

**Bioinformatics:
a Servant or the Queen of Molecular Biology?**

Pavel Pevzner

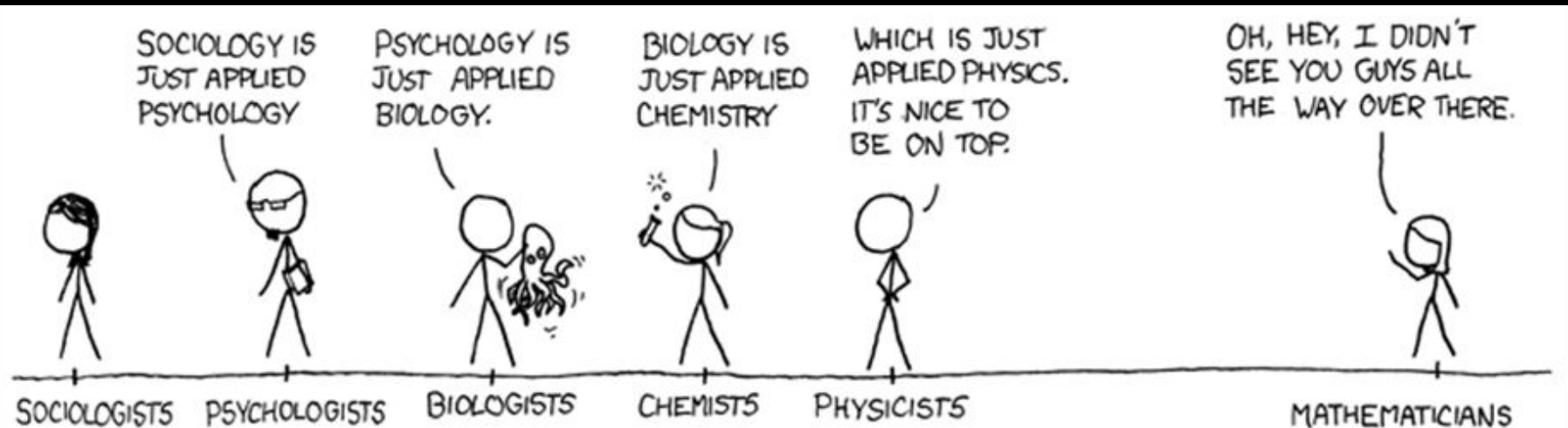
**Department of Computer Science and Engineering
University of California at San Diego**

Mathematics is the queen of the sciences...

Carl Friedrich Gauss

Mathematics is the queen of the sciences...
She often condescends to render service to
astronomy and other natural sciences, but in
all relations, she is entitled to first rank.

Carl Friedrich Gauss



Bioinformatics is the queen of **biology**... She often condescends to render service to **genomics, proteomics and other omics**, but in all relations, she is entitled to first rank.

Bioinformatics is the queen AND servant of biology...



Dogma

- **Bioinformatics: Queen or Servant?**
- **Bioinformatics refutes biological dogmas**
 - Protein half-life and the N-end rule
 - Master Alu Theory
 - Random Breakage Model
 - ...



- **A 1000-year old dogma**

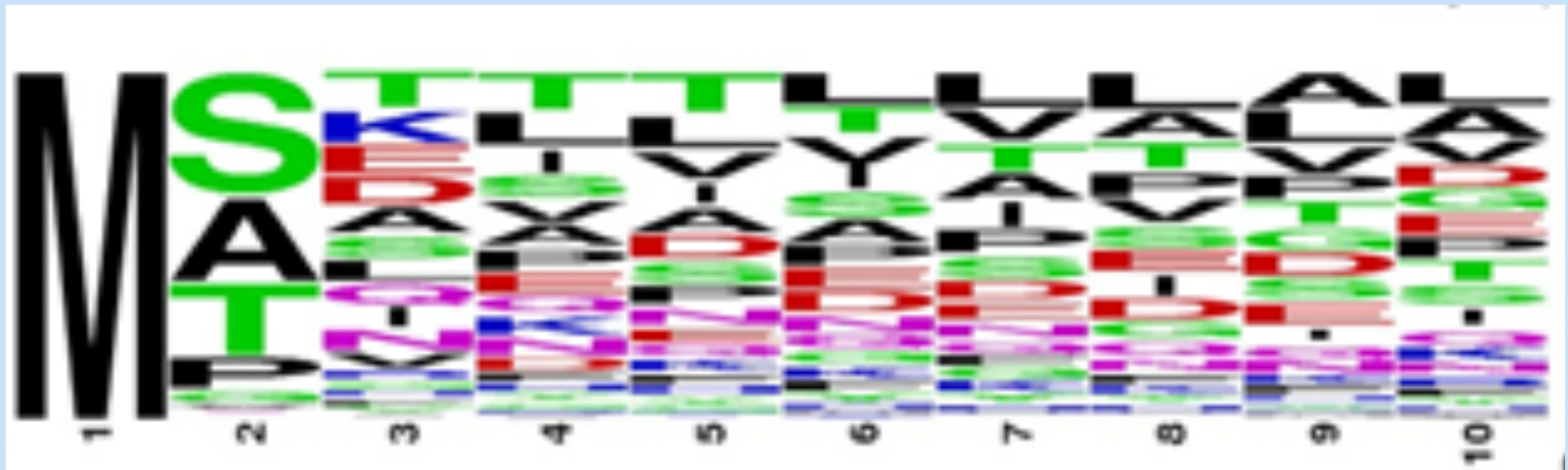
What controls protein half-lives?

Protein half-lives vary by five orders of magnitude (from seconds to days).

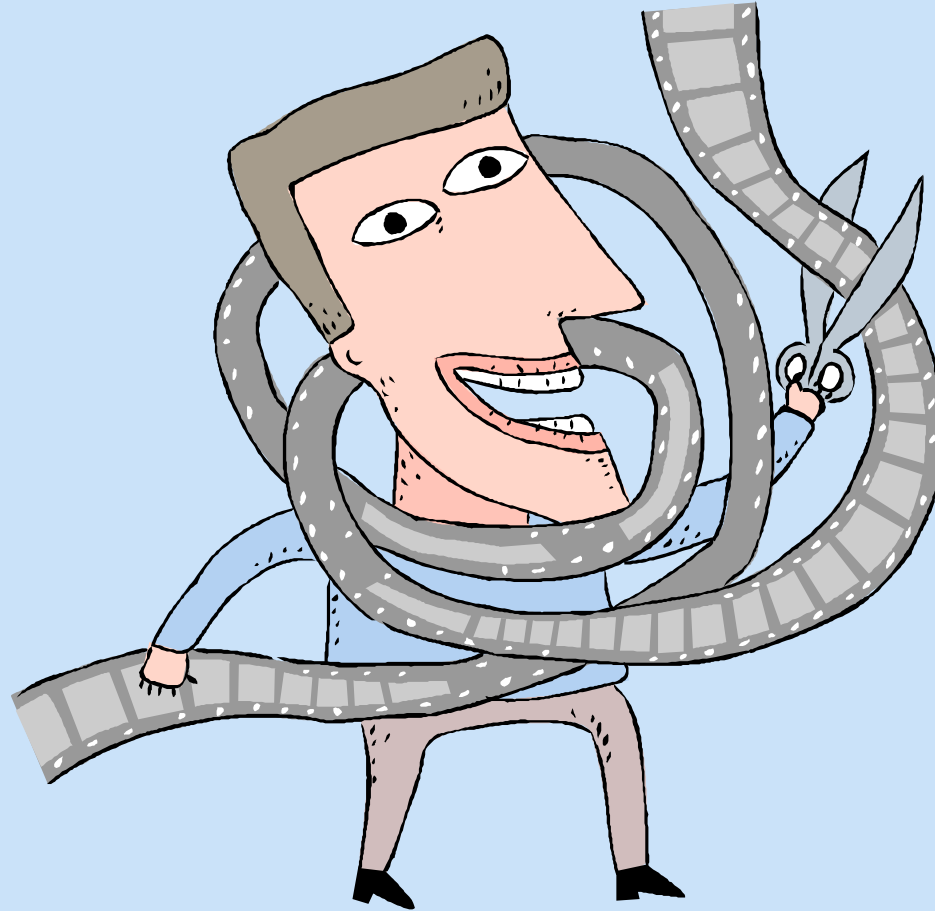


Lens proteins do not turnover leading to cataracts

All proteins start from Methionine



N-terminal Methionine Excision (NME)



Universal presense of NME in all species suggests that it is critically important (knock-outs are lethal)

2nd aa in the protein controls NME

7 amino acid triggering NME

13 other amino acids

M**G**XXXXXXXX
M**A**XXXXXXXX
M**P**XXXXXXXX
M**S**XXXXXXXX
M**T**XXXXXXXX
M**V**XXXXXXXX
M**C**XXXXXXXX

M**D**XXXXXXXX
M**N**XXXXXXXX
M**L**XXXXXXXX
M**I**XXXXXXXX
M**Q**XXXXXXXX
M**E**XXXXXXXX
M**H**XXXXXXXX
M**M**XXXXXXXX
M**F**XXXXXXXX
M**K**XXXXXXXX
M**Y**XXXXXXXX
M**W**XXXXXXXX
M**R**XXXXXXXX

2nd aa in the protein controls NME

7 amino acid triggering NME: 13 other amino acids

GXXXXXX
AXXXXXX
PXXXXXX
SXXXXXX
TXXXXXX
VXXXXXX
CXXXXXX

MDXXXXXX
MNXXXXXX
MLXXXXXX
MIXXXXXX
MQXXXXXX
MEXXXXXX
MHXXXXXX
MMXXXXXX
MFXXXXXX
MKXXXXXX
MYXXXXXX
MWXXXXXX
MRXXXXXX

What NME has to do with protein half-life?



Varshavsky's Contributions

WE ARE WRITING TO EXPRESS OUR enthusiasm that the discovery of the ubiquitin conjugation system has been acknowledged with the award of the Nobel Prize in

Chemistry to three outstanding biochemists: Avram Hershko and Aaron Ciechanover of the Technion Institute in Israel, and Irwin Rose of the University of California at Irvine ("Gold medal from cellular trash," *G. Vogel, News Focus*, 15 Oct., p. 400). Unraveling the chemistry that underlies the attachment of ubiquitin to proteins that are destined to be degraded was a magnificent achievement and is fully deserving of this recognition.

The mechanism of ubiquitin conjugation and its role in proteolysis was selected for recognition by the Nobel Committee in large part because of the vital role that the ubiquitin-proteasome system (UPS) plays in the physiology of cells and organisms. Investigations on the physiological functions of the UPS, which dominate current research in this field, were pioneered largely by Alexander Varshavsky of the California Institute of Technology. Several core principles that guide our current understanding of the ubiquitin system had their origins in Varshavsky's work, including the following: (i) the UPS is the predominant mechanism for selective protein turnover in the cytoplasm and is essential for cellular function; (ii) in addition to its role in turning over damaged proteins, the UPS controls diverse physiological processes such as the cell

cycle, DNA repair, and stress responses; (iii) ubiquitin ligases (E3s) are highly specific receptors that underlie the remarkable specificity of ubiquitination by binding to defined sequences within proteins (degrons); and (iv) a ubiquitin chain linked via the lysine-48 residue of ubiquitin governs targeting of substrates to the proteasome for degradation. In addition to these fundamental contributions, Varshavsky ushered the ubiquitin field into the age of molecular genetics by identifying mutants and characterizing

“ We suggest that the impact of Varshavsky's work on the physiology of the ubiquitin system and its relationship to fundamental processes such as mitosis and chromosome segregation justifies serious consideration for a future Nobel Prize in Physiology or Medicine. ”

—BAUMEISTER ET AL.

Baumeister et al., *Science*, 2004

ioemag.org on March 12, 2009

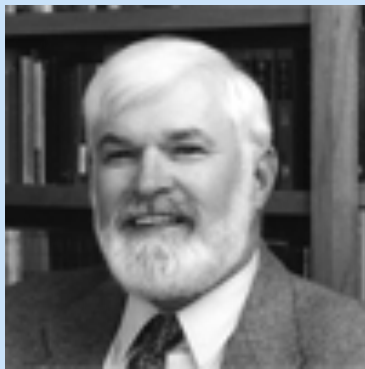
Varshavsky's pioneering work included the **N-end theory** connecting NME and protein half-life (Varshavsky, *Science*, 1986)

NME regulates protein half-life!

NME regulates protein half-life

Varshavsky et al., *Science* 1986

Arfin&Bradshaw, *Biochemistry* 1988

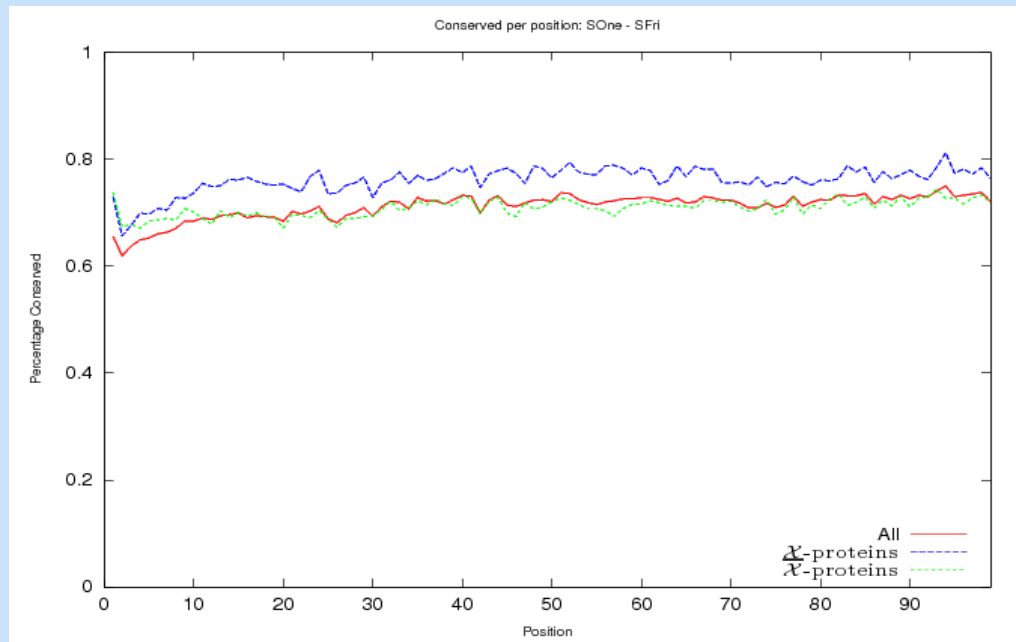


Evolutionary conservation of NME

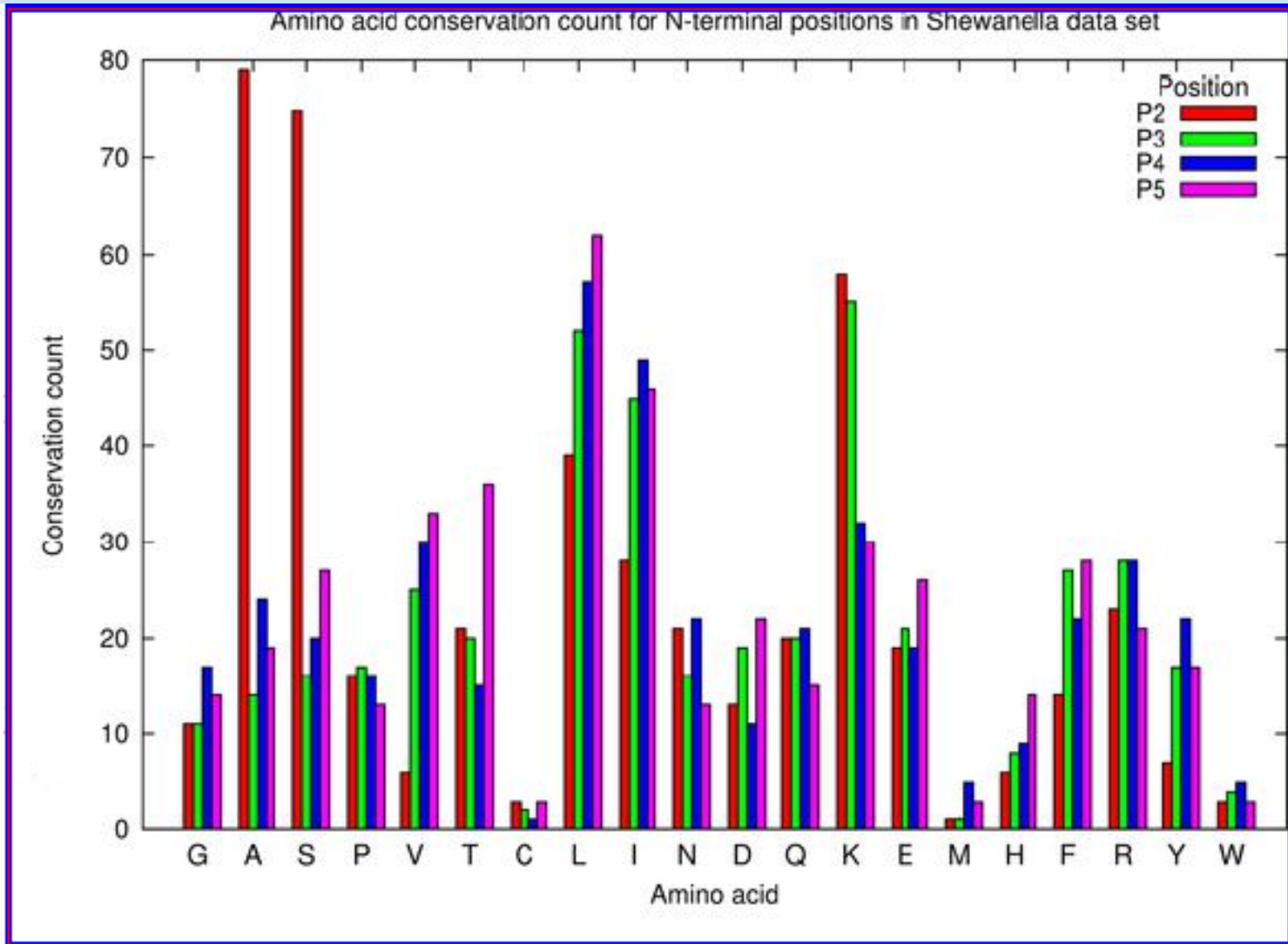
- If NME is important for the proper functioning of a particular protein, it should be evolutionarily **conserved**: position 2 should be more conserved than other positions in the 2-letter **blue-red** alphabet (**7 NME-triggering** aa vs **13 others**)

Position 2 is not conserved!

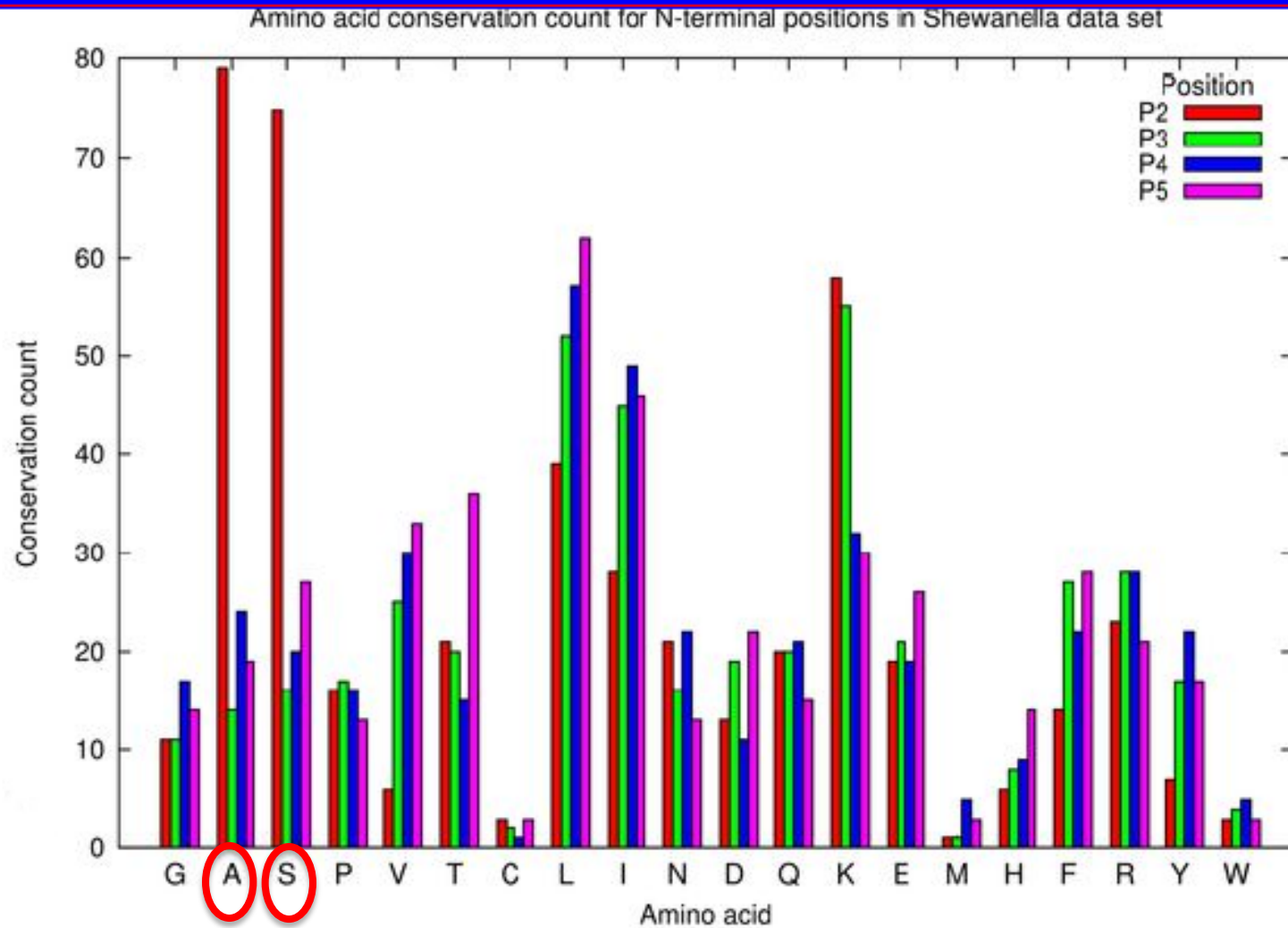
- If NME is important for the proper functioning of a particular protein, it should be evolutionarily **conserved**: position 2 should be more conserved than other positions in the 2-letter **blue-red** alphabet (**7 NME triggering** aa vs **13 others**)



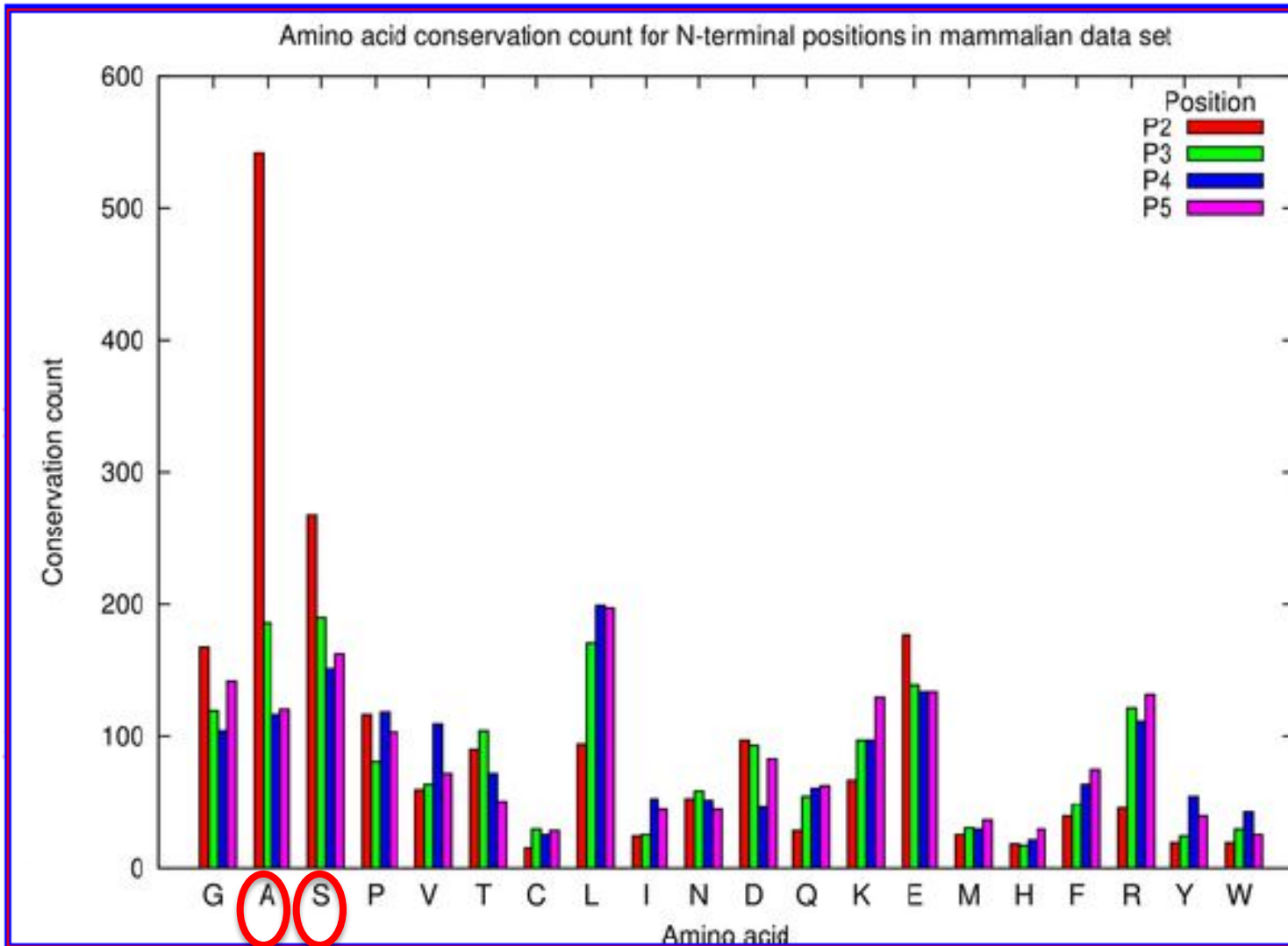
Conservation of INDIVIDUAL amino acids



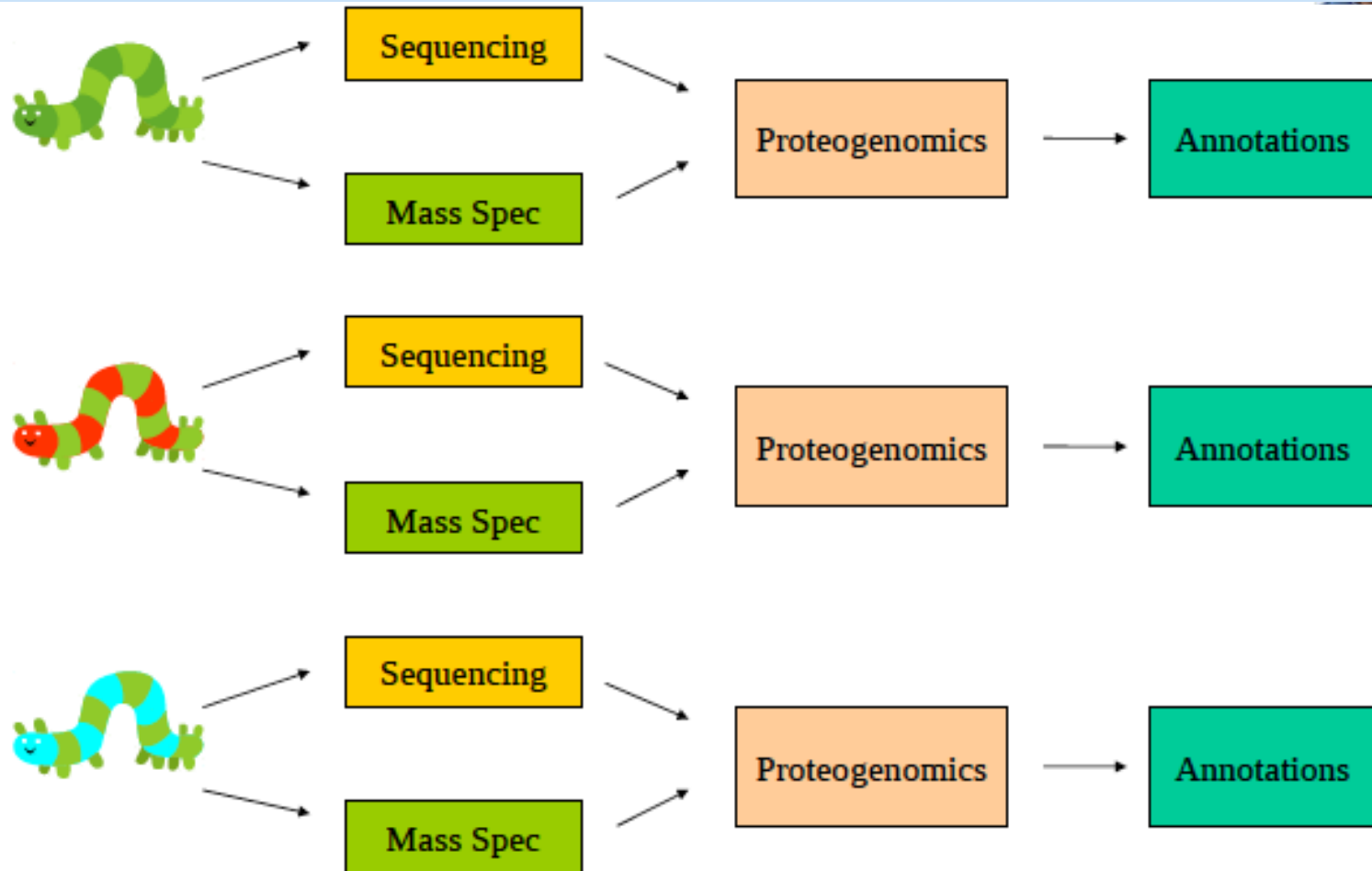
Alanine and Serine!



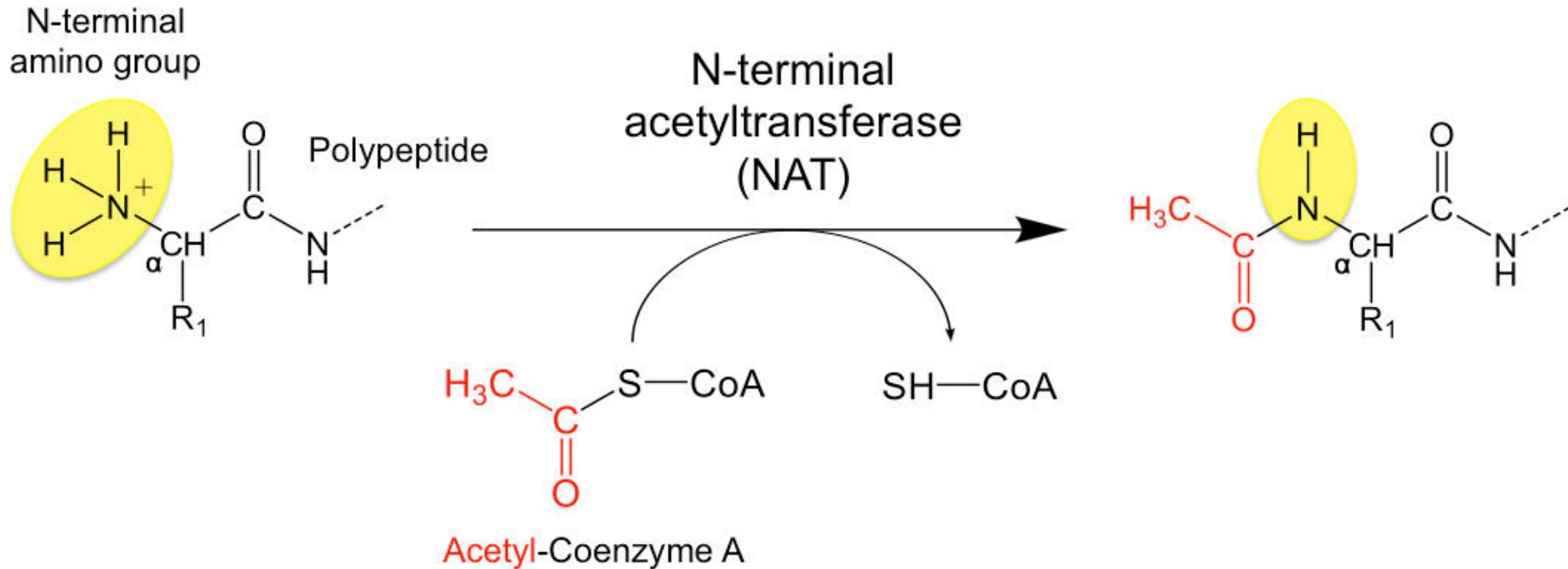
Alanine and Serine!



Comparative Proteogenomics



N-terminal acetylation



Fifty years after its discovery, the function of the N-terminal acetylation remains enigmatic.

Arnesen, *PLoS Biology* 2011

NME has nothing to do with protein half-life!

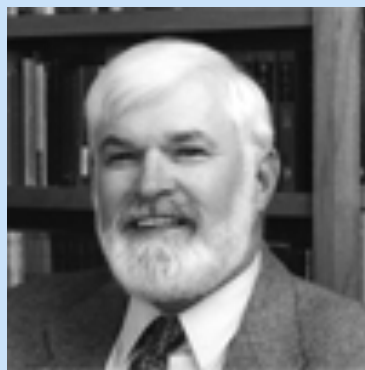
NME regulates protein half-life

Varshavsky et al., *Science* 1986
Arfin & **Bradshaw**, *Biochemistry* 1988

NME has nothing to do with protein half-life

- NME exposes Ala and Ser for N-terminal acetylation
- **Guilt by association:** Other 5 aa triggering NME are bystanders

Bonissone, Gupta, **Bradshaw**, PP.
Molecular & Cell Proteomics 2013



The Master Alu Repeat theory

A single master Alu repeat have colonized the human genome.

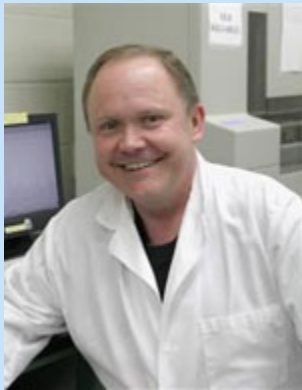
Shen,Deininger,**Batzer**
J. Mol. Evolution 1991

Many Alu repeats have colonized the human genome

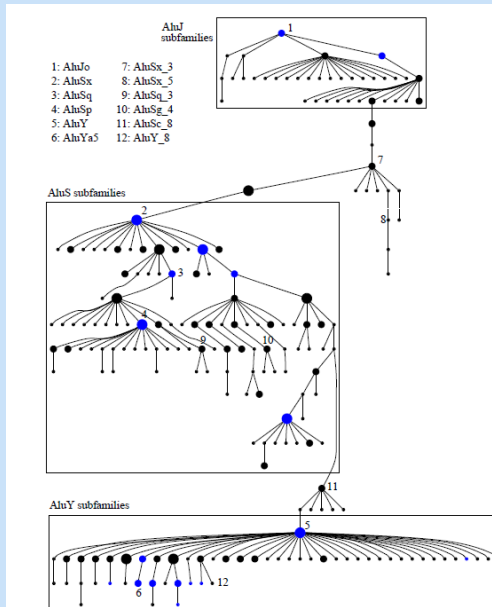
Price, Eskin, PP
Genome Research
November 2004

Many Alu repeats have colonized the human genome

Cordeau,Hedges,**Batzer**
Trends in Genetics
October 2004



Mark Batzer



Random Breakage Model of Chromosome Evolution

Random Breakage Model:
rearrangements occur at
random positions in genome
(no fragile regions).

Ohno, *Science* 1973

Nadeau&Taylor, *PNAS* 1984



Fragile Breakage Model
genome is a mosaic of solid
and fragile region

PP&Tesler *PNAS* 2004



Genomic distance theorem:

Hannenhalli and PP. *FOCS* 1995, *STOC* 1995



- Bioinformatics: Queen or Servant?
- Bioinformatics refutes biological dogmas
- **A 1000-year old dogma**





laurentius. cc.

Education is the one major activity which has thus far not applied ingenuity to the solution of its problems.

Sydney Pressey, 1932

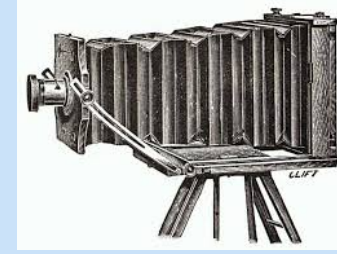
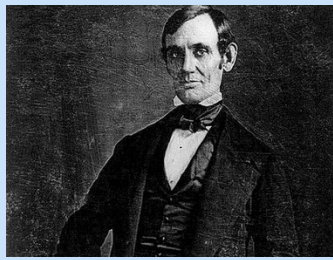


THE NEW FILM BY JOEL AND ETHAN COEN

A SERIOUS MAN



Daguerreotypist



1839

Match maker



1995

Travel agent



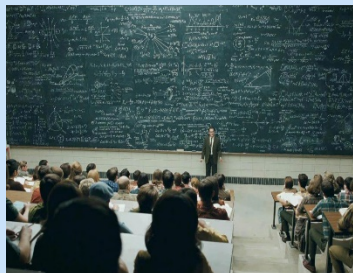
1996

Taxi driver

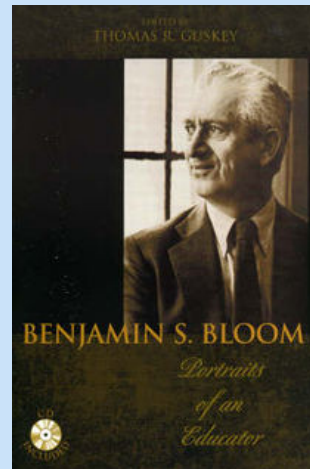


2025?

Professor???

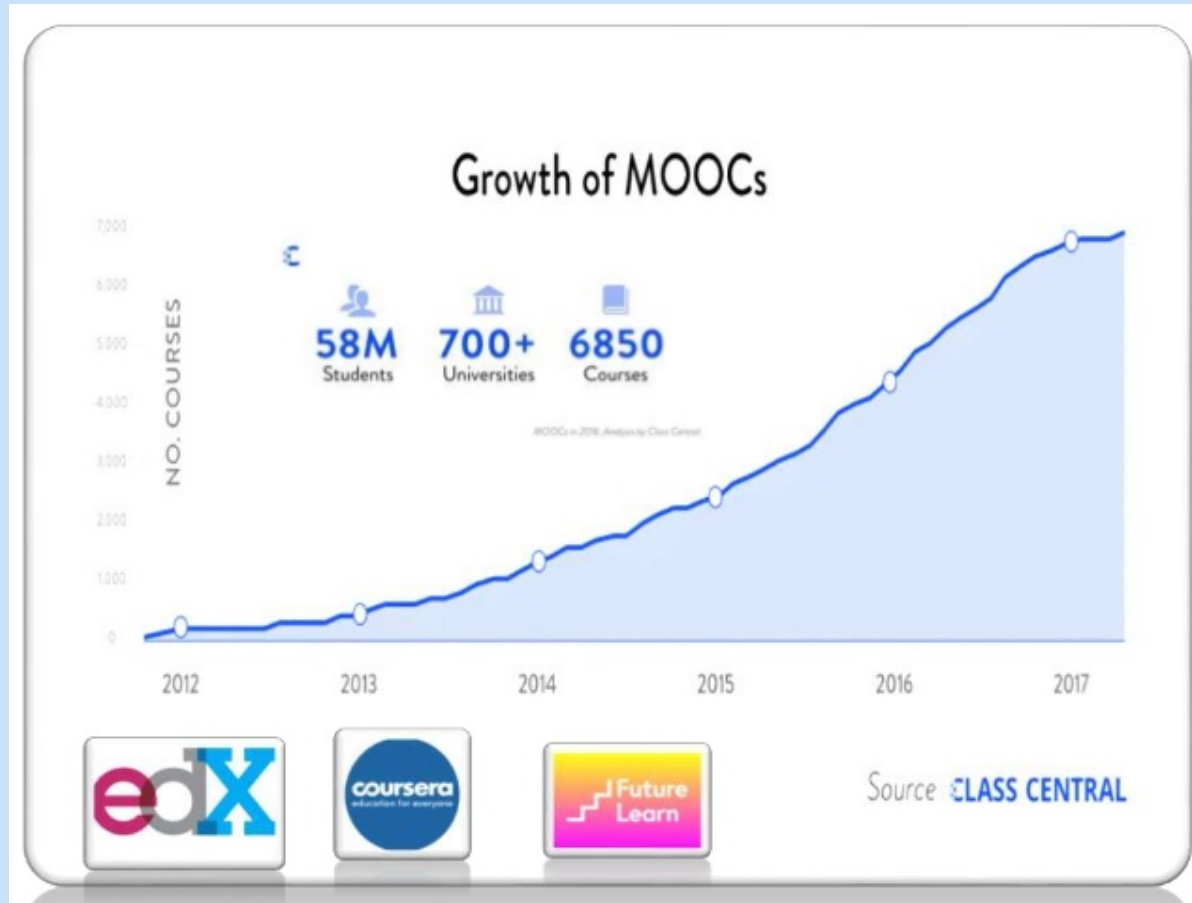


- **Bioinformatics: Queen or Servant?**
- **Bioinformatics refutes biological dogmas**
- **A 1000-year old dogma**
 - **Classroom: an Optimal Learning Experience**



- What Is Wrong with MOOCs?
- Massive Adaptive Interactive Texts (MAITs)
- MAIT: a Solution of the Bloom Conundrum?

Massive Online Open Courses (MOOCs) Revolution



Massive Online Open Courses (MOOCs) Revolution

THE CHRONICLE
of Higher Education

Slate

The New York Times

A MOOC Delusion:
Why Visions to Educate the
World are Absurd

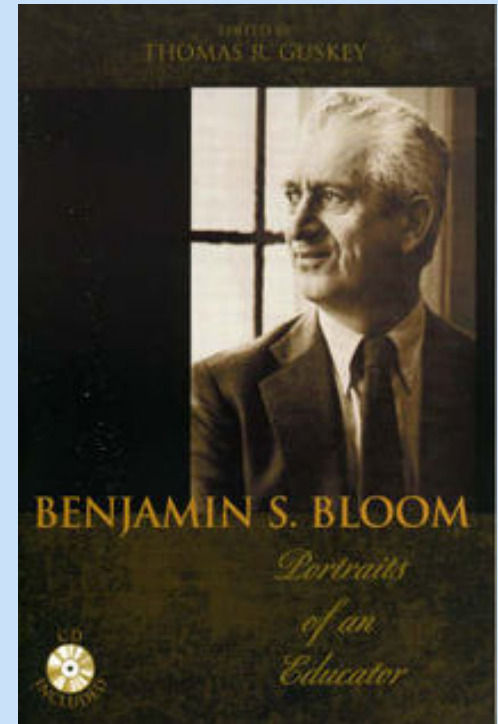
The MOOC Racket:
Widespread online-only higher ed
will be disastrous for students —
and most professors.

**Professors at San Jose State
Criticize Online Course**



Classroom instructions provide optimal form of learning!

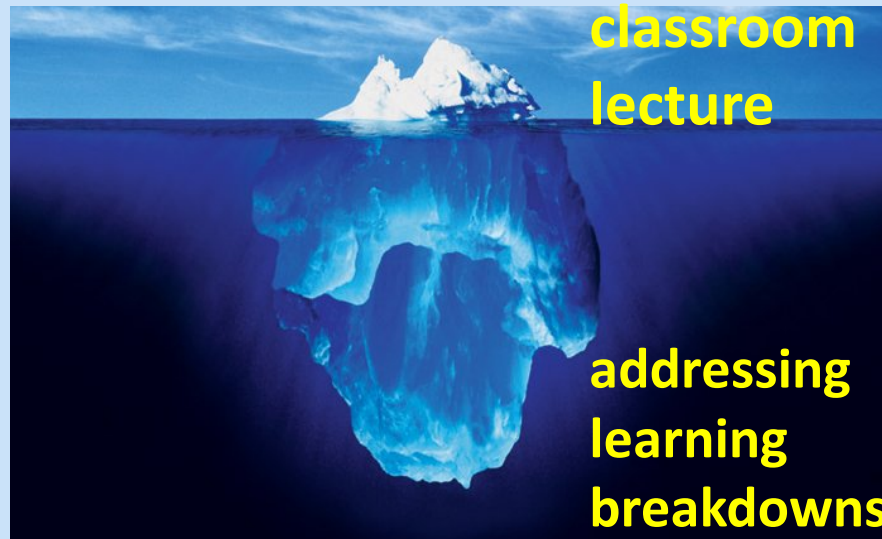
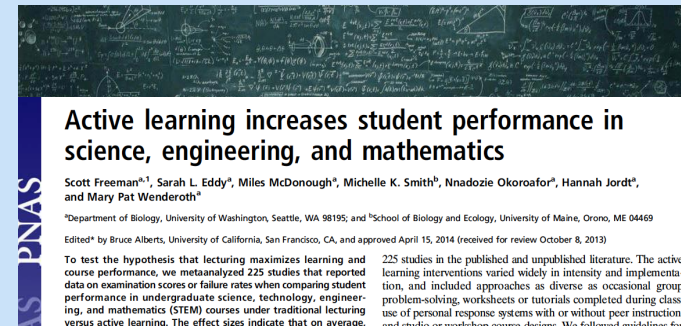
- In 1984, Benjamin Bloom demonstrated that traditional one-to-many classroom is minimally effective.



- Bloom: *It makes no sense to expect all students to take the same amount of time to achieve the same educational objectives*

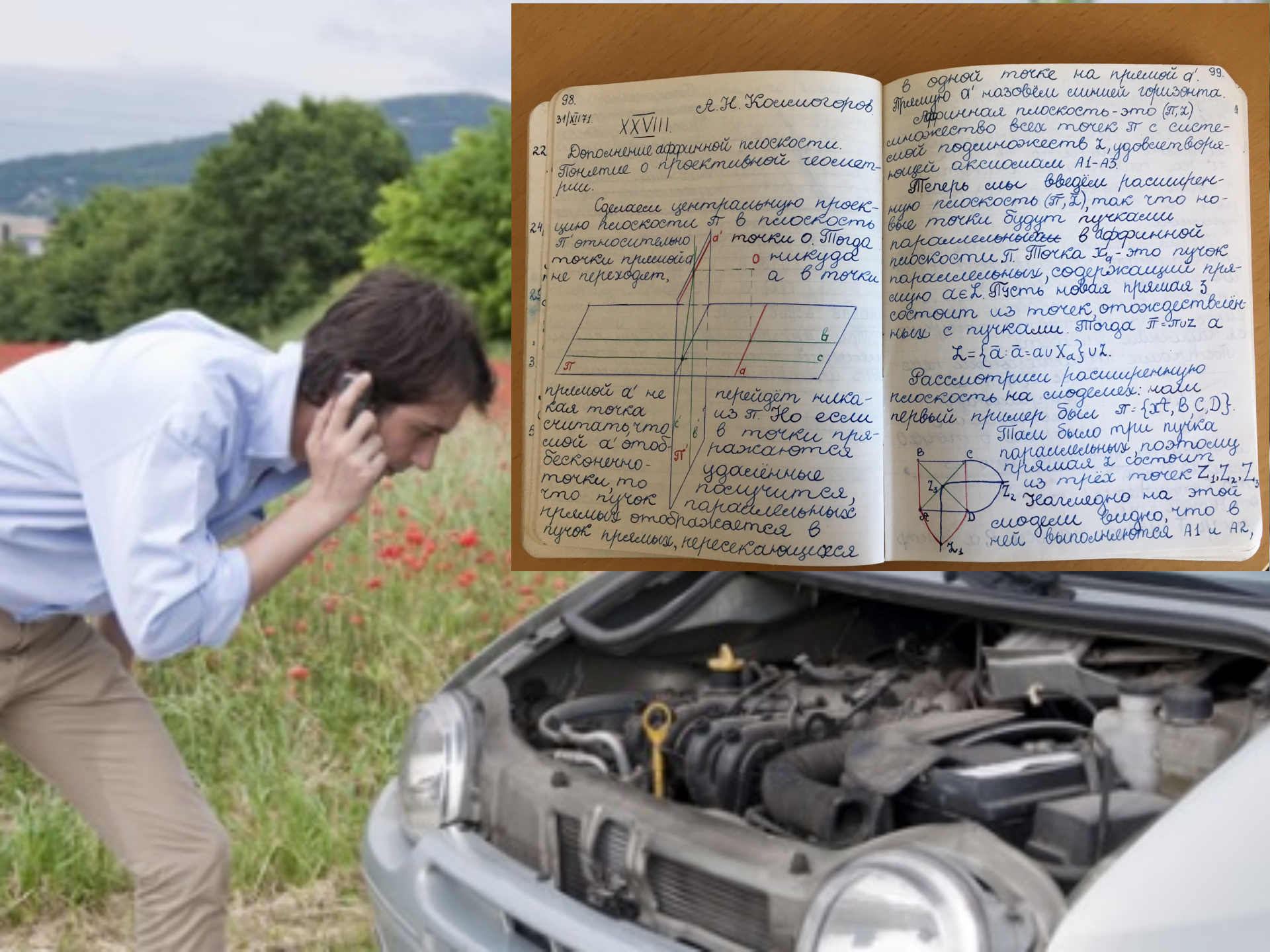
Classroom versus MOOC

- *Classroom instruction, in difference from MOOCs, uniquely personalizes learning.*
- **Really? In a 300-student classroom???**
- What is about the evidence otherwise:
Freeman, 2014 *Proc. Natl. Acad. of Science*
- Science/Technology/Engineering/Math education is a non-stop series of painful **learning breakdowns**



My Struggle at School





98.

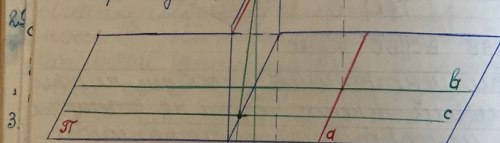
31/XII/71

XXVIII.

Л.Н. Кошмогоров

22. Дополнение аффинной плоскости.
Понятие о проективной геометрии.

Сделаем центральную проекцию плоскости π в плоскость π' относительно точки O . Тогда точки прямой a не переходят, а в точки



прямой a' некая точка. Скажем, что прямой a' отодвинем бесконечно-то, что пучок прямой отображается в пучок прямой, перейдет ника- из π . Но если в точки прямой a удалимся, пучок перейдет, параллельные пучок прямой, пересекающиеся

в одной точке на прямой a' .
Прямую a' назовем линией горизонта.

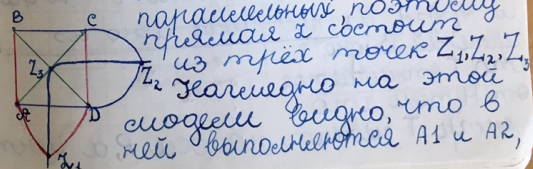
Аффинная плоскость - это (π, ℓ) множество всех точек π с системой координат ℓ , удовлетворяющей аксиоме А1-А3.

Теперь мы введем расширенную плоскость (π, ℓ) , так что новые точки будут пучками параллельных в аффинной плоскости π . Пучок X_a - это пучок параллельных, содержащий прямую $a \in \ell$. Пусть новая прямая ℓ состоит из точек отождествленных с пучками. Тогда $\pi = \pi \cup \ell$

$$\ell = \{\bar{a} : a = avX_a\} \cup \ell.$$

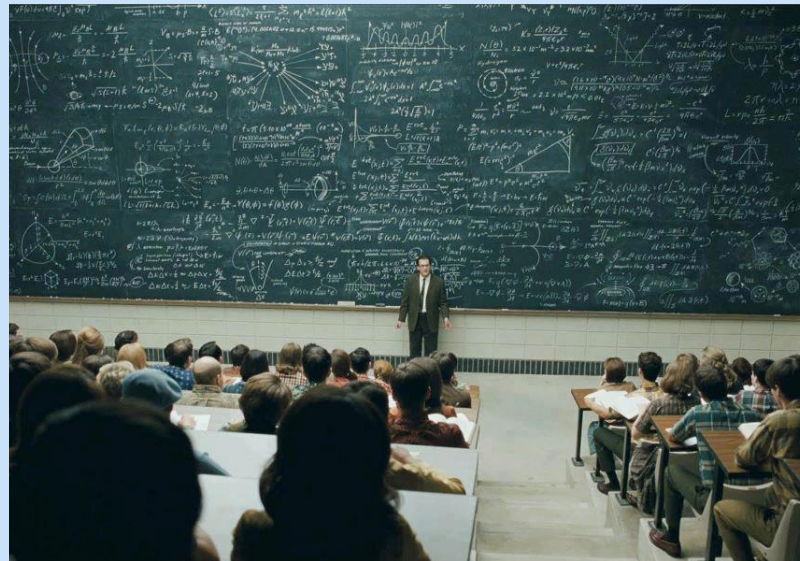
Рассмотрим расширенную плоскость на примере: найм первый пример для $\pi = \{x, y, z, d\}$.

Получим три пучка параллельных, поэтому прямая ℓ состоит из трех точек Z_1, Z_2, Z_3 . Каждому на этой модели видно, что в ней выполняются А1 и А2,

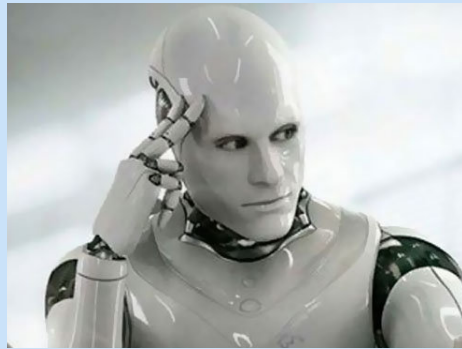


Limitations of a traditional classroom

- Most instructors teach to only a certain percentile of the class
- Instructors lack information about the learning breakdowns of *individual* students.
- *Individual* students do not receive the immediate feedback to address their learning breakdowns.



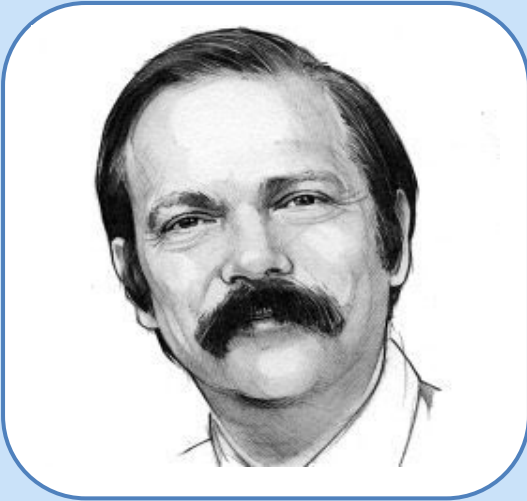
- **Bioinformatics: Queen or Servant?**
- **Bioinformatics refutes biological dogmas**
- **A 1000-year old dogma**
 - Classroom: an Optimal Learning Experience
 - **What Is Wrong with MOOCs?**



- Massive Adaptive Interactive Texts (MAITs)
- MAIT: a Solution of the Bloom Conundrum?



Today, the lion's share of MOOCs are
Massive Open Online **Hoarding** Classes



Vardi 2012

*"If I had my wish, I would wave a wand and make MOOCs **disappear**."*



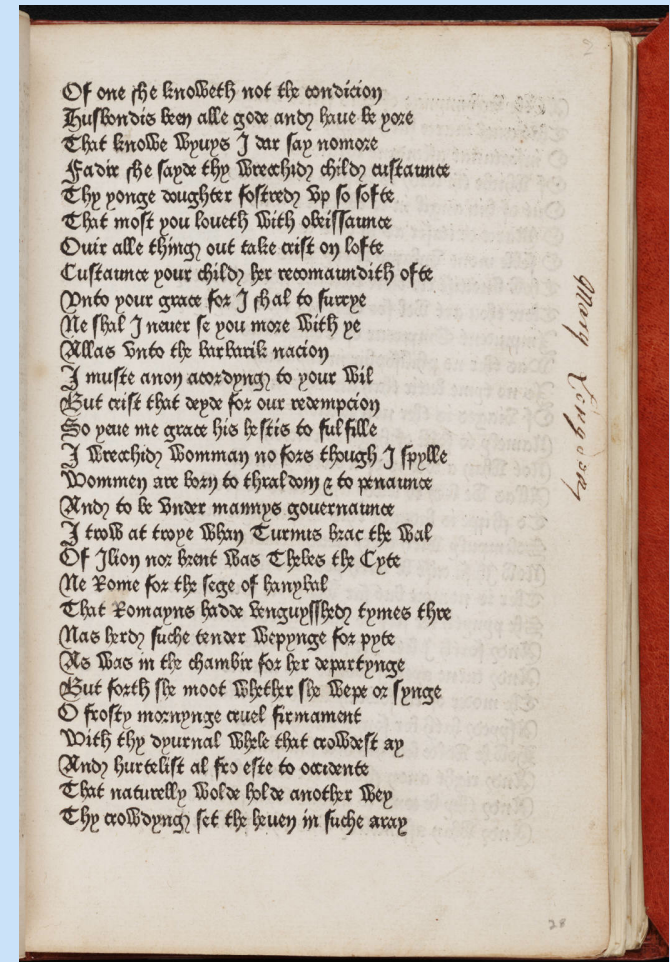
Trithemius 1492

*"The printed book is made of paper, and like paper, will soon **disappear**."*

Which Would You Prefer?



OR



Which Would You Prefer?



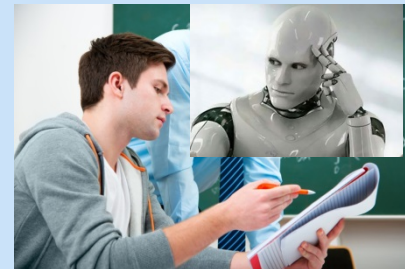


From MOOC to MAIT

Massive
Open
Online
Course

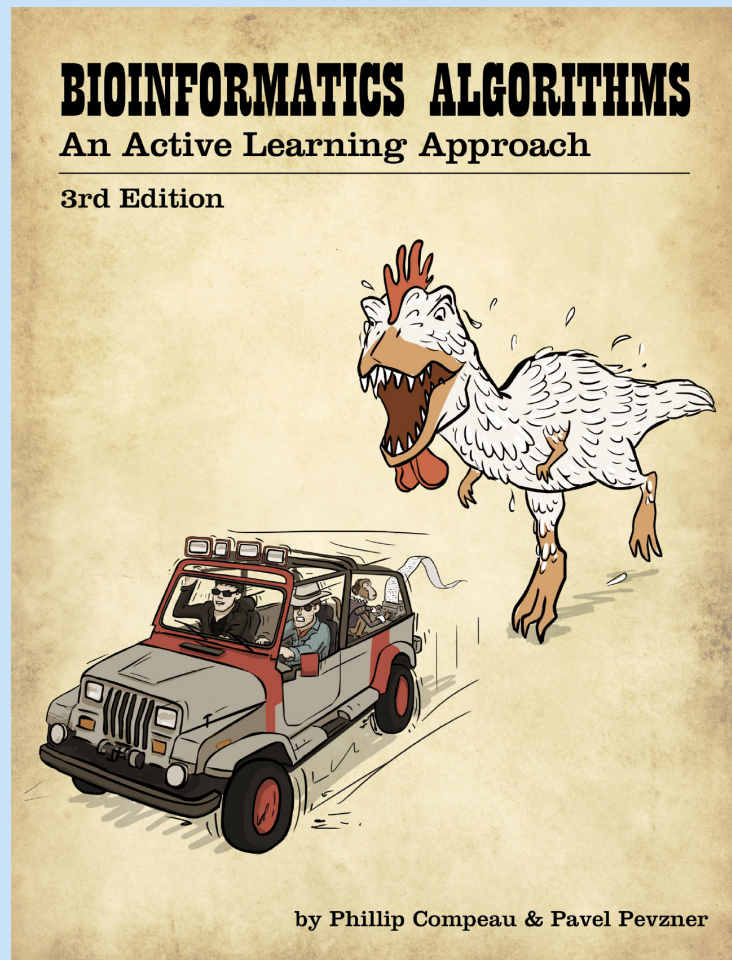


Massive
Adaptive
Interactive
Text

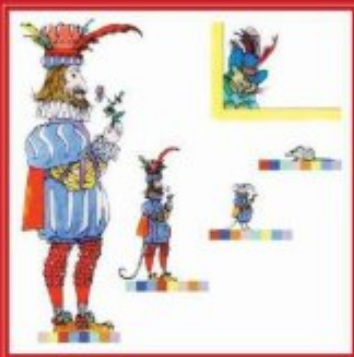


Towards Massive Adaptive Interactive Text (MAIT)

- Launched as a Coursera MOOC in 2013
- MOOCBook published in 2014 (3rd edition published in 2018)



Copyrighted Material



Pavel A. Pevzner

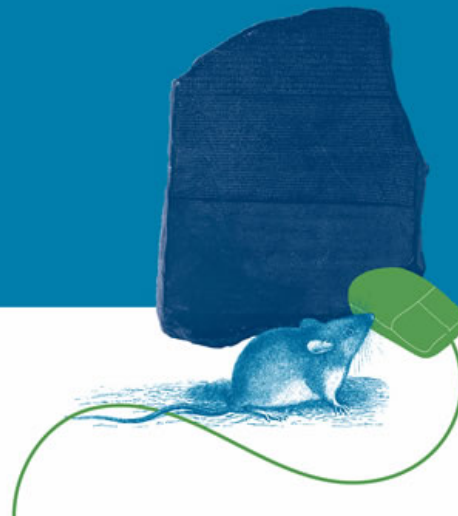
Computational Molecular Biology

An Algorithmic Approach

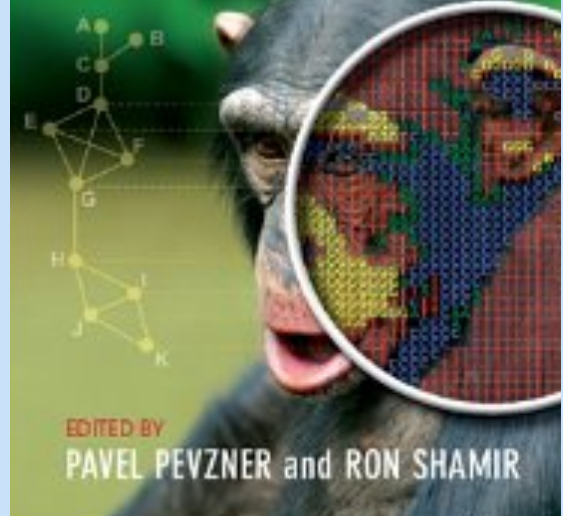
Copyrighted Material

AN INTRODUCTION TO BIOINFORMATICS ALGORITHMS

NEIL C. JONES AND PAVEL A. PEVZNER



BIOINFORMATICS FOR BIOLOGISTS

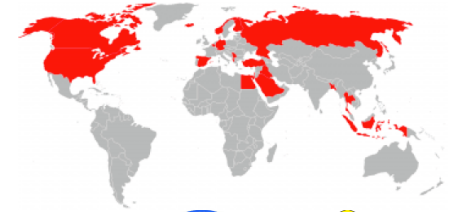


EDITED BY
PAVEL PEVZNER and RON SHAMIR

Adopted at 100+ Universities in 30+ Countries

UC San Diego

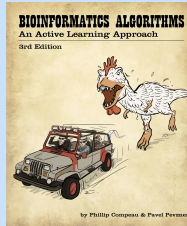
Carnegie Mellon University



UNIVERSITY OF ICELAND



End of a Classroom As I Know It: My MAIT is a Better Teacher than Me!



CMSC423, Bioinformatic
Algorithms, Databases and Tools
(Fall 2014)

Course Calendar

Course Logistics

Resources

Syllabus



CMSC423, Bioinformatic Algorithms, Databases and Tools (Fall 2014)

Course Information

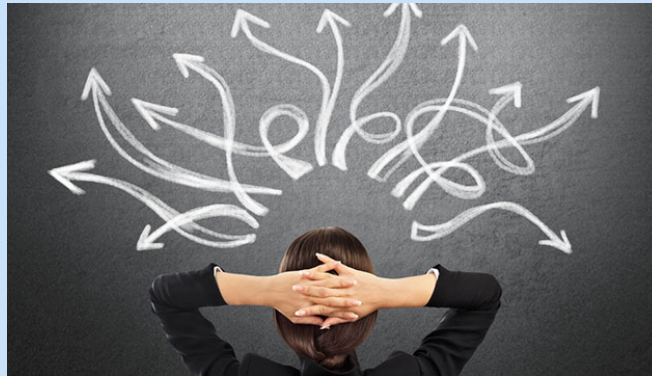
- Instructor:
[Héctor Corrada Bravo](#)
Center for Bioinformatics and Computational Biology
Department of Computer Science
hcorrada@umiacs.umd.edu
Office: 3114F Biomolecular Sciences Building
Phone Number: 301-405-2481
- Lecture Meeting times

Monday and Wednesday, 3:30pm-4:45pm
Room CSI 1121
- Office Hours: Friday 1:00pm-2:00pm AVW 3223 and by appointment
- TA: Wikum Dinalankara
wikum@cs.umd.edu
Office Hours: Thursday 1:00pm-3:00pm BSB 3119
- Textbook information: We will be using [Compeau, P. and Pevsner, P., Bioinformatics Algorithms, An Active Learning Approach. Active Learning Publishers, La Jolla, CA, 2014.](#)

From class syllabus:

“Class meeting times will not be lecture time”

- **Bioinformatics: Queen or Servant?**
- **Bioinformatics refutes biological dogmas**
- **A 1000-year old dogma**
 - Classroom: an Optimal Learning Experience
 - What Is Wrong with MOOCs?
 - **Massive Adaptive Interactive Text (MAIT)**



- MAIT: a Solution of the Bloom Conundrum?

Large Development Team



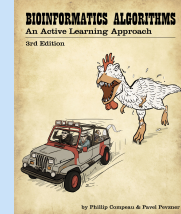
- Kolowich 2013: Average time spent developing MOOC is **~100s hours**



Large Development Team



- Kolowich 2013: Average time spent developing MOOC is **~100s hours**
- **7,000+ hours** already spent on



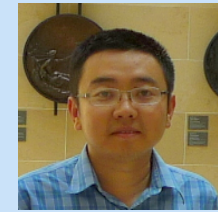
Bioinformatics Algorithms Development Team



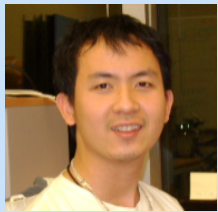
Nikolay Vyahhi
Rosalind founder,
Stepic CEO, Coursera
co-instructor



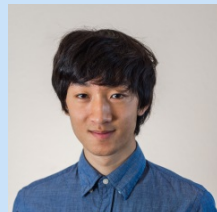
Olga Botvinnik
course/graphics
development



Yu Lin, Ph. D.
course
development



Son Pham, Ph. D.
invited lecturer,
course developer



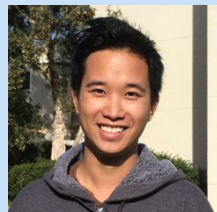
Max Shen
course/software
development



Robin Betz
course
development



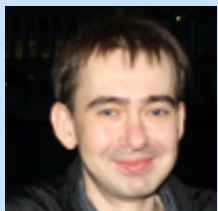
Kai Zhang
chief assessment
programmer



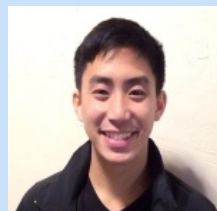
Vu Ngo
course/software
development



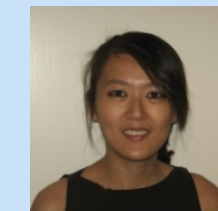
Mark Mammel
Coursera TA



Alexei Balandin
chief Rosalind
programmer



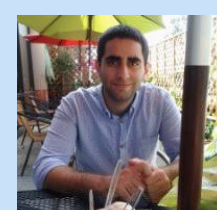
Jeffrey Yuan
course/software
development



Isabel Lupiani
Coursera TA



Randy Christopher
resident artist



Niema Moshri
course
development



Glenn Tesler
content
feedback

Rosalind: Joint Russian-US Educational Project

- Online bioinformatics platform founded by

- Nikolay Vyahhi (Stepik)
- Phillip Compeau (UCSD)



- **100+** autograded programming challenges
- **53,000** students from **120** countries
- Nearly **million** programs submitted by students
- **Used by 150+ instructors in offline classes**



Interactivity: Asking Questions to Detect and Prevent Educational Breakdowns

- Students face **Exercise Breaks**, **STOP and Think** questions, **Code Challenges**, and **Detours** at every page.
- Solutions to these challenges are immediately and automatically graded
- As soon as a student fails, he/she is directed to a module (**Charging Station**) addressing the detected learning breakdown



WHERE IN THE GENOME DOES DNA REPLICATION BEGIN?
Algorithmic Warmup



ARE THERE FRAGILE REGIONS IN THE HUMAN GENOME?
Combinatorial Algorithms



Traditional Classroom Lecture

Lecture “Finding Hidden Messages in DNA” (first 20 min)



Introduction



Hidden Messages in
the Replication
Origin



Some Hidden
Messages are More
Surprising than Others



An Explosion of
Hidden Messages



Homework (coding challenges)



Pattern
Count



Frequent
Words



Reverse
Complement



Pattern
Matching



Clump
Finding



Text Lesson



Coding Challenge



Towards MAIT

Massive Adaptive Interactive Text (MAIT) with embedded coding challenges



Introduction

Hidden Messages in
the Replication
Origin

Some Hidden
Messages are More
Surprising than Others

An Explosion of
Hidden Messages

Pattern
Count

Frequent
Words

Reverse
Complement
Matching

Clump
Finding



Text Lesson

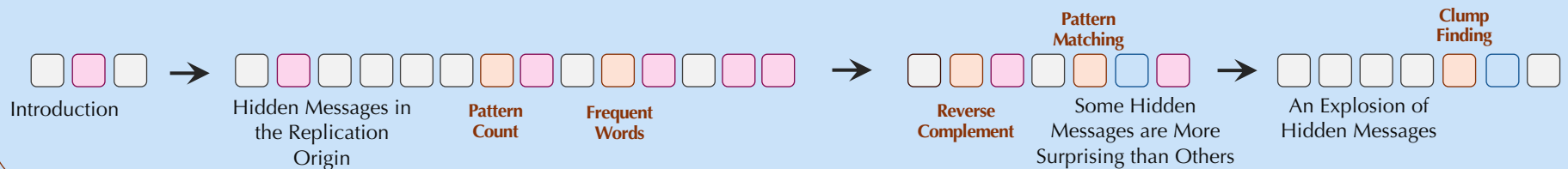


Coding Challenge



Towards MAIT

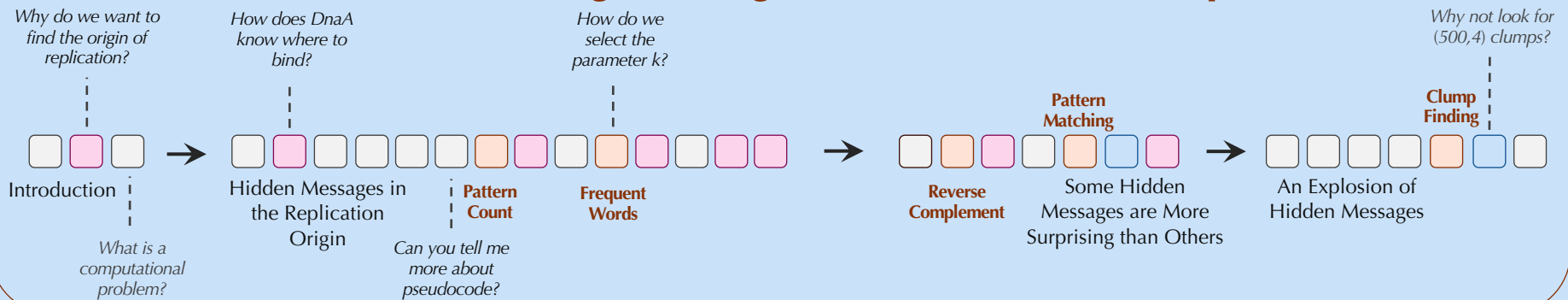
MAIT with embedded coding challenges/exercises/discussion questions



-  Text Lesson
-  **Discussion Question**
-  Exercise
-  Coding Challenge

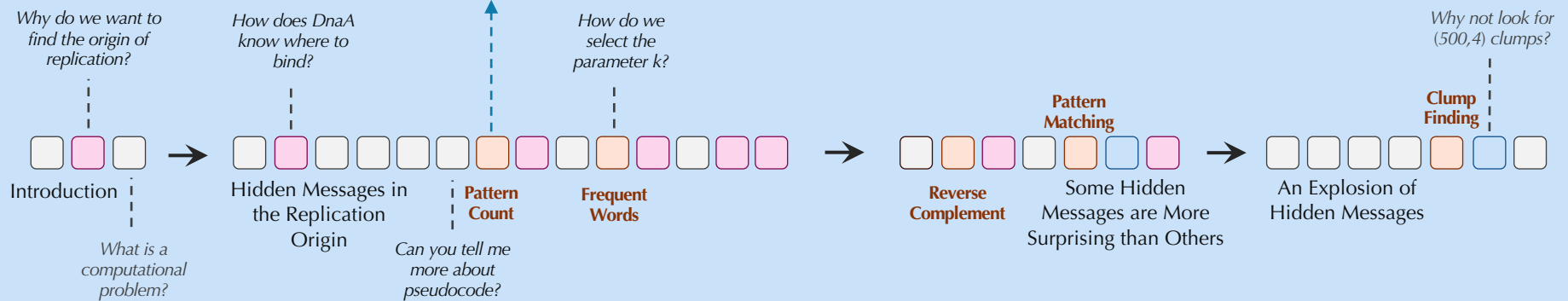
Towards MAIT

MAIT with embedded coding challenges/exercises/discussion questions/FAQs



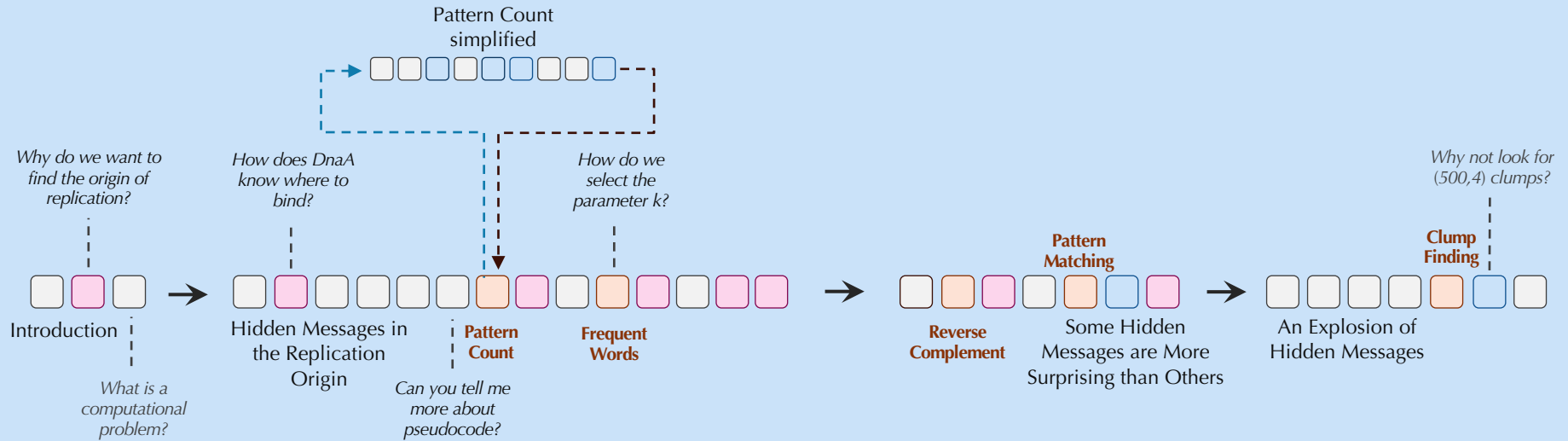
- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ



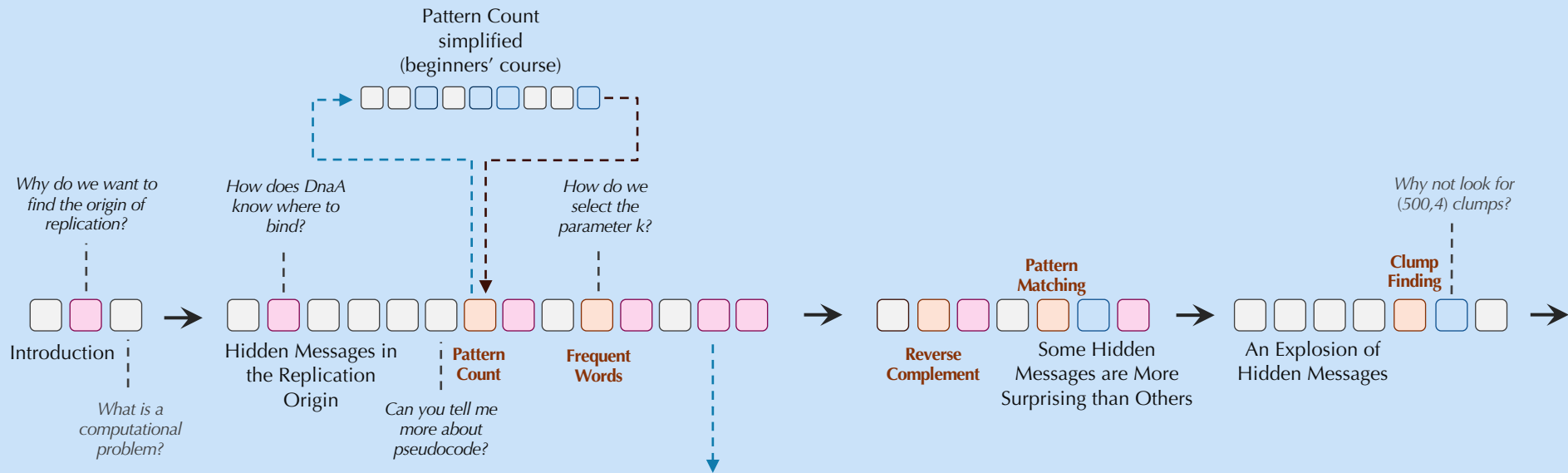


A MOOC for tens of thousands of students is a perfect way to form a **compendium of learning breakdowns** (8500 discussion posts for the first run of *Bioinformatics Algorithms* on Coursera).

Learning Breakdown Resolved!

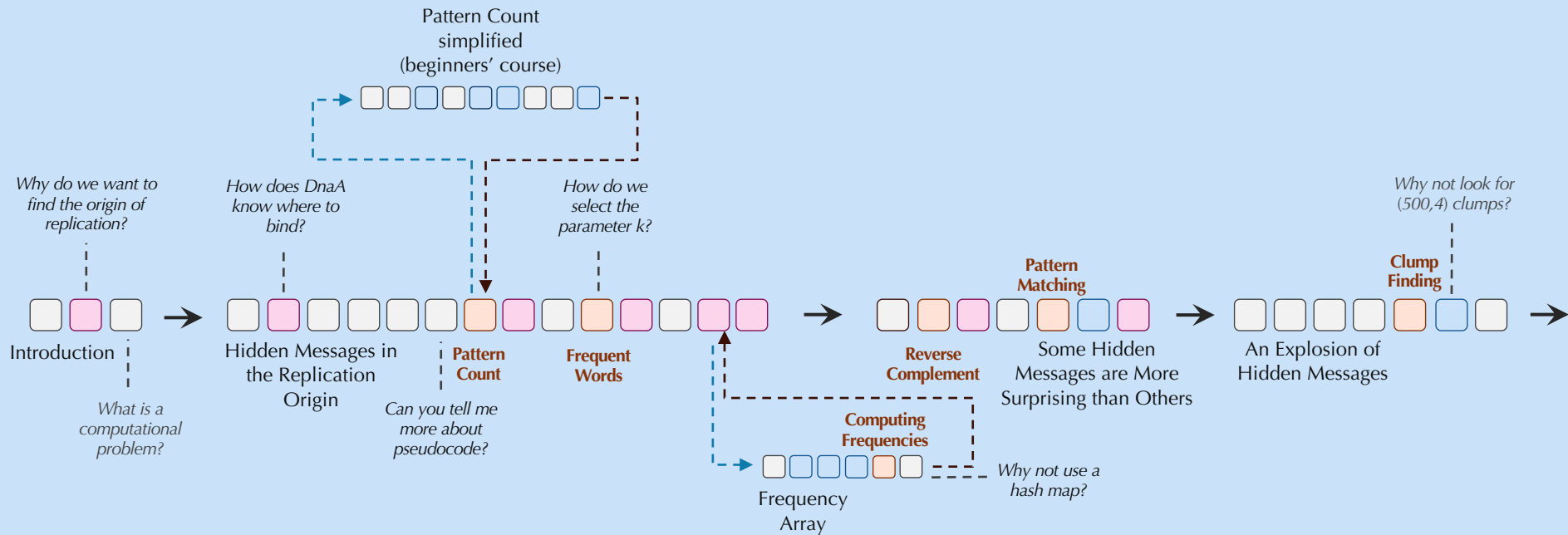


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- Learning Breakdown
- Resolved Breakdown**

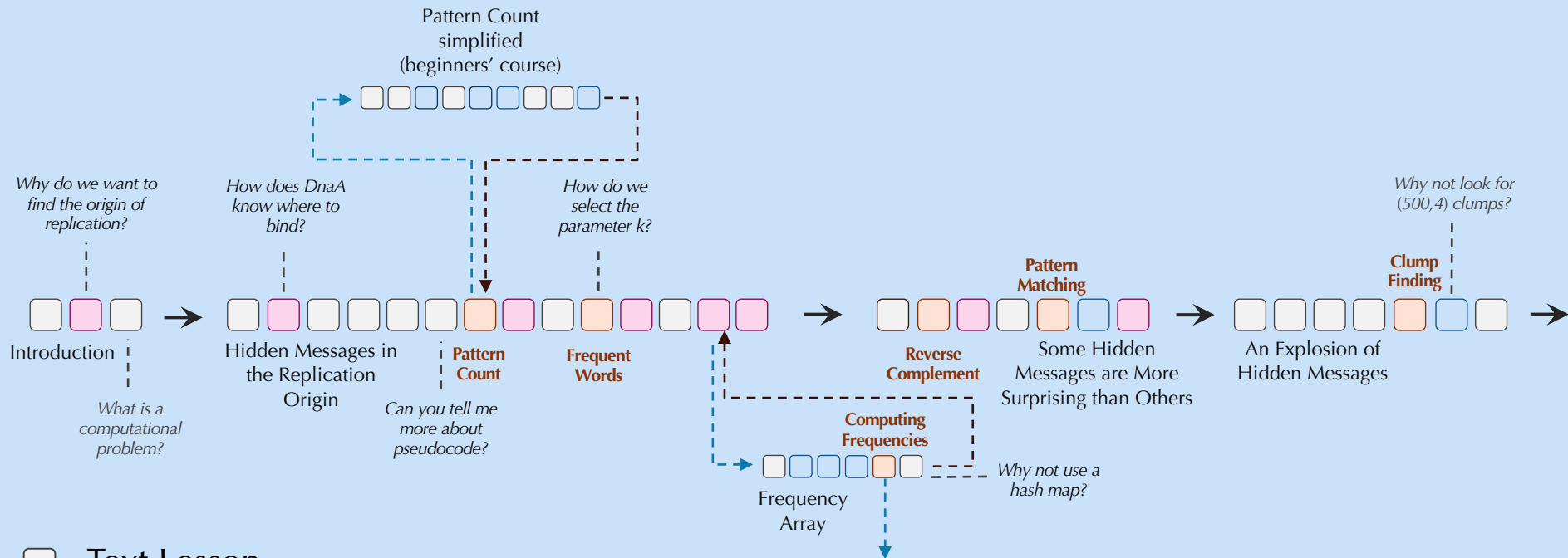


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- - -> Learning Breakdown
- - -> Resolved Breakdown



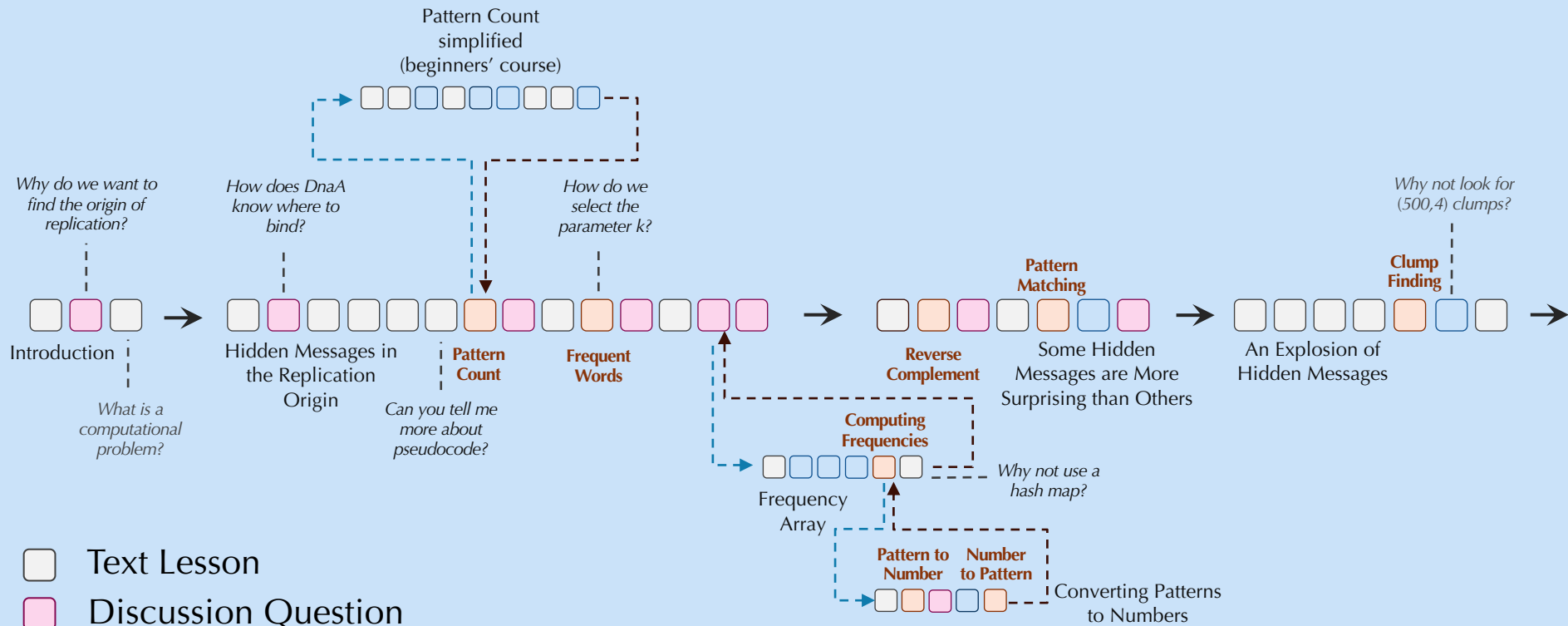


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- > Learning Breakdown
- > Resolved Breakdown

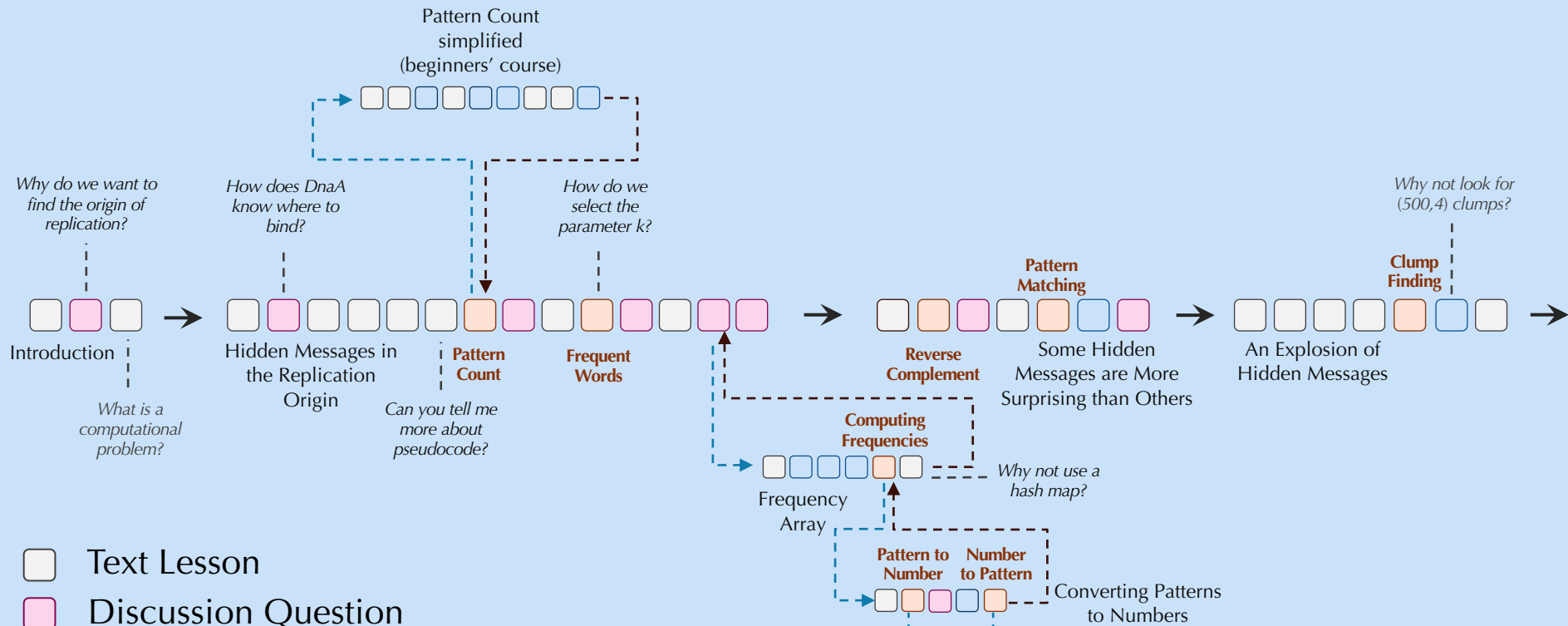


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- > Learning Breakdown
- > Resolved Breakdown



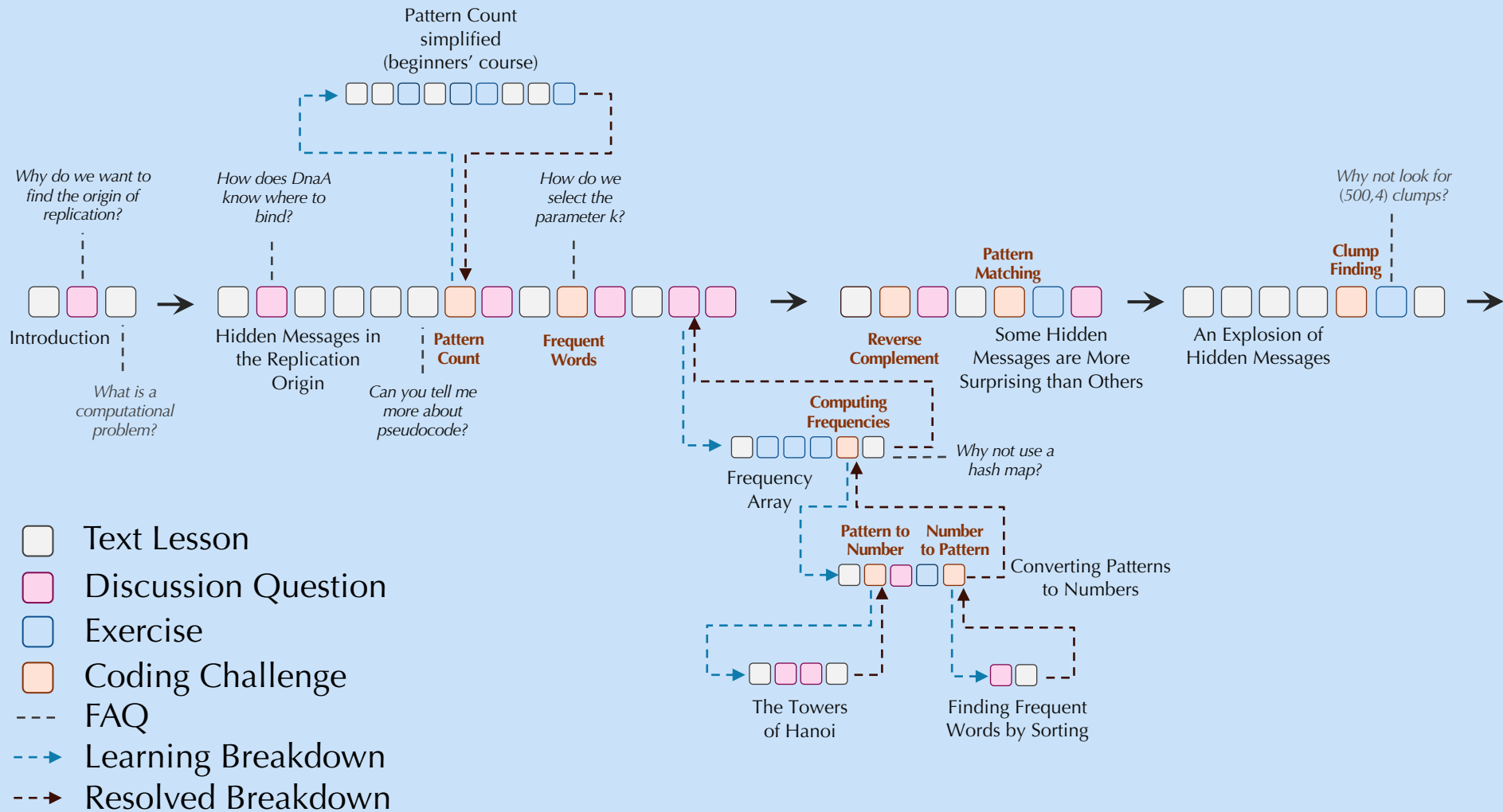


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- > Learning Breakdown
- > Resolved Breakdown

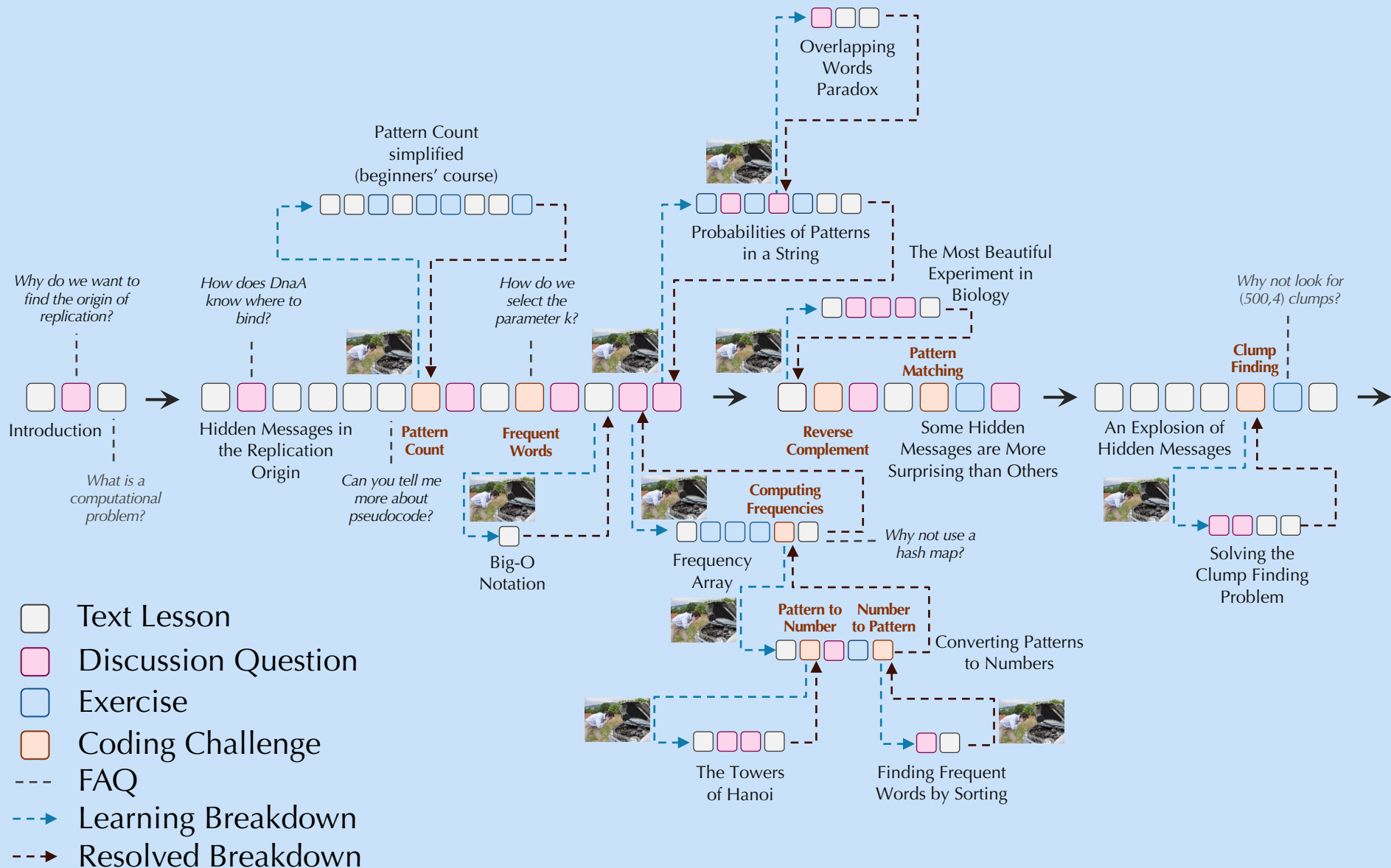


- Text Lesson
- Discussion Question
- Exercise
- Coding Challenge
- FAQ
- - -> Learning Breakdown
- - -> Resolved Breakdown

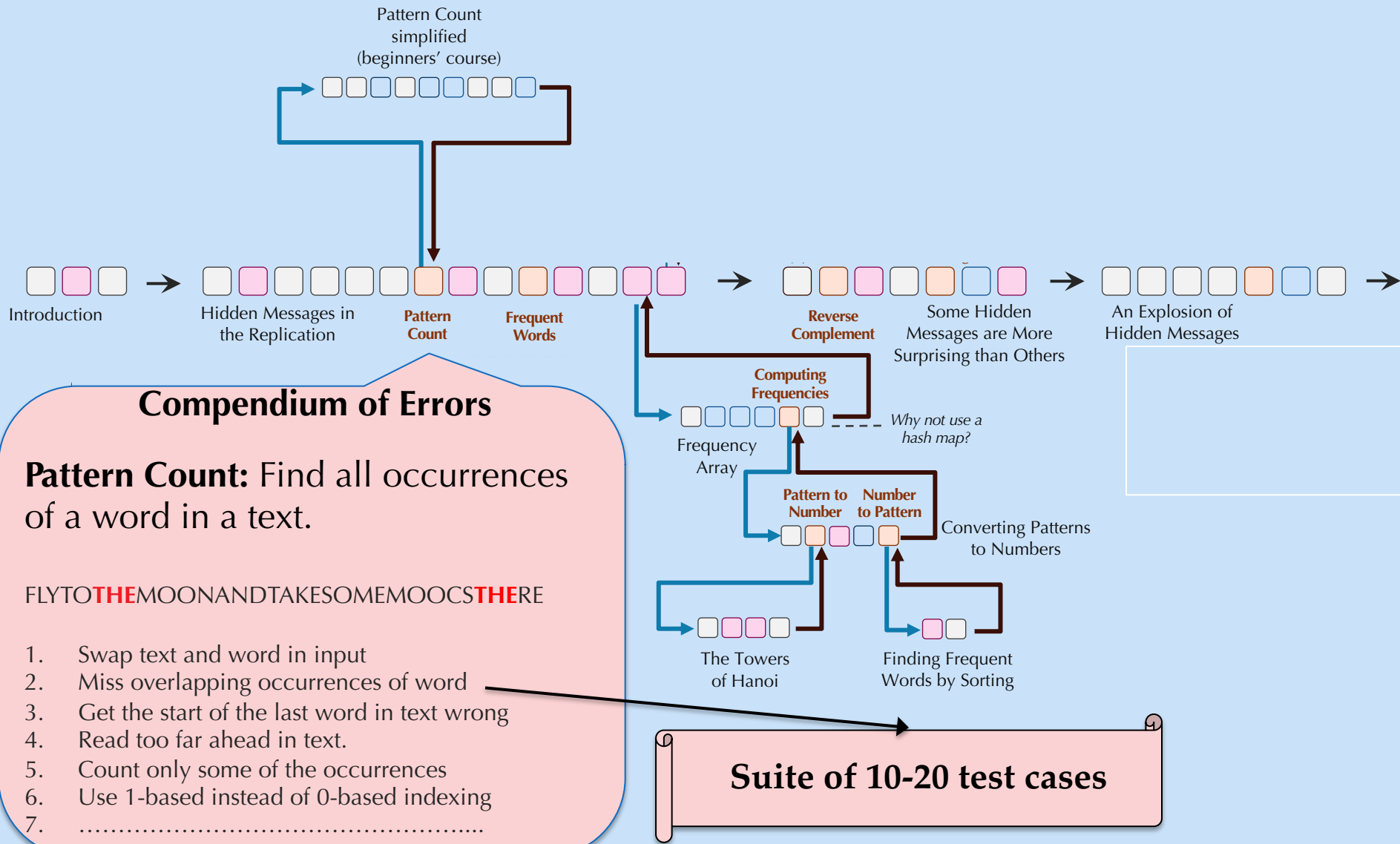




Ten Learning Breakdowns in 20 Minutes...

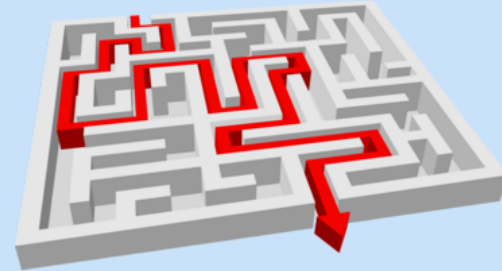


Individual Learning Path (3-5 hours)



Towards DIGITAL Science of Learning

- Compendium of Learning Breakdowns and Compendium of Errors are the initial steps towards mapping each learner to his/her unique learning path through a MAIT.
- Learning paths collected for 1000s of students will transform educational psychology into a Digital Science of Learning and will enable Machine Learning approaches to improving education.



Adaptive MOOC Introduction to Genomic Data Science

debuted at  2017

Modularity:

MAIT as a hub for a community of educators

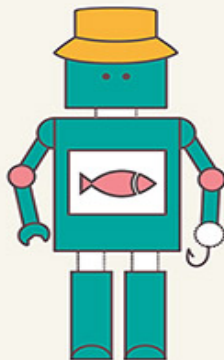
- MAITs should modularize content and supplement it by additional crowd-sourced modules.



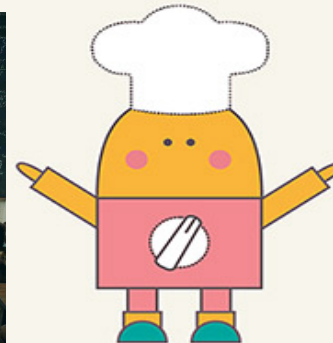
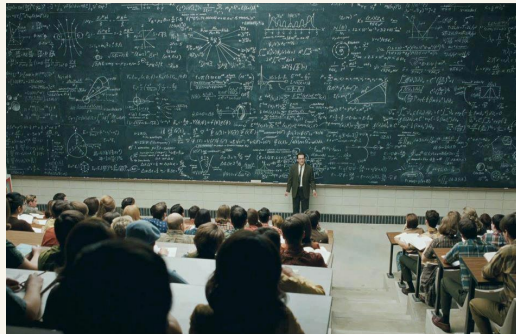
- Bioinformatics: Queen or Servant?
- Bioinformatics refutes biological dogmas
- **A 1000-year old dogma**
 - Classroom: an Optimal Learning Experience
 - What Is Wrong with MOOCs?
 - Massive Adaptive Interactive Text (MAIT)
 - **MAIT: a Solution of the Bloom Conundrum?**



Taxi Drivers



Fishermen



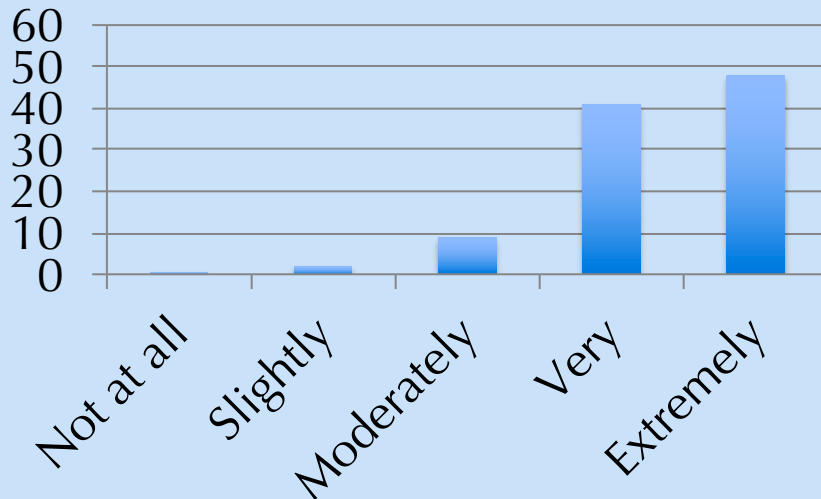
Bakers



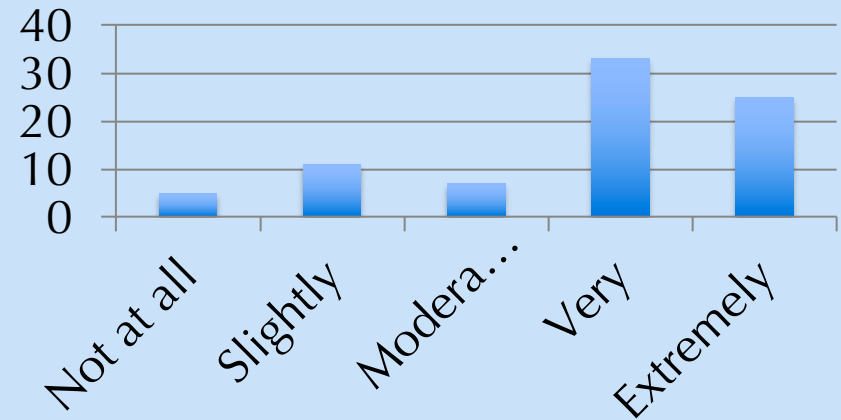
Fast Food Cooks

MAIT vs MOOC

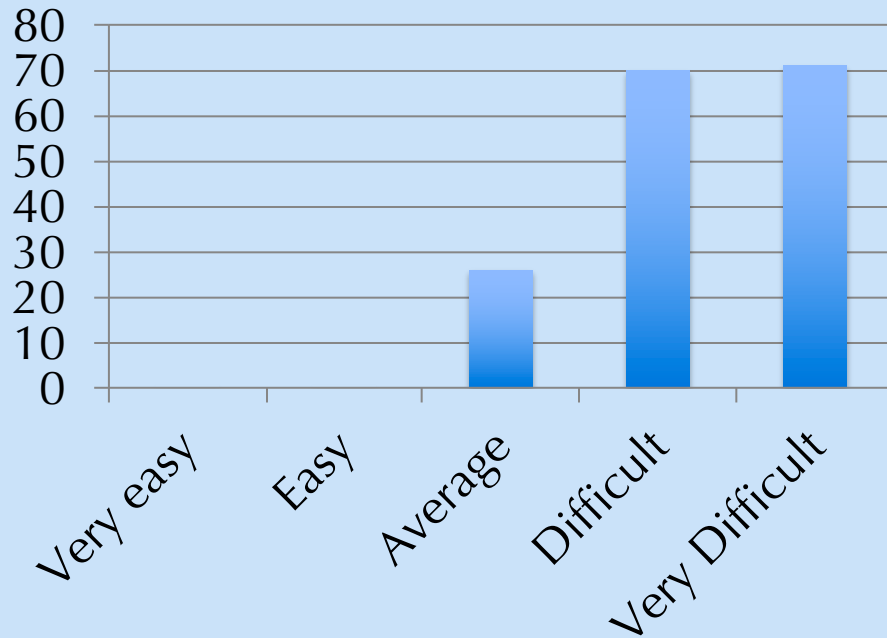
How valuable is the **interactive text** in helping you learn?



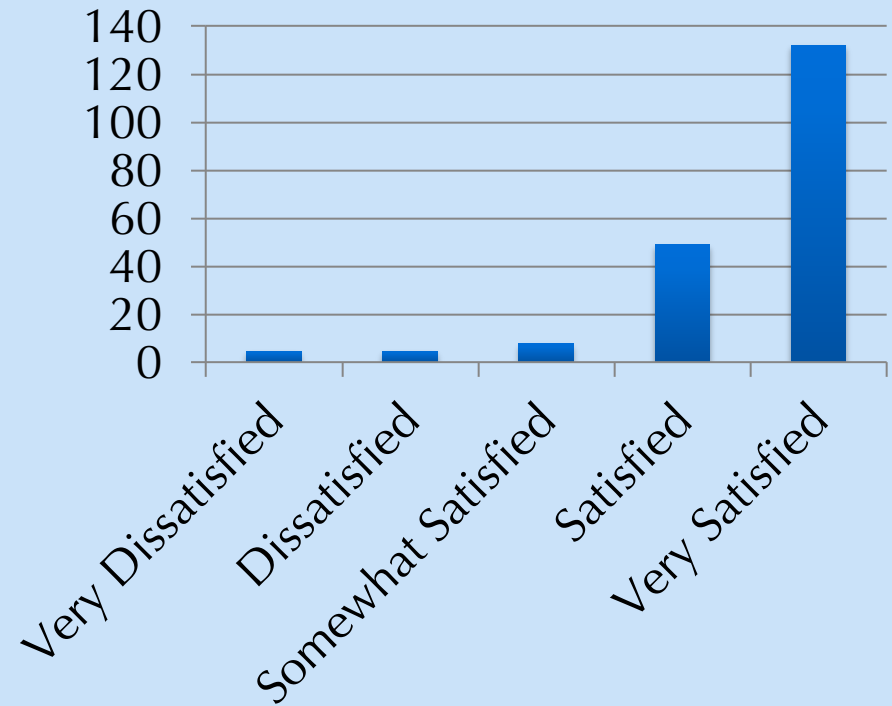
How valuable are the **lecture videos** in helping you learn?



Relative Difficulty



Overall Satisfaction



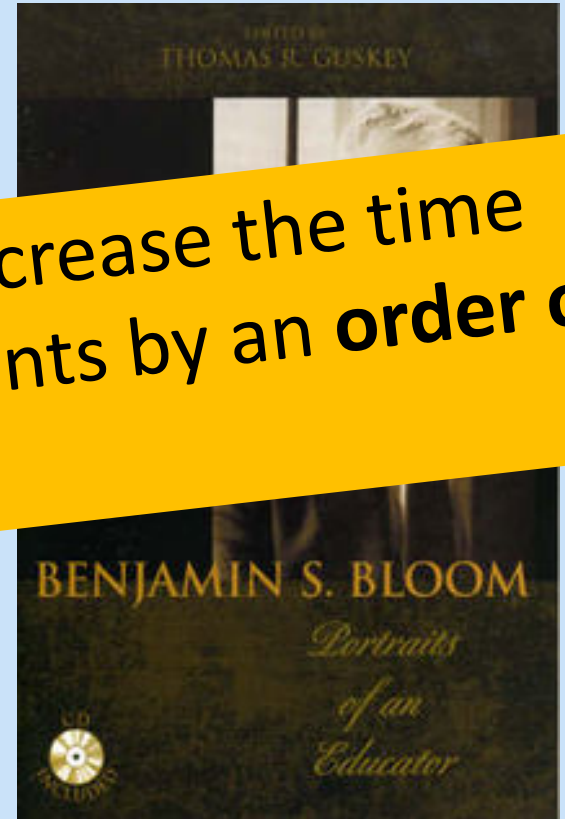
12 hours work a week on average!

MAITs and the Bloom Conundrum

- A MAIT does not just *facilitates* a professor's transition toward a flipped classroom; it *necessitates* this transition.

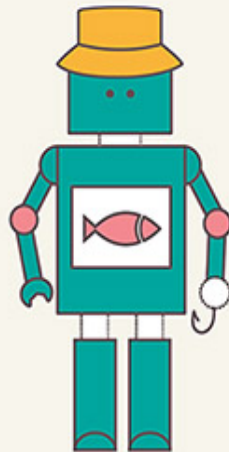
A MAIT allows professors/TAs to increase the time for real communication with students by an **order of magnitude**

testing so that they can interact with students.

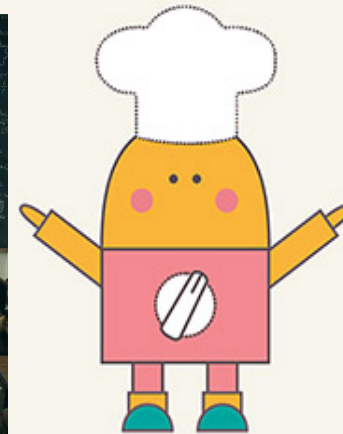
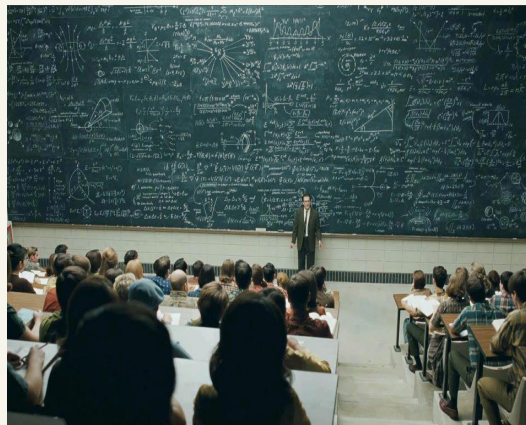




Taxi Drivers



Fishermen



Bakers



Fast Food Cooks

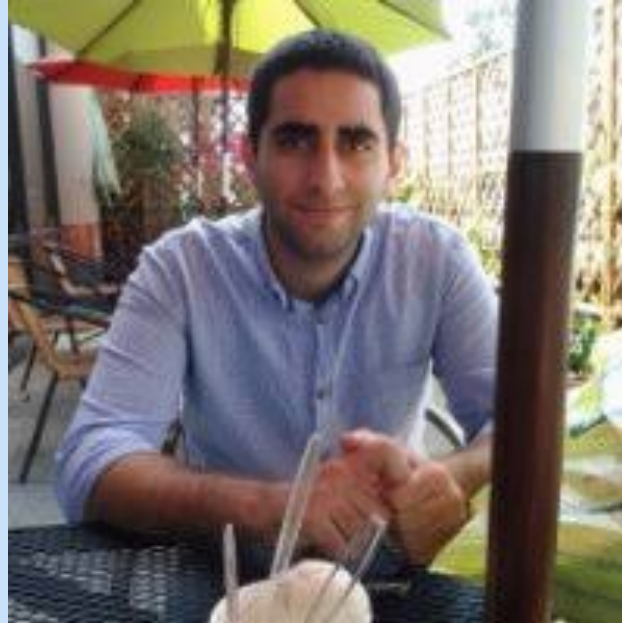


Acknowledgements



Phillip Compeau
Carnegie Mellon University

Acknowledgements



Niema Moshri
UCSD

Analyze Your Genome!

debuted at  in 2018