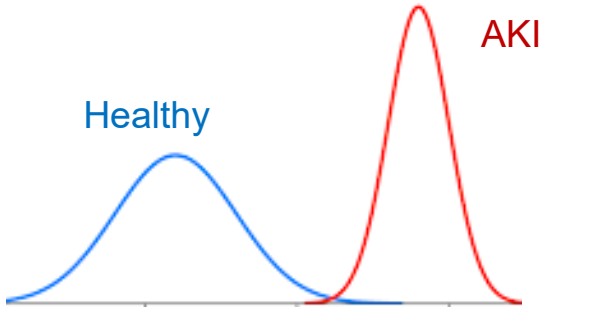
n = Query list *(e.g. upregulated genes)*

K = genes annotated in the pathway/set tested *(e.g. Glycolysis)*

k = ***n*** ∩ ***K***

- In a population *N* = 70 million
  - There are estimated *n* = 1 million with spontaneous nosebleed
  - There are estimated *K* = 20 million exposed to high PM2.5
  - There are estimated *k* = 300,000 with both
- How would you quantify the association between PM2.5 and nosebleed?
  - ***Overrepresentation analysis with hypergeometric distribution***

Unwrapping “canned” statistics



- How would you to quantify the ability of *X* to distinguish healthy and AKI?
  - How about $\text{Score}(X) = \text{Mean}_1 - \text{Mean}_2$ or $\text{Abs}(\text{Mean}_1 - \text{Mean}_2)$?
  - How about $\text{Score}(X) = \frac{\text{Mean}_1 - \text{Mean}_2}{\sqrt{\frac{1}{n}(\text{Variance}_1 + \text{Variance}_2)}}$?
  - You have rediscovered ***independent two-sample t-test*** !!

A peek behind popular statistical tests

- **Mann-Whitney U test:**
  (aka Wilcoxon rank sum)

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j), \quad S(X,Y) = \begin{cases} 1, & \text{if } X > Y, \\ \frac{1}{2}, & \text{if } Y = X, \\ 0, & \text{if } X < Y. \end{cases}$$

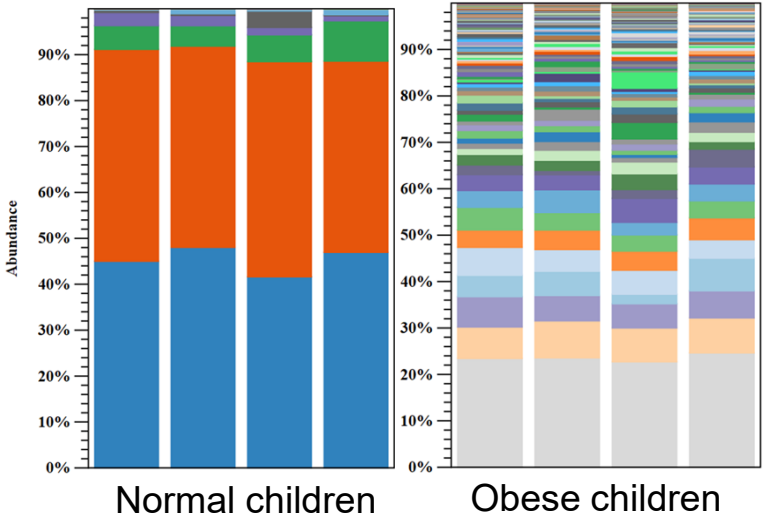
- **Wilcoxon signed-rank test:**

1. Compute $|X_1|, \ldots, |X_n|$.
2. Sort $|X_1|, \ldots, |X_n|$, and use this sorted list to assign ranks $R_1, \ldots, R_n$

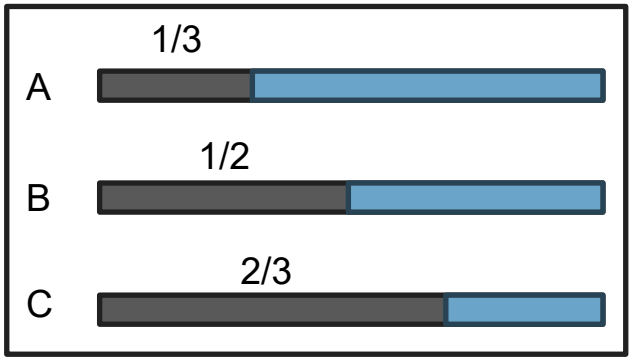
$$T = \sum_{i=1}^{N} \operatorname{sgn}(X_i) R_i.$$

- **Sign test**: Each observation is equally likely to be + or –
  - Pr(*k* positive values from *N* observations) = Binomial(*N*, *k*, *p* = 0.5)

Turning verbal description into mathematical formula

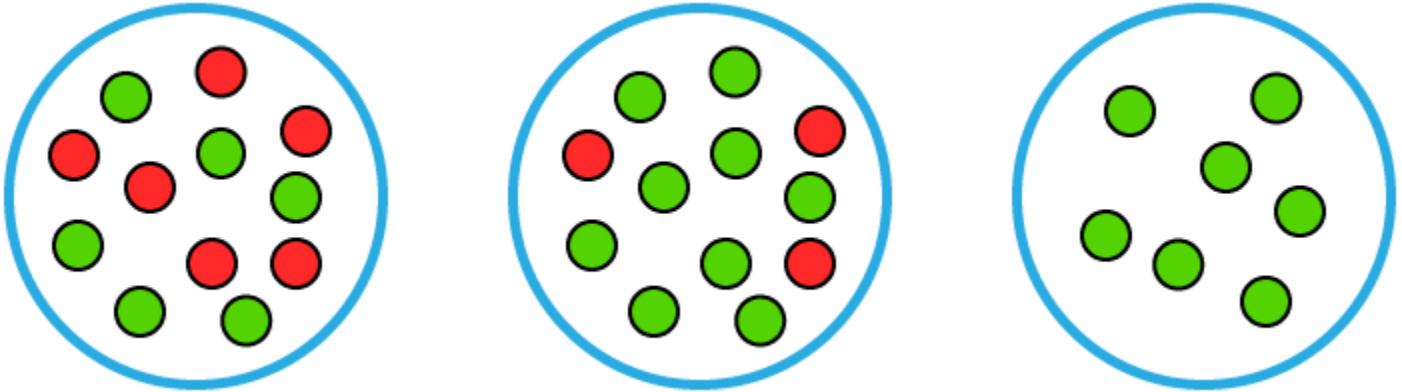


Colors indicate different microbial taxa in children gut microbiome



- How would you quantify the diversity of microbiome?
  - Number of different taxa = n
  - Let p_1, ..., p_n be the taxa frequency, how should diversity change with these parameters?
  - What should be the relationship between **score(A)**, **score(B)**, and **score(C)**?

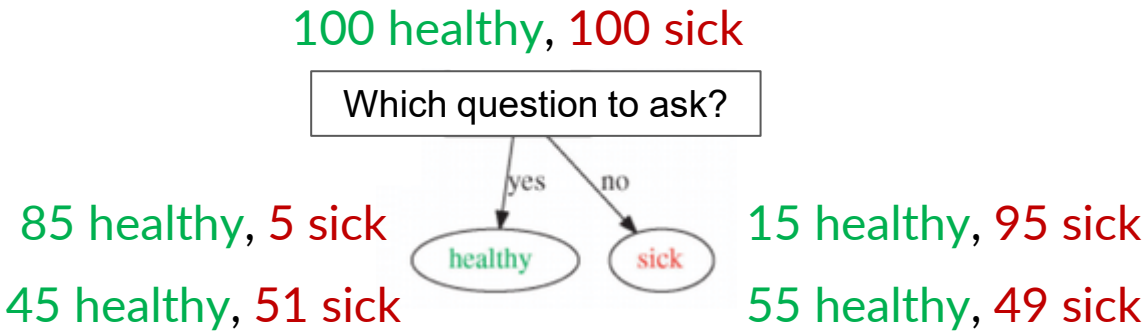
Entropy quantifies purity of a mixture



https://www.javatpoint.com/entropy-in-machine-learning

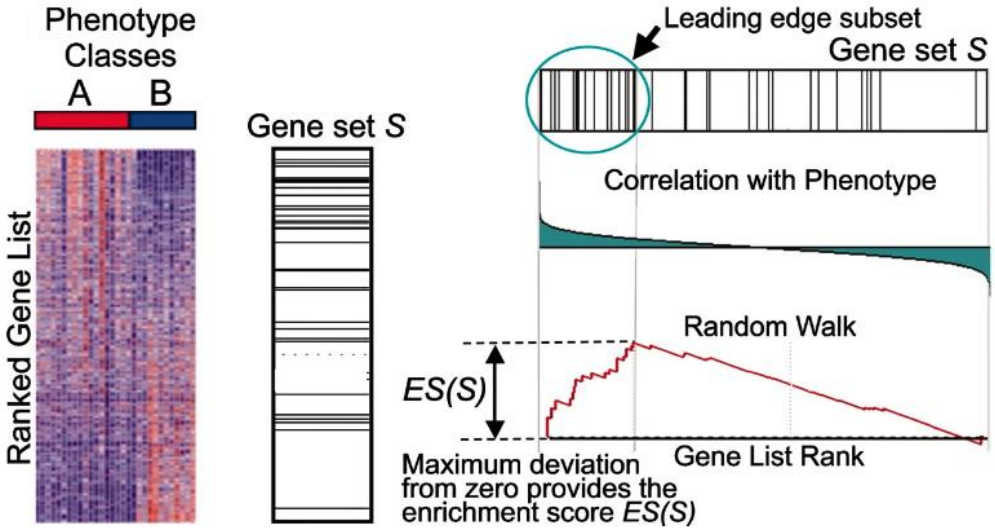
- **Entropy** = $-p_1 \log_2(p_1) - p_2 \log_2(p_2) - \cdots - p_n \log_2(p_n)$
  - Left: $-\frac{6}{12}\log_2\left(\frac{6}{12}\right) - \frac{6}{12}\log_2\left(\frac{6}{12}\right) = \log_2(2) = 1$
  - Middle: $-\frac{3}{12}\log_2\left(\frac{3}{12}\right) - \frac{9}{12}\log_2\left(\frac{9}{12}\right) = 0.811278$
  - Right: $-\frac{0}{12}\log_2\left(\frac{0}{12}\right) - \frac{12}{12}\log_2\left(\frac{12}{12}\right) = 0$

Constructing a decision point



- Identify questions that produce the best separation of patients
  - Gini impurity: $\sum p(1-p)$
  - Entropy: $-\sum p \ln(p)$
- This is the principle behind decision tree model

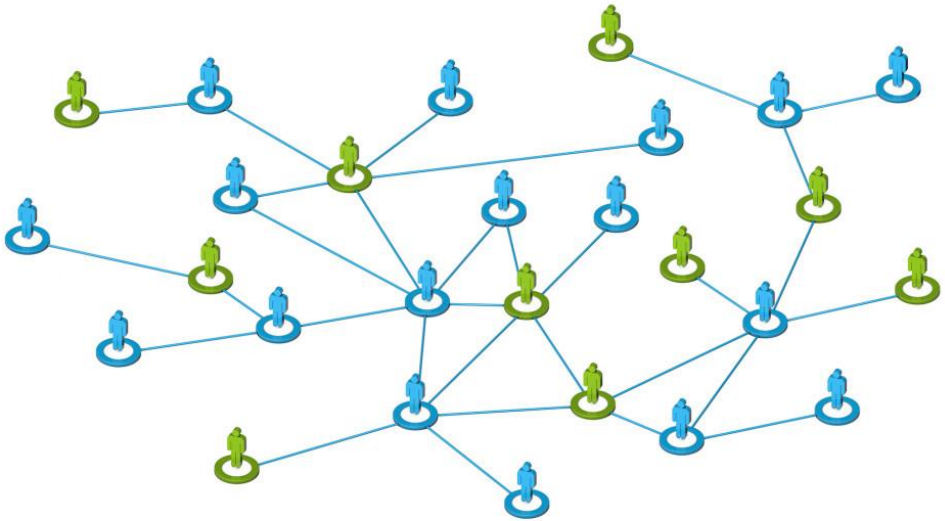
Let's craft a statistical test



Subramanian *et al*. PNAS. 102:15545-15550 (2005)

- Sort 1,000 patients from high to low evaluated risks
- Found that female patients are concentrated at the top
- How will you test whether female patients have higher risks?

Permutation test

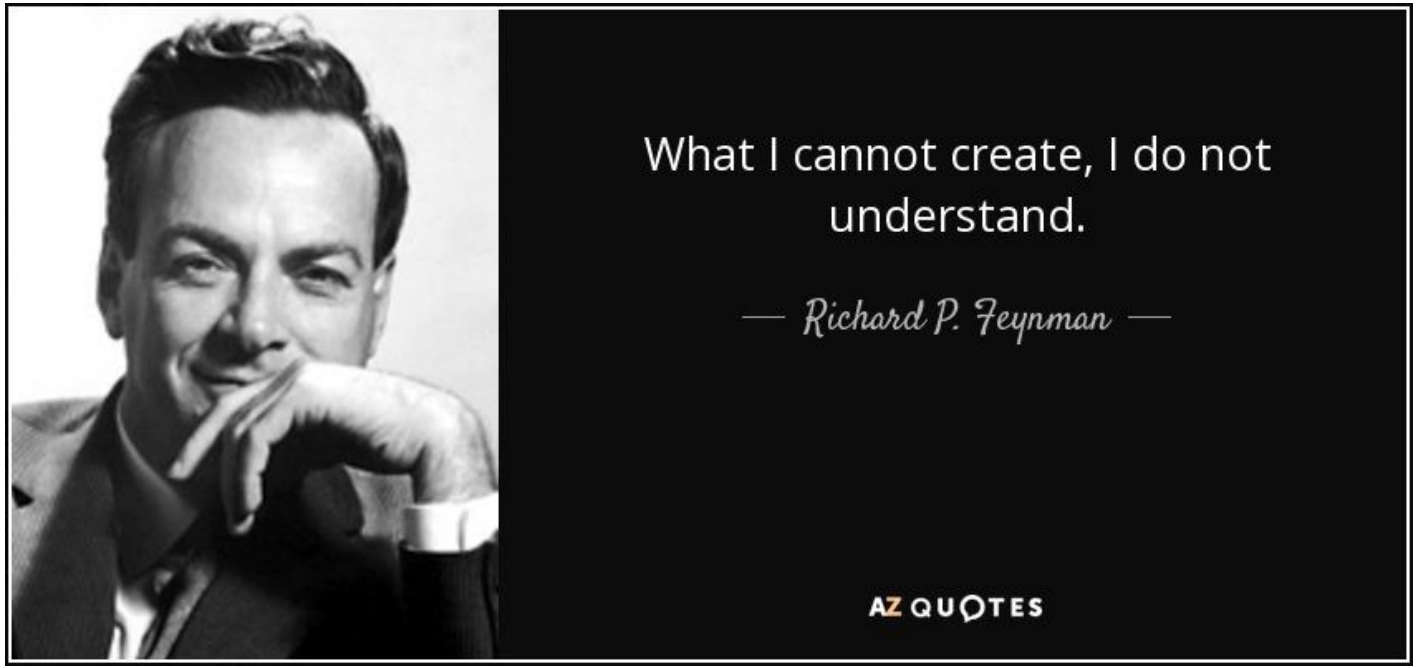


- Generate 1 million networks with the same number of people and friendship connections
- Count the number of connections between friends voting for the same party
- Compare to the original count

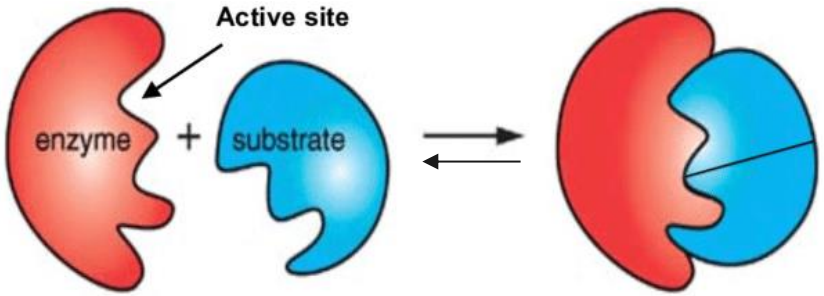
- **Hypothesis**: FB friends tend to vote for the same political party
- How can you test this statement?

Mechanistic modeling

Importance of modeling



Enzyme substrate binding model

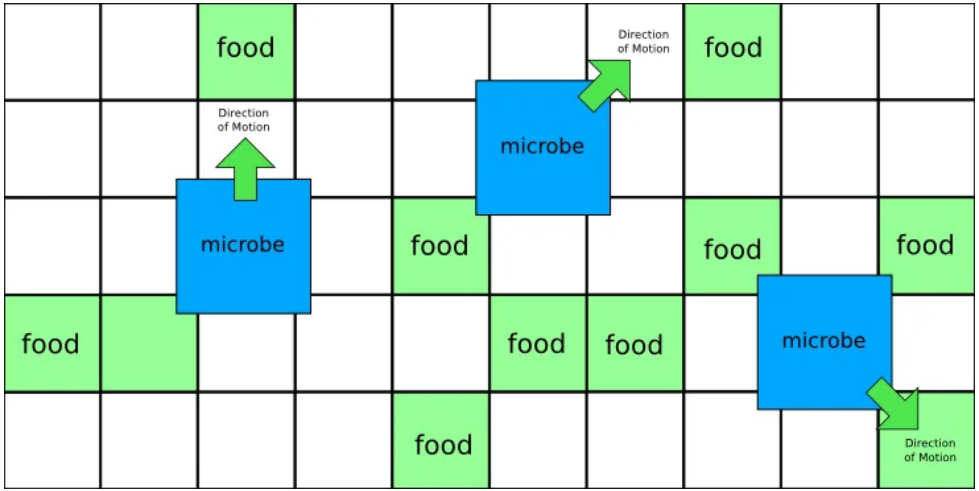


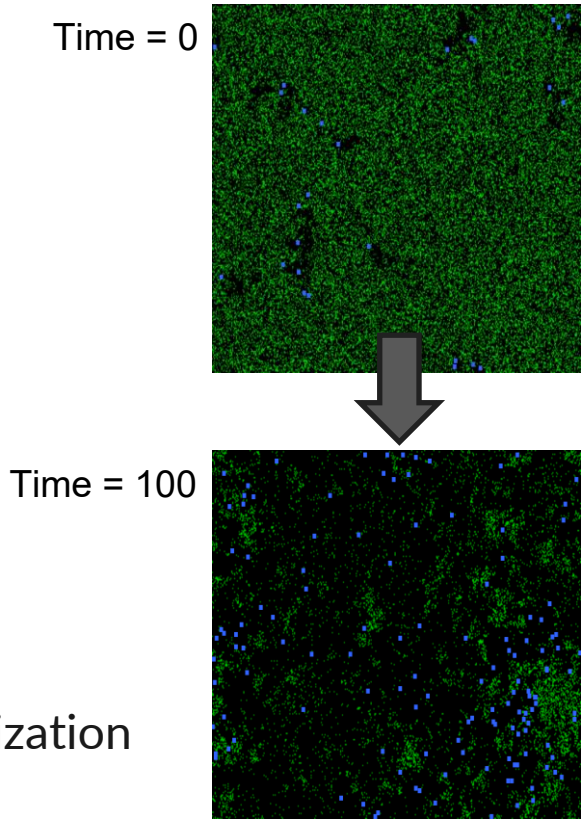
Graham Hutchings. “Development of new highly active nano gold catalysts for selective oxidation reactions” (2014)

- $K_{dissociation}$ = [E][S] / [E-S] at equilibrium, [ ] = concentration

- Association = P(*E* collides with *S* in space) x $Pr_{binding}$ ~ k_1 [E][S]
- Dissociation = Number of E-S molecules x $Pr_{dissociation}$ ~ k_2 [E-S]
- At equilibrium, Association = Dissociation
  - k_1 [E][S] = k_2 [E-S] → $K_{dissociation}$ = k_2 / k_1 = [E][S] / [E-S]

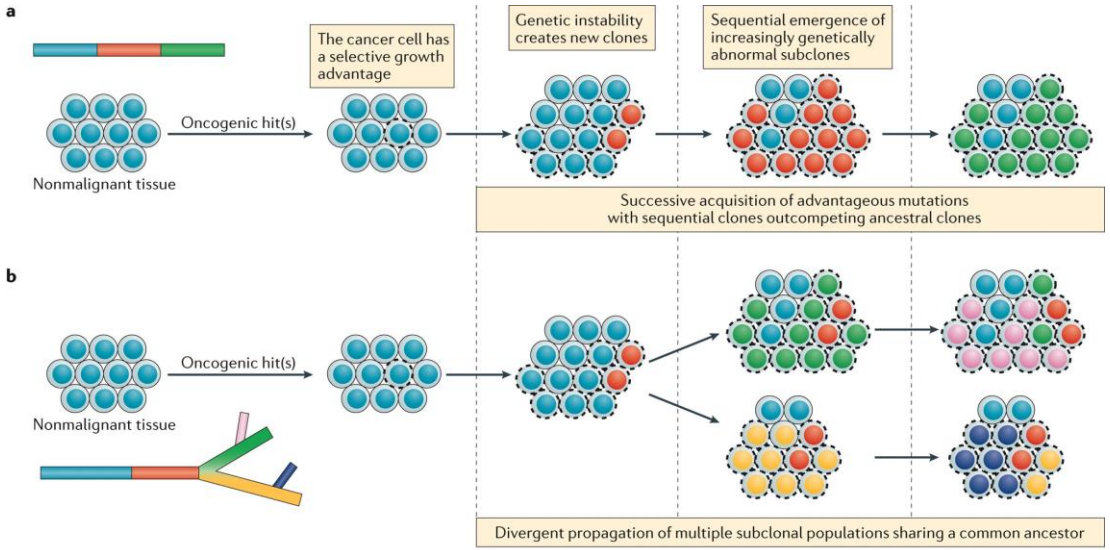
Evolutionary simulation





- Simulate bacteria movement behavior and food utilization
  - Random exploration versus guided chemotaxis
- **https://beltoforion.de/en/simulated_evolution/**

Spatial model for tumor growth

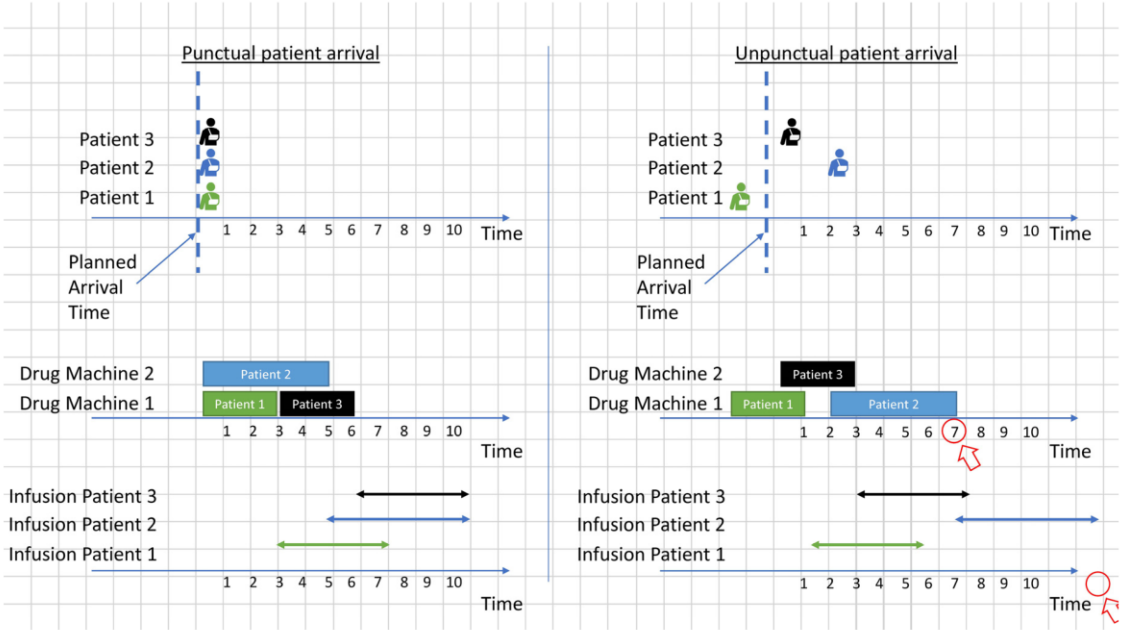


Dagogo-Jack and Shaw, Nat Rev Clin Oncol (2017)

- Possible actions: **Mutate**, **Divide**, **Die**
  - What are the parameters influencing these actions?
  - Competition for resources: tumor volume, relative division rates, ...

Empirical, data-driven modeling

Substituting theory with data

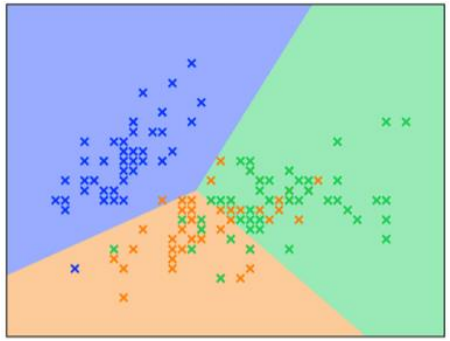


Hadid, M. et al. Int J Environ Res Public Health 2022

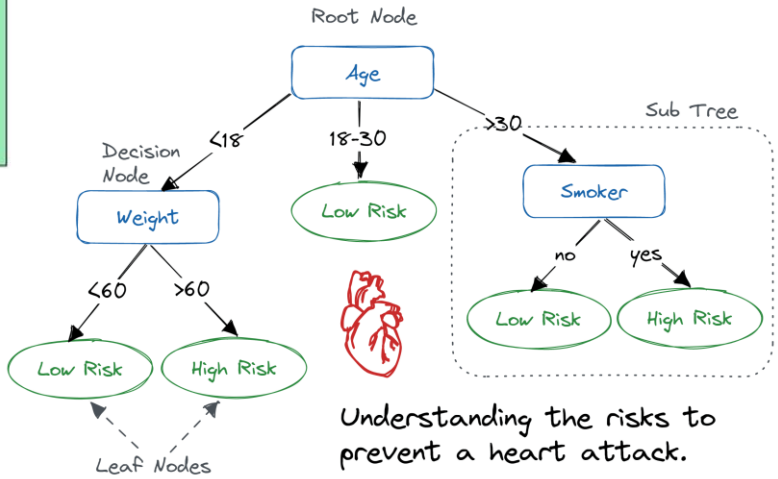
- Some systems cannot be captured by any theoretical distributions
- Build an empirical distribution for sampling using past observations

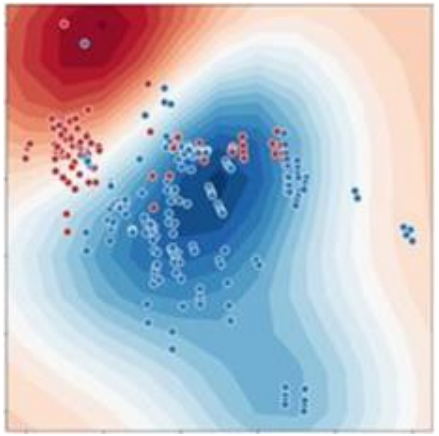
Classical machine learning models

Linear: Score = $(input_1 \times w_1) + \ldots + (input_n \times w_n)$



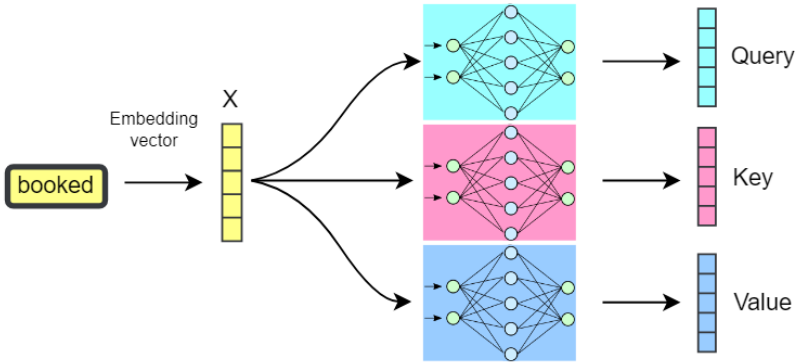
Tree: Collection of decisions



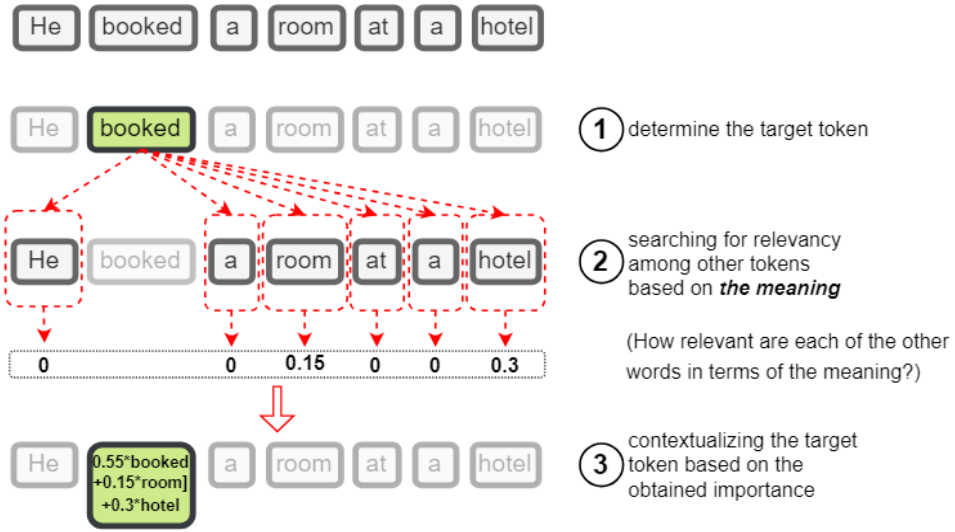


Neighbor-based: Predict using similarity to past observations

Attention mechanism in transformer architecture

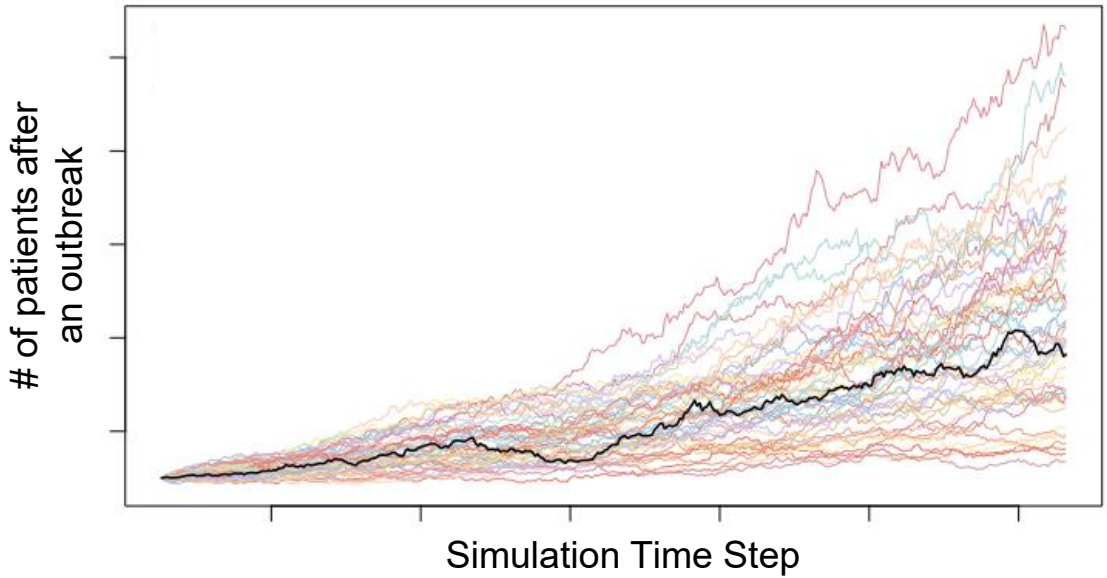


https://epichka.com/blog/2023/qkv-transformer/



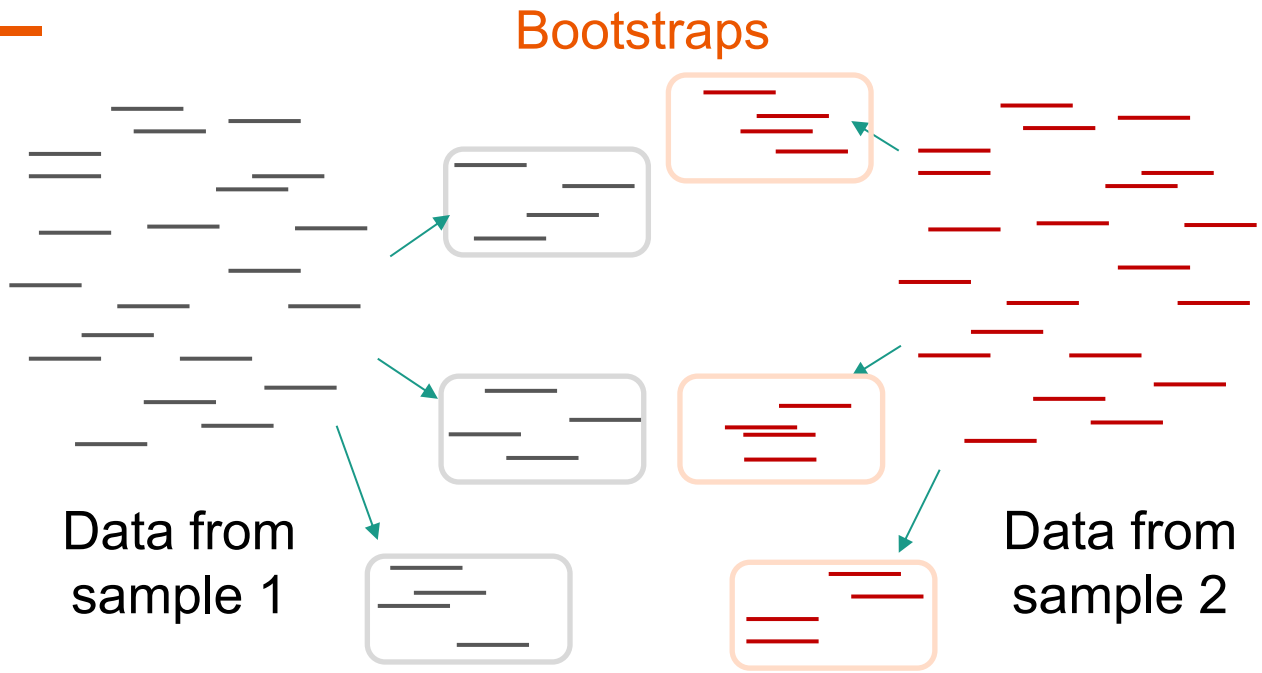
- Generate Query, Key, and Value for each input token
- Use Query-Key similarity to derive attention weights
- Apply weights to the corresponding Value to get the output

Accounting for uncertainty: Monte Carlo simulation



- Different initial conditions can lead to different outcomes
- At each step, multiple outcomes have a probability to occur
- Simulate a population and analyze

Variance estimate: Bootstrapping



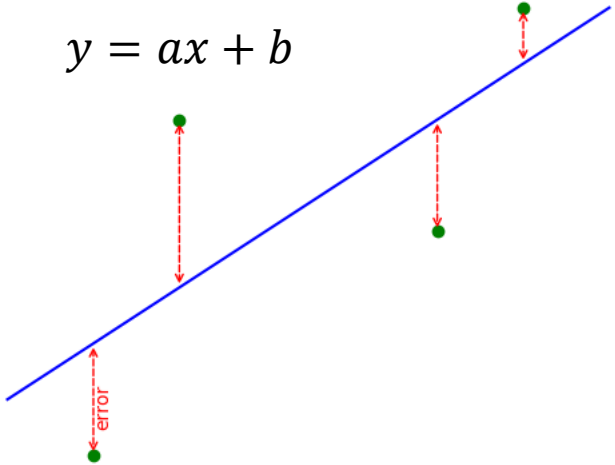
- Instead of using the whole data to perform one calculation
- Sampling from data to perform multiple calculations → Estimate variance

Optimization perspective

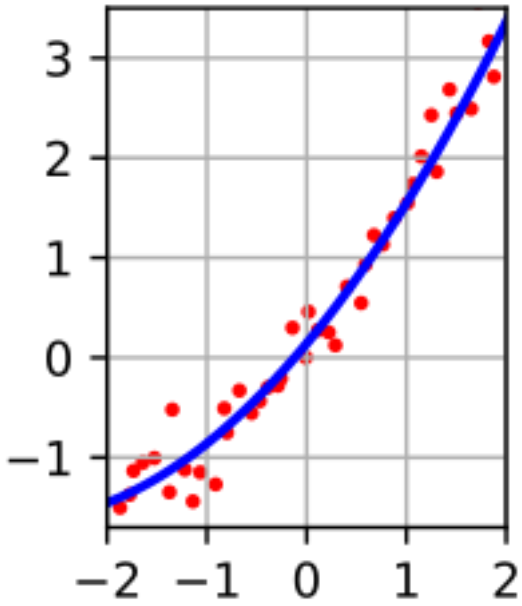
Every algorithm involves optimization

- **Curve fitting** = minimize square error between curve and data points
- **Drug-protein docking** = minimize energy of the drug-protein complex
- **Hospital queuing** = minimize average patient wait time
- **Cell type annotation** = maximize similarity to known cell types
- **Sequence alignment** = minimize number of mismatched position

Curve fitting with least square



https://en.wikipedia.org/wiki/Least_squares

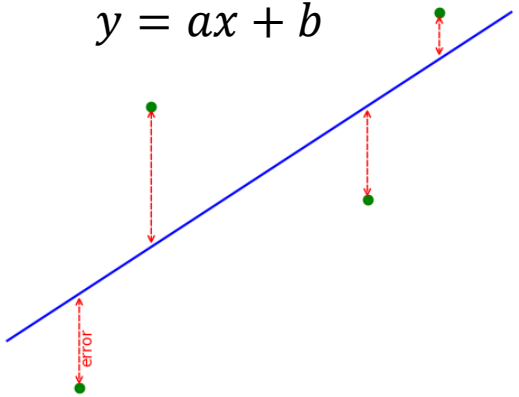


$y = ax^2 + bx + c$

- Finding the **best** *a*, *b*, and *c* that make the curve fit the observations
- Minimize least square error $\frac{1}{n}\left((\hat{y}_1 - y_1)^2 + \cdots + (\hat{y}_n - y_n)^2\right)$

Some optimization approaches

- Minimize $\mathrm{L}(a,b) = \frac{1}{n}\left((\hat{y}_1 - y_1)^2 + \cdots + (\hat{y}_n - y_n)^2\right)$
- Randomly search for (a, b)
  - Heuristic guess: $a = \frac{\sum y_i}{\sum x_i}$
- Use calculus
  - At optimal, slopes are zero
  - Solve $\frac{\partial \mathrm{L}(a,b)}{\partial a} = 0$ and $\frac{\partial \mathrm{L}(a,b)}{\partial b} = 0$



Newton-Ralphson method

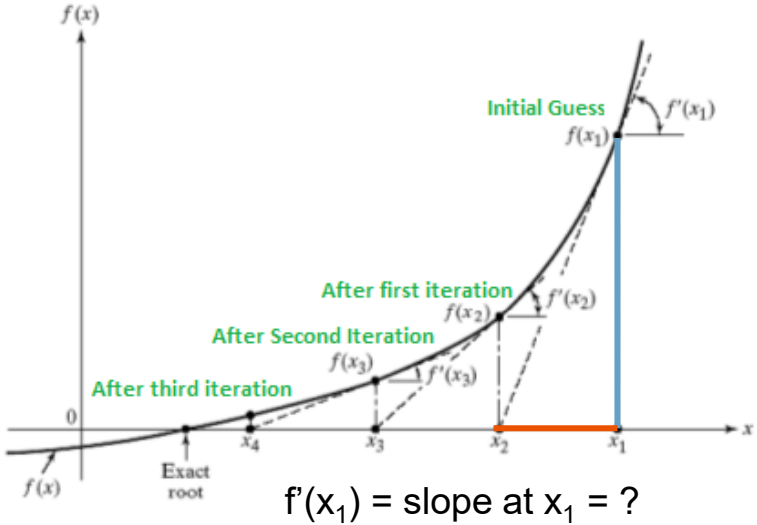
- Want to maximize L(x) by solving

$$\mathrm{f(x)} = \frac{d\mathrm{L}(x)}{dx} = 0$$

- Start with an initial guess x_1
- Calculate $\mathrm{f}(x)$ and $\mathrm{f}'(\mathrm{x})$ at x_1
- Define $x_2 = x_1 - \frac{f(x_1)}{f\prime(x_1)}$

- Repeat the process for x_3, x_4, …
- Stop when x_i converges to a value

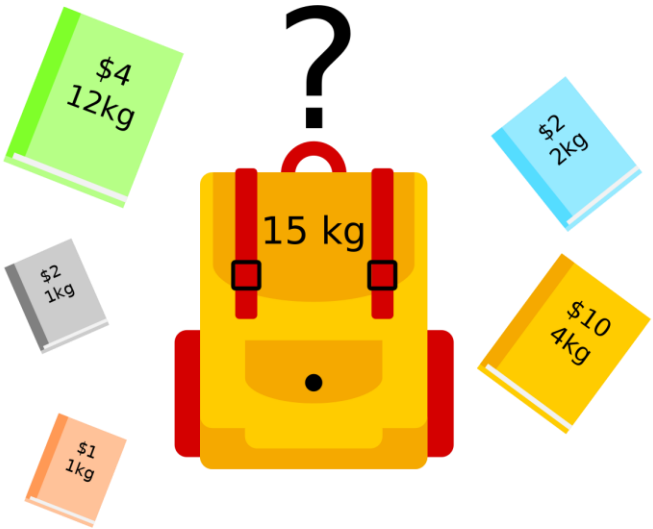
f(x) = the first derivative of the objective L(x)



f'(x_1) = slope at x_1 = ?

https://predictivehacks.com/newton-raphson-method-in-python/

Combinatorial decision: Knapsack problem

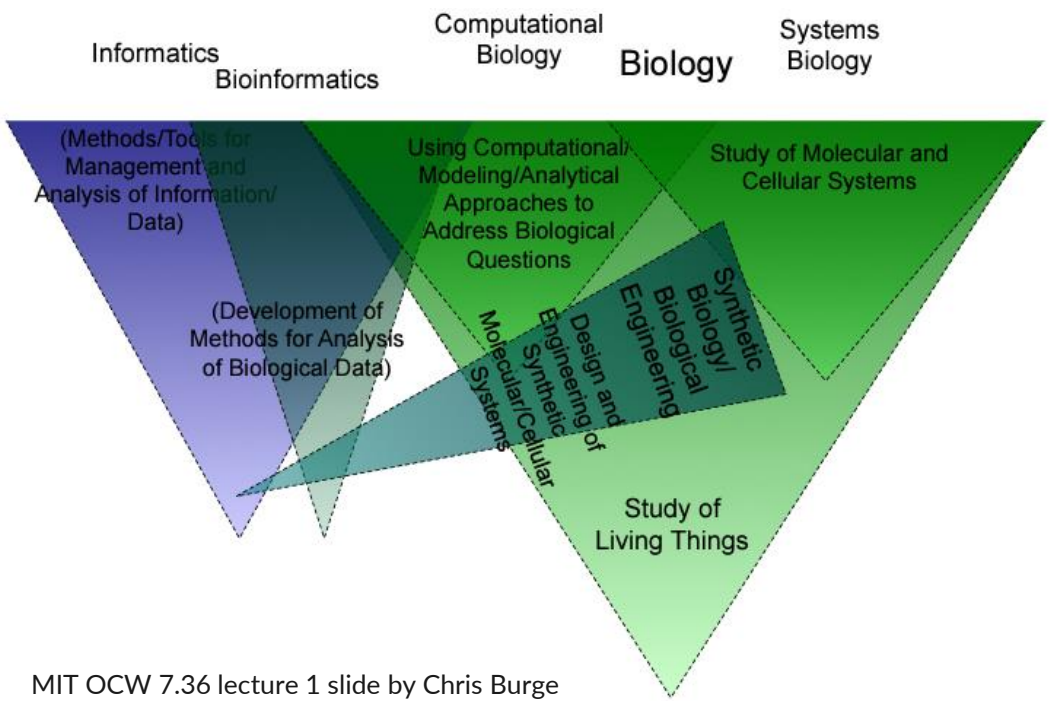


https://en.wikipedia.org/wiki/Knapsack_problem

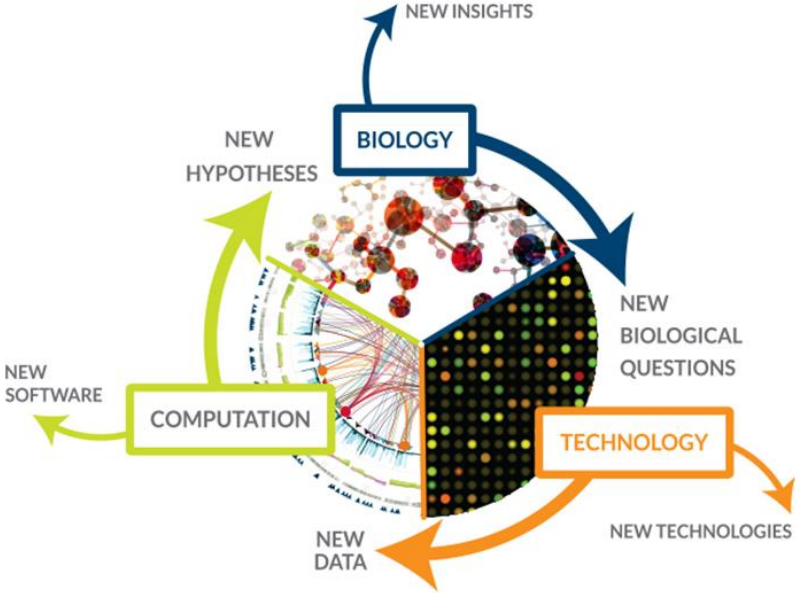
- Limited bag capacity
- Want to pick as much value as possible without exceeding the bag capacity
- General optimization setting
  - **Objective** = Total value
  - **Constraint** = Total weight ≤ W
- Evolution is a combinatorial optimization of the genome

Computational thinking in biology

What is computational biology?



MIT OCW 7.36 lecture 1 slide by Chris Burge

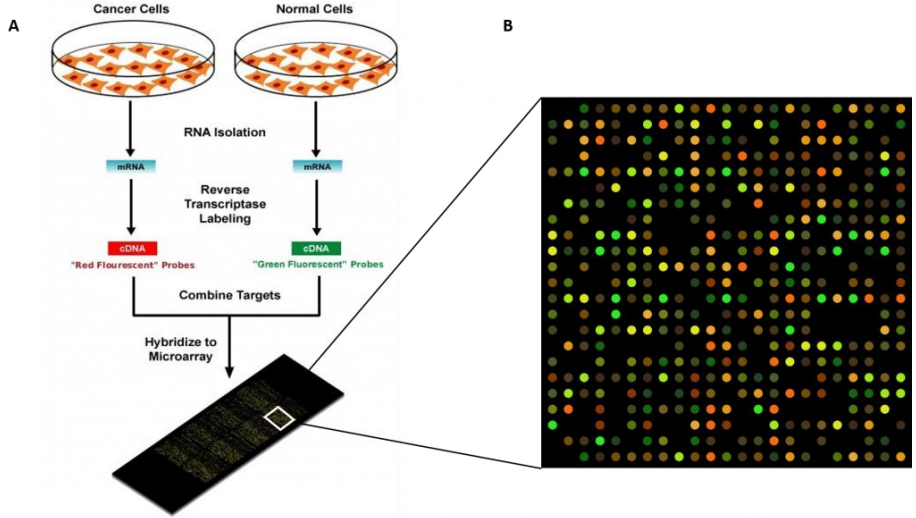


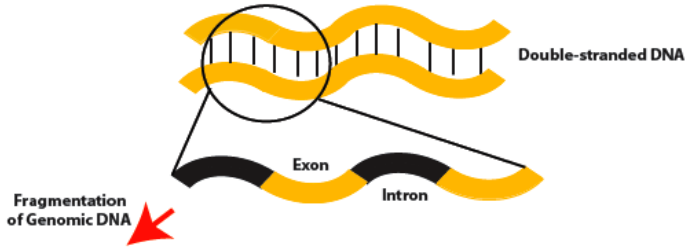
www.systemsbiology.org/about/what-is-systems-biology

The Big Bang moment of biology

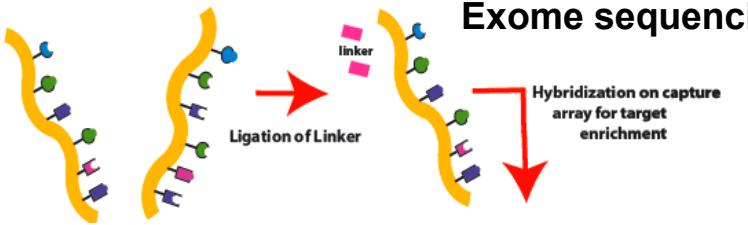
BLAST

Oligonucleotide microarray



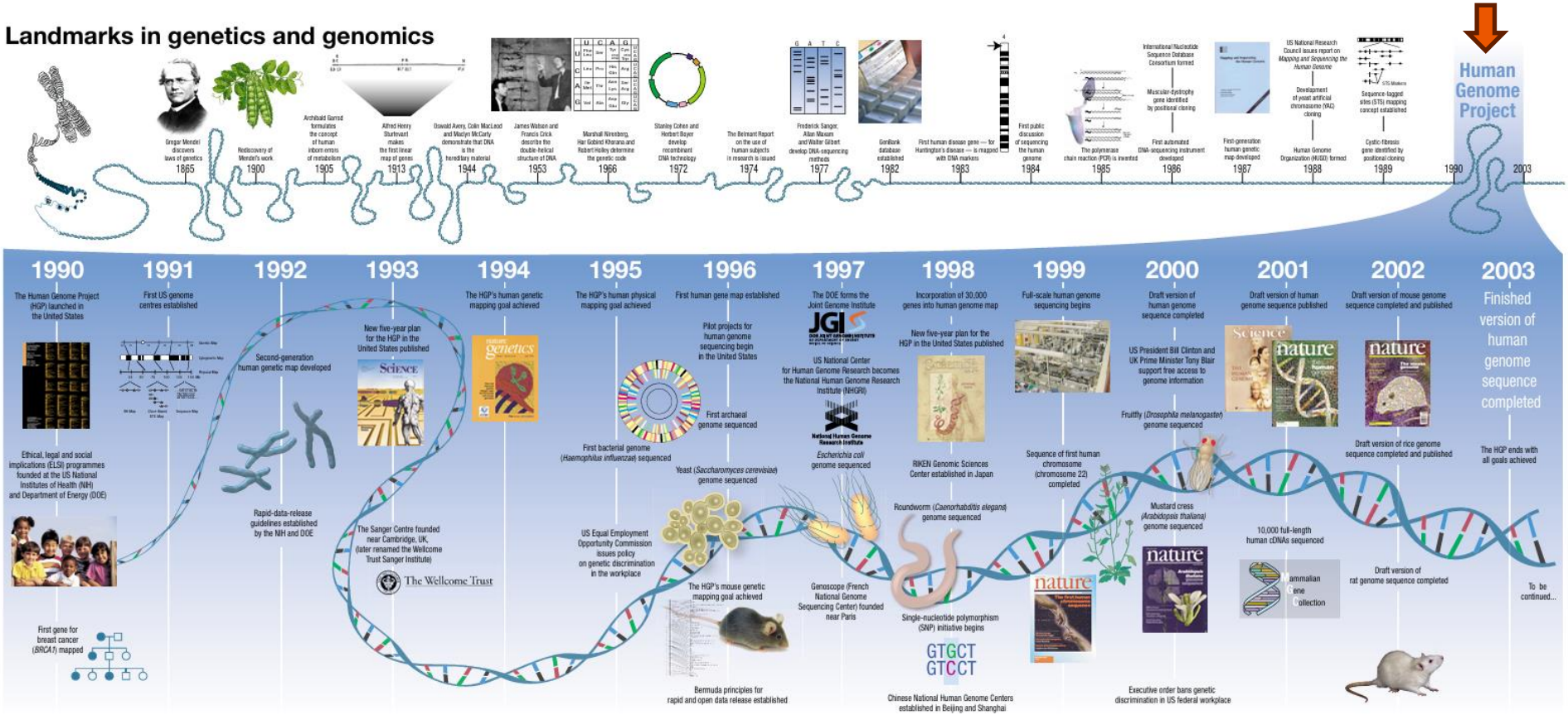


Exome sequencing



What does these techniques have in common?

Landmarks in genetics and genomics



The Evolution of Mass Spectrometry

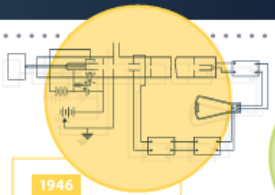
A journey through time and technology

References

1. Wikimedia Commons, Jeff Dahl
2. Courtesy of Pacific Northwest National Laboratory
3. Wikimedia Commons, JEOL USA

1898
Wilhelm Wien discovers that beams of charged particles are deflected by a magnetic field.

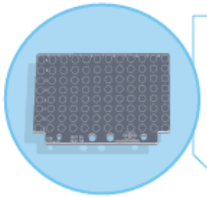
1919
Francis Aston produces a velocity focusing mass spectrograph.



1946
William Stephens reveals the concept of a time-of-flight mass analyzer.

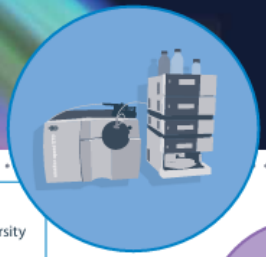
1969
The American Society for Mass Spectrometry (ASMS) is formed.

1984
John Bennett Fenn and his team ionize biomolecules using electrospray.



1987
Koichi Tanaka ionizes intact proteins using a method later known as soft laser desorption (SLD).

2004
A research group at the University of Virginia, presents electron transfer dissociation (ETD).

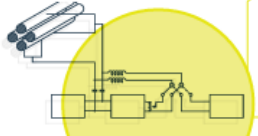


2013
Professor Zoltan Takats and his team present the concept of the iKnife, an intelligent knife that can analyze tissue during surgery.

1898	1910	1919	1943	1946	1953	1969	1974	1984	1985	1987	1999	2004	2005	2013	2014

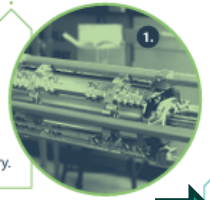
1910
J.J. Thomson separates particles with different mass-to-charge ratios.

1943
Consolidated Engineering Corporation (CEC) becomes the first organization to successfully market mass spectrometers.

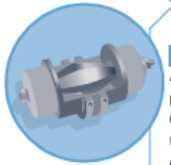


1953
Wolfgang Paul and Helmut Steinwedel present the concept of a quadrupole mass filter.

1974
Melvin Comisarow and Alan Marshall present Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry.



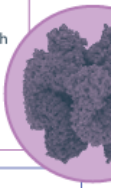
1985
Franz Hillenkamp, Michael Karas, and their colleagues present matrix-assisted laser desorption/ionization (MALDI).



1999
Alexander Makarov presents the concept of the Orbitrap™ mass analyzer at the ASMS annual conference.

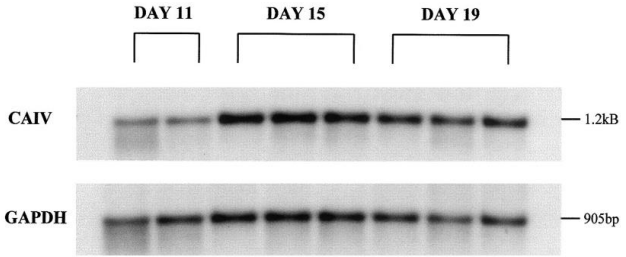
2005
Direct Analysis in Real Time (DART) is presented by Robert B. Cody at the January 2005 ASMS Sanibel Conference.

2014
Two separate research groups publish draft maps of the human proteome based on mass spectrometry.



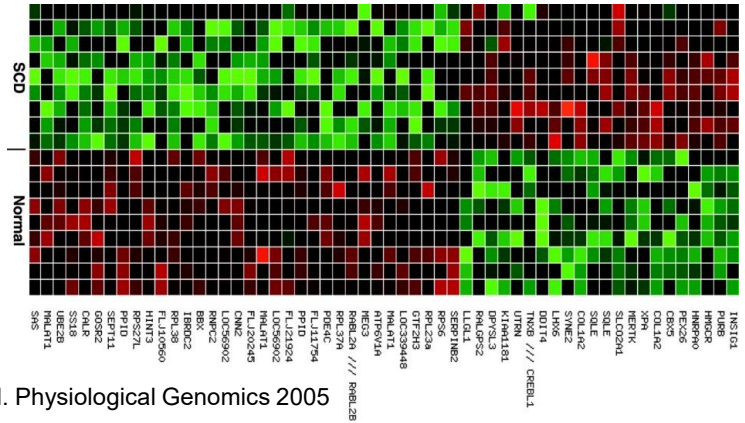
Digital transformation of biology

Qualitative experiment



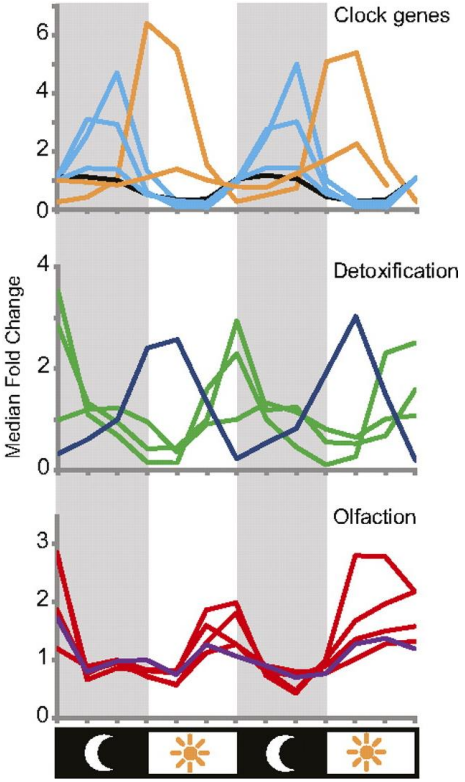
Rosen et al. Am J of Physiology 2001

High-throughput data



Klings et al. Physiological Genomics 2005

Time-series data

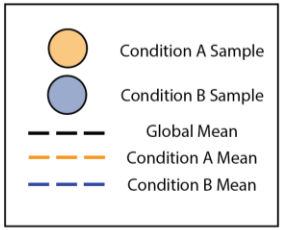


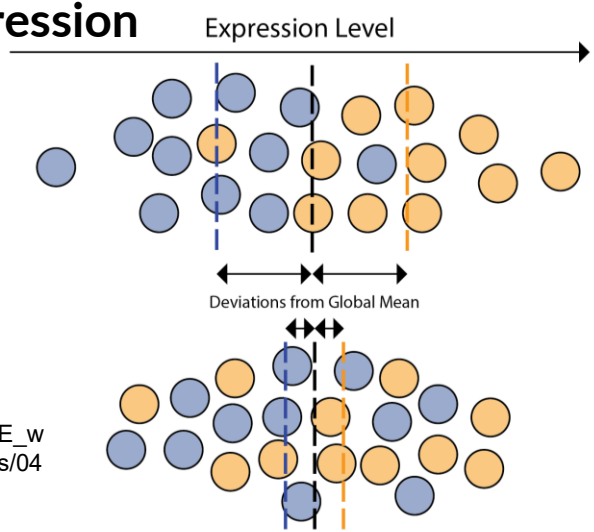
Rund et al. PNAS 2011

- Paradigm shift from data digitization and introduction of quantitative, high-throughput assays
- From hypothesis-driven to data-driven

Statistical approaches

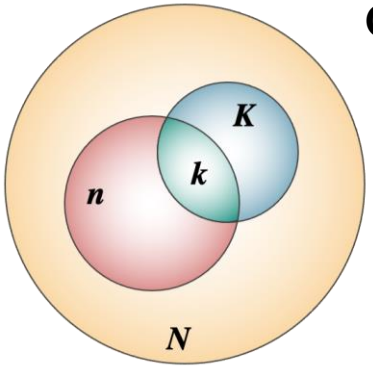
Differential Expression





https://hbctraining.github.io/DGE_workshop_salmon_online/lessons/04a_design_formulas.html

Overrepresentation



$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

N = Background *(e.g. Transcriptome ~ 40.000 genes)*

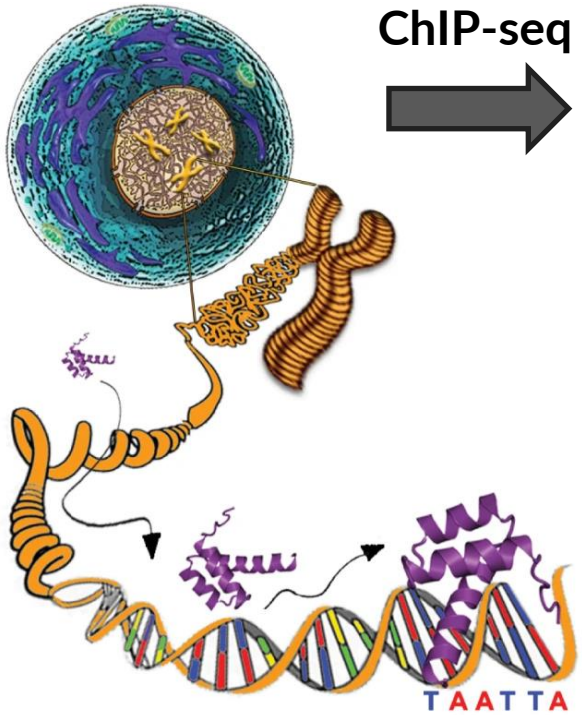
n = Query list *(e.g. upregulated genes)*

K = genes annotated in the pathway/set tested *(e.g. Glycolysis)*

$k = n \cap K$

- Theoretical models for data distribution
- Direct probability calculation to determine significance

Bioinformatics = algorithms for biological data



Castellanos *et al*. Nat Comm 2020

ChIP-seq

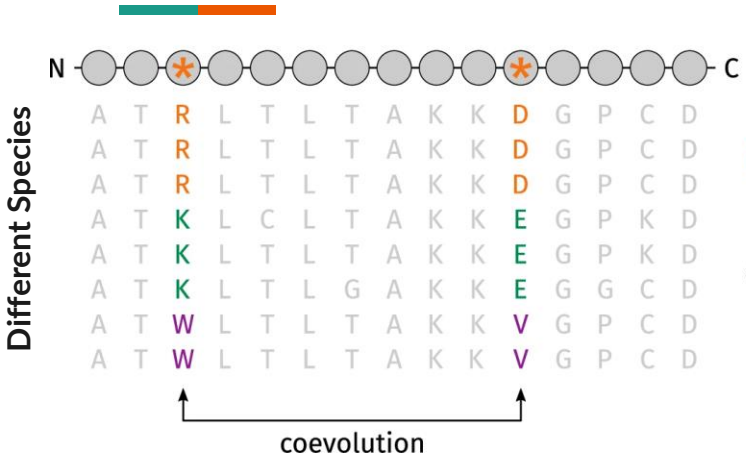
Gibbs Sampling

```
                CGGGGCTATcCAgCTGGGTCGTCACATTCCCCTT
TTTGAGGGTGCCCAATAAggGCAACTCCAAAGCGGACAAA
                   GGATGgAtCTGATGCCGTTTGACGACCTA
                AAGGAaGCAACcCCAGGAGCGCCTTTGCTGG
AGATTATAATGTCGGTCCtTGgAACTTC
CAACTGAGATCATGCTGCATGCcAtTTTCAAC
    TACATGATCTTTTGATGgcACTTGGATGAGGGAATGATGC
```

An Introduction to Bioinformatics Algorithm by Jones and Pevzner

- Use **algorithm** to find the optimal solutions with minimal resource requirement
- Use **statistics** to test whether the optimal solution is a *true signal*

Capturing co-evolution with mutual information

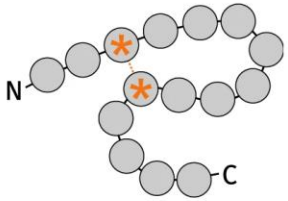


$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{(X,Y)}(x,y) \log \left(\frac{P_{(X,Y)}(x,y)}{P_X(x)\, P_Y(y)} \right)$$

Mutual Information

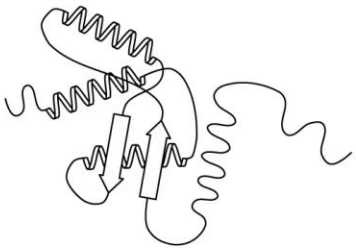
inference

constraint

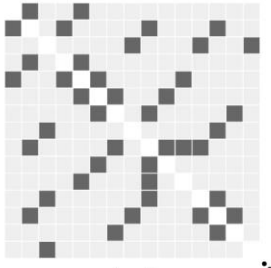


contact in 3D

Positions in protein(s) co-evolve because they interact in some fashions – usually physical bindings



reconstruct

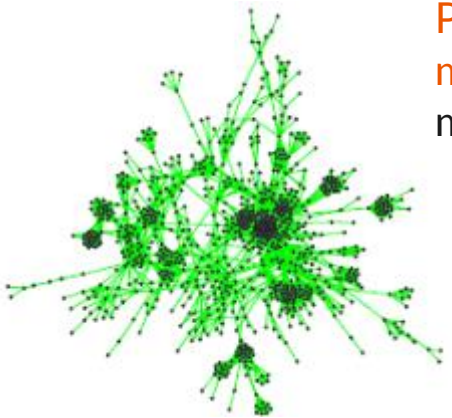


contact map

Co-evolution can be quantified and used to guide biological inferences

Bittrich *et al.* Sci Rep 2019

Using graph theory to identify important genes

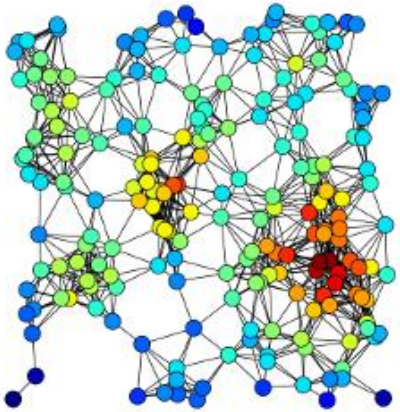


Proteins involved in many interactions might be important

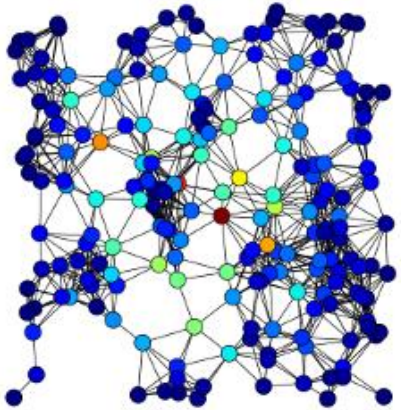
Proteins that connect other proteins might be important

Node = Protein
Edge = Protein-protein interaction

Degree



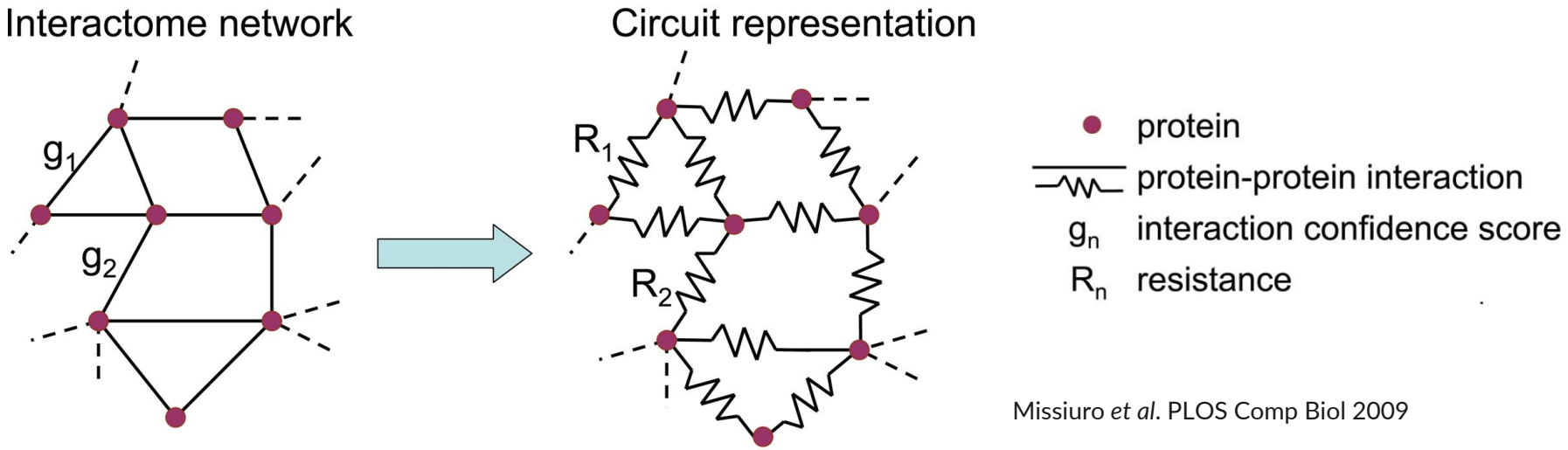
Betweenness



Images from wikipedia

- Genes and proteins work together as a single interaction network

Electrical circuit model for biological signal flows

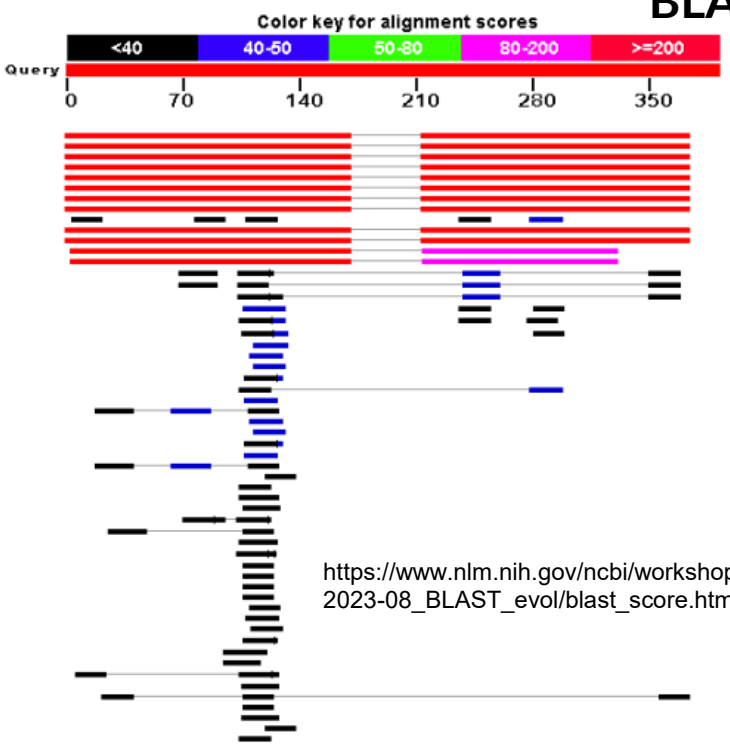


Missiuro *et al*. PLOS Comp Biol 2009

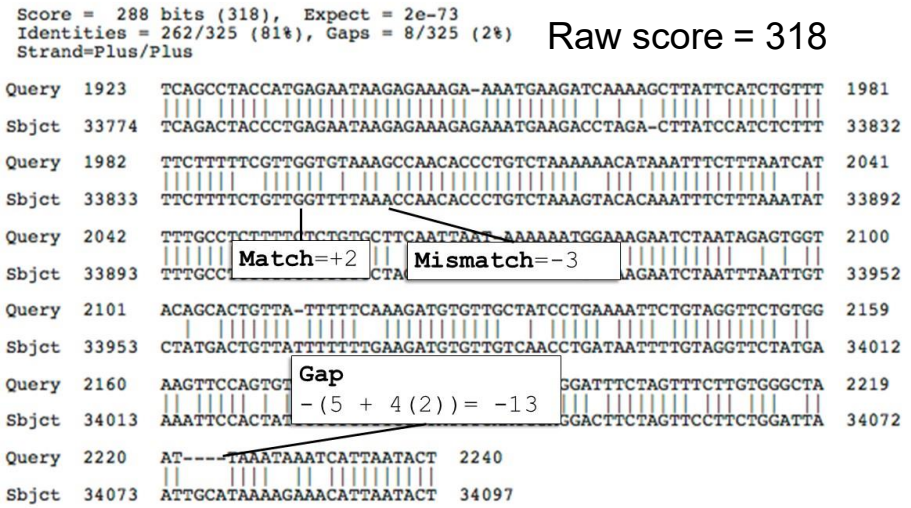
To determine how information flows between two nodes

Apply Kirchoff's laws to calculate the current through the circuit

Collaboration between algorithm and statistics



https://www.nlm.nih.gov/ncbi/workshops/2023-08_BLAST_evol/blast_score.html



- **The best solution can be wrong if the true answer is not among the options**
- Statistics can identify improbable answers

Overview of this year’s contents

Module 1: DNA sequencing and applications

2	3	18-Aug-25	Mon	10:00-11:30	Sequence alignment	
	4	20-Aug-25	Wed	10:00-11:30	DNA sequencing techniques and applications	**Assignment 1**
3	5	25-Aug-25	Mon	10:00-11:30	Sequencing data processing workflows	
	6	27-Aug-25	Wed	10:00-11:30	Genome assembly and annotation	
4	7	1-Sep-25	Mon	10:00-11:30	Variant calling and genome-wide association study	
	8	3-Sep-25	Wed	10:00-11:30	Phylogenetics and evolutionary models	**Assignment 2**
5	9	8-Sep-25	Mon	10:00-11:30	Metagenomics	
	10	10-Sep-25	Wed	10:00-11:30	ChIP-seq and DNA motif detection	

- What can be inferred from sequence data?
- DNA sequencing: which platform to use?
- Evolutionary perspective
- Genomics and metagenomics
- Discovery of DNA binding and regulatory motifs

Module 2: Transcriptomics

6	11	15-Sep-25	Mon	10:00-11:30	Transcriptomics techniques	
	12	17-Sep-25	Wed	10:00-11:30	Differential expression analysis	**Assignment 3**
7	13	22-Sep-25	Mon	10:00-11:30	Functional enrichment analysis	
	14	24-Sep-25	Wed	10:00-11:30	Single-cell and spatial techniques	
8	15	29-Sep-25	Mon	10:00-11:30	Single-cell data processing	

- Why is transcriptomics one of the most popular omics?
- How to identify differentially expressed genes?
- How to narrow down on affected pathways?
- Benefits of single-cell and spatial omics
- A quick look at single-cell transcriptomics data

Module 3: Broad fields in computational biology

	16	1-Oct-25	Wed	10:00-11:30	Proteomics and mass spectrometry	
9	17	6-Oct-25	Mon	10:00-11:30	Introduction to systems biology and dynamics	**Assignment 4**
	18	8-Oct-25	Wed	10:00-11:30	Multi-omics integration	
10	19	13-Oct-25	Mon		Holiday	
	20	15-Oct-25	Wed	10:00-11:30	Biological networks	
11	21	20-Oct-25	Mon	10:00-11:30	Chromatin conformation capture	**Assignment 5**
	22	22-Oct-25	Wed	10:00-11:30	RNA and protein structure models	
12	23	27-Oct-25	Mon	10:00-11:30	Online tools and resources	
	24	29-Oct-25	Wed	10:00-11:30	Microscopy data analysis	

- Study of proteins
- Systems and dynamics modeling
- Biological networks
- Structural models
- Imaging

Module 4: Computational tools

13	25	3-Nov-25	Mon	10:00-11:30	Essential statistics for computational biologist	**Assignment 6**
	26	5-Nov-25	Wed	10:00-11:30	Omics data visualization	
14	27	10-Nov-25	Mon	10:00-11:30	Introduction to machine learning and AI	
	28	12-Nov-25	Wed	10:00-11:30	Machine learning applications	
15	29	17-Nov-25	Mon	10:00-11:30	Foundational AI models in biology	
	30	19-Nov-25	Wed	10:00-11:30	Biomarker discovery	

- Statistics, revisited
- Data visualization
- Machine learning and AI
- Biomarker discovery

Assignments: What to expect?

- Analyze data using online tools and pre-built software like BLAST, MEGA, MaxQuant, Web-Gestalt, CellProfiler, Cytoscape, etc.
- Identify key details in research articles
- Design experiments using omics techniques and bioinformatics tools
- Critique LLM’s answers
- Edit R and Python scripts to analyze data

Any question or comment?

See you next time