Introduction to Bioinformatics

Winter 2024

- **Lecturer: Dr. Dvir Aran**

- **Teaching Assistance:**

  Almog Angel

  Ziv Cohen

- Course web site: http://webcourse.cs.technion.ac.il/236523

גל איזנקוט ז״ל

סטודנט למדעי המחשב
ומדעי החיים, עם התמחות
בביו-אינפורמטיקה
אוניברסיטת תל-אביב

Introduction – Dr. Dvir Aran

THE HEBREW UNIVERSITY OF JERUSALEM

Computer Science
Computational Biology
Mathematics
Faculty of Medicine

Asaf Hellman
Epigenetics

Shai Shalev-Shwartz
Machine learning

Anthem | ai

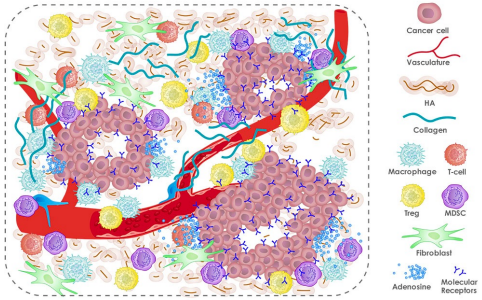
Anthem ISRAEL

Data scientist
Clinical Informatics
Digital Medicine

STANFORD UNIVERSITY

UCSF

Atul Butte
Translational Bioinformatics
Clinical Informatics



Biomedical Data Science
Aran Lab @ Technion

Course Information

- Lectures

  Wednesday, 10:30 – 12:30
- Recitations: Almog Angel
  - Wednesday, 14:30-15:30
  - Thursday, 13:30-14:30
- Course Website

  https://webcourse.cs.technion.ac.il/236523/

Course Structure and Requirements

- Class Structure
  - 2 hours lecture
  - 1 hour tutorial
- Home-work
  - Homework assignments will be given every second week.
  - The homework will be done in pairs.
  - 4/4 homework assignments must be submitted.
  - A final project will be conducted in pairs.
- Final projects will be analyzing genomic data and presenting your research and results.

Course Regulations

- Staff
  - Dr. Dvir Aran
  - TA: Almog Angel
  - HW: Ziv Cohen
- Reception hour
  - Wednesday, 12:30
  - By appointment
- Please avoid emails regarding the course not through represantatives.

Grading

- 40% homework assignments (10 points each)
- 60% final project

Course Objectives

- Learn underlying ideas of common algorithms in bioinformatics.
- Learn to translate a biological problem into a computational problem.
- Learn to read scientific papers, propose and conduct independent research.

As part of the course we will introduce the R programing language and learn to analyze DNA and RNA sequencing data and conduct basic statistical and machine-learning analyses.

Class structure

- Lectures:
  1. Problem in biology.
  2. Computer science concept.
  3. Applying the computer science concept to solve the biology problem.
- Recitations:
  - Hands-on analysis of biological data
  - R, statistics, algorithms.

Course subjects

1. Comparative genomics (2 classes)
2. Gene expression (3 classes)
3. Genetics (2 class)
4. Frontiers in bioinformatics:
   - Cancer genomics
   - Single-cell genomics
   - Genome engineering

Dynamic programming, linear algorithms hash functions, clustering algorithms, differential expression analysis, deconvolution, burrows-wheeler transformation, machine learning algorithms, energy functions optimization, deep learning, statistical hypothesis testing, dimensionality reduction, population structure algorithms, genome wide association studies, polygenic risk scores and more…

Aligning expectations

Evolution | Sequencing technologies | CRISPR | immunology | Cancer | Research

Statistics | Machine learning | Genetic diseases | Gene expression | Algorithms | Molecular biology

What is Computational Biology/Bioinformatics?

What is Computational Biology/Bioinformatics?

Computational biology and **bioinformatics** is an interdisciplinary field that develops and applies **computational methods** to analyze large collections of biological data, such as genetic sequences, cell populations or protein samples, to make new predictions or **discover new biology**.

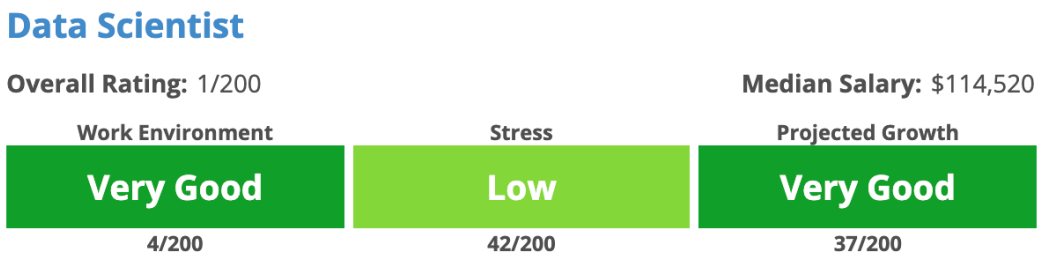
https://www.nature.com/subjects/computational-biology-and-bioinformatics

Why Study Computational Biology?

Interdisciplinary

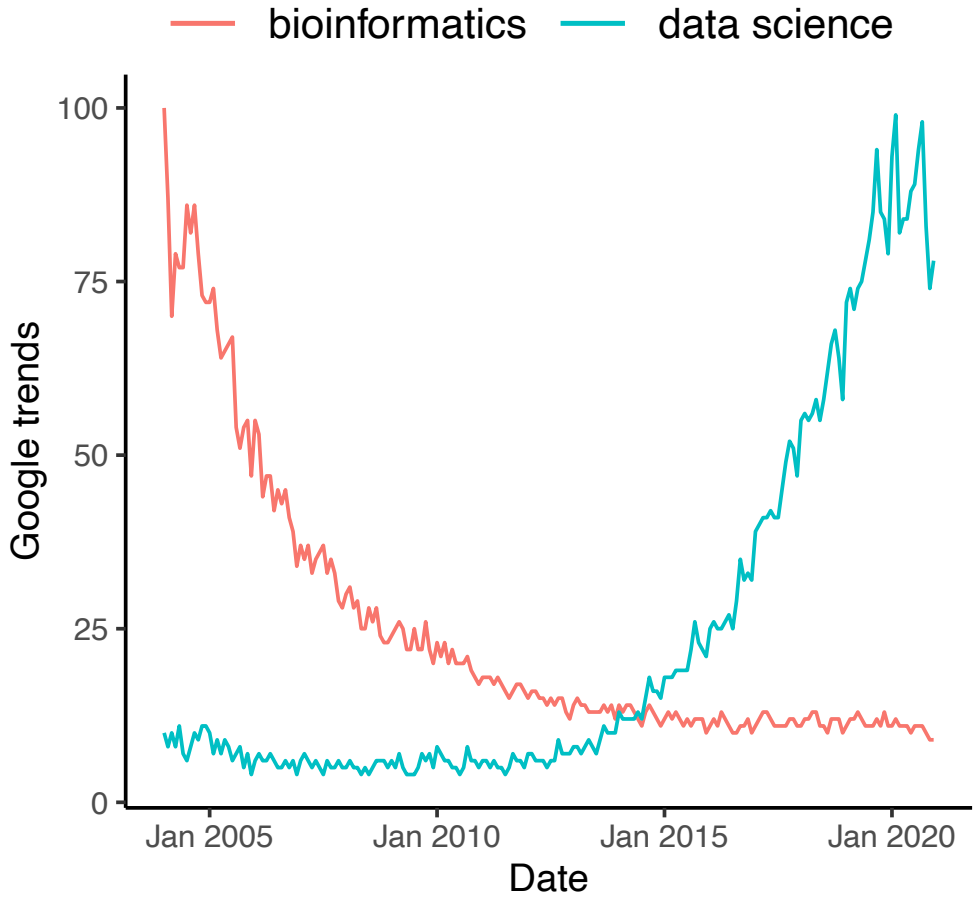
- Biology
- Computer Science
- Data Science
- Mathematics
- Statistics

Best Jobs	Worst Jobs
1. Data scientist	224. Taxi Driver
2. Statistician	223. Logging Worker
3. University Professor	222. Newspaper Reporter
4. Occupational Therapist	221. Retail Salesperson
5. Genetic Counselor	220. Enlisted Military Personnel
6. Medical Services Manager	219. Correctional Officer
7. Information Security Analyst	218. Disc Jockey
8. Mathematician	217. Nuclear Decontamination Technician
9. Operations Research Analyst	216. Advertising Salesperson
10. Actuary	215. Broadcaster
11. Software Developer	

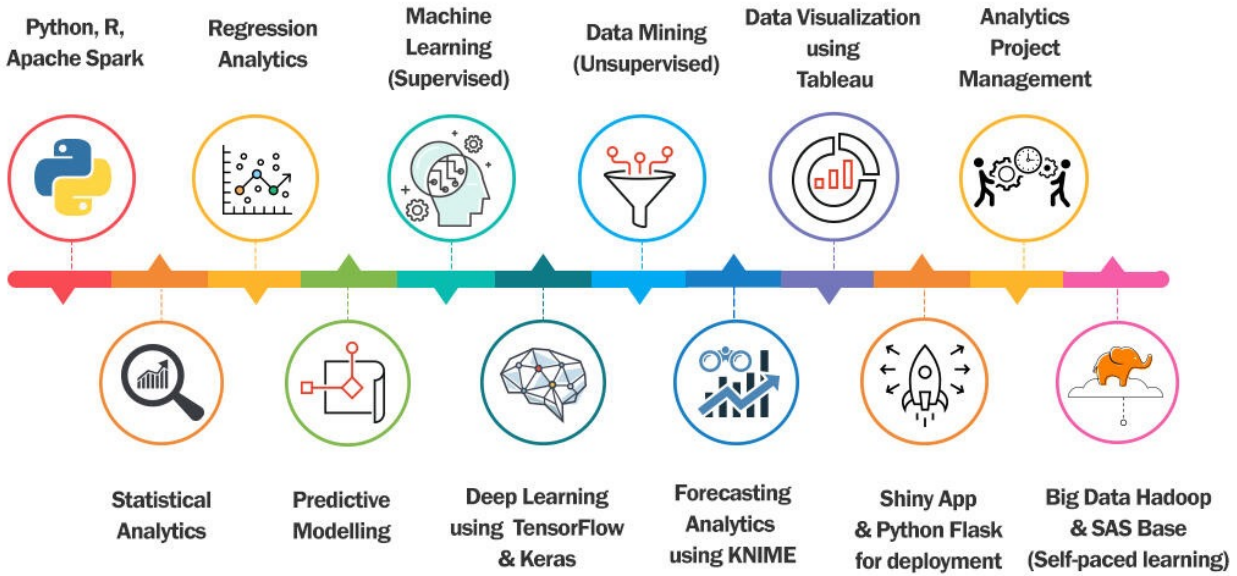


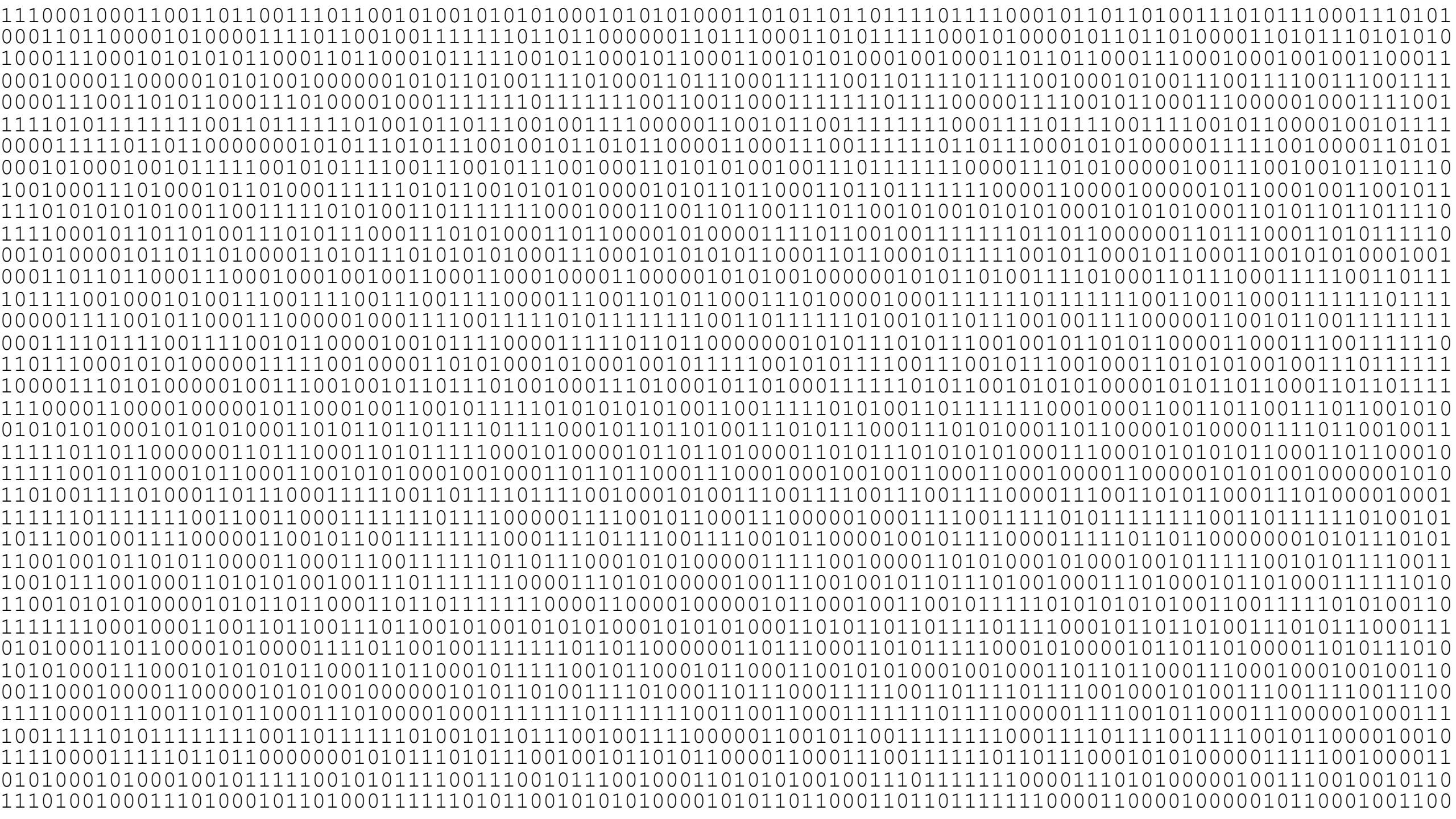
https://www.careercast.com/jobs-rated/2019-jobs-rated-report

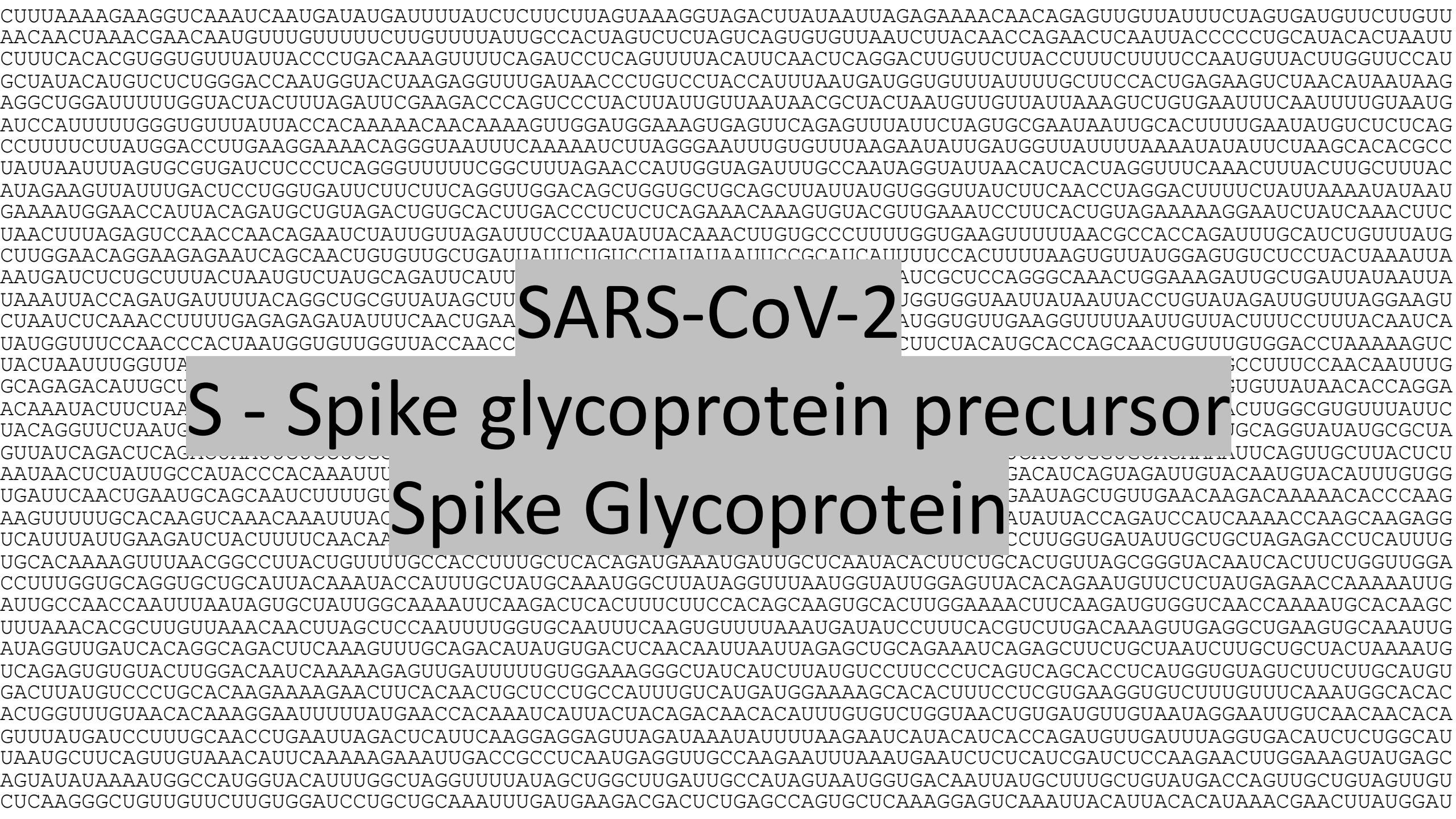
Bioinformatics or Biomedical Data Science?

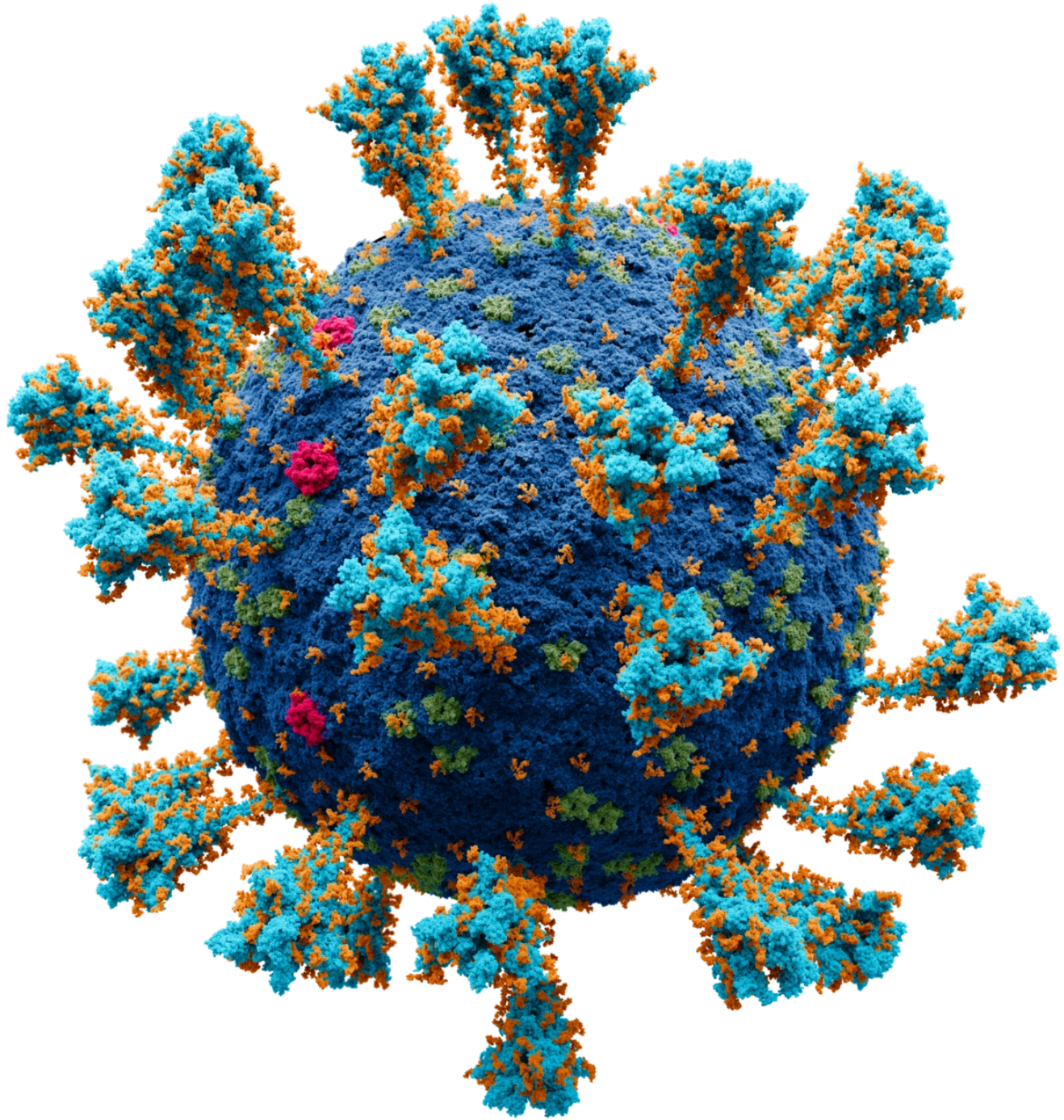


- Bioinformatics
- Translational Bioinformatics
- Clinical Research Informatics
- Imaging Informatics
- Medical informatics
- Clinical Informatics
- Health informatics
- Public Health Informatics



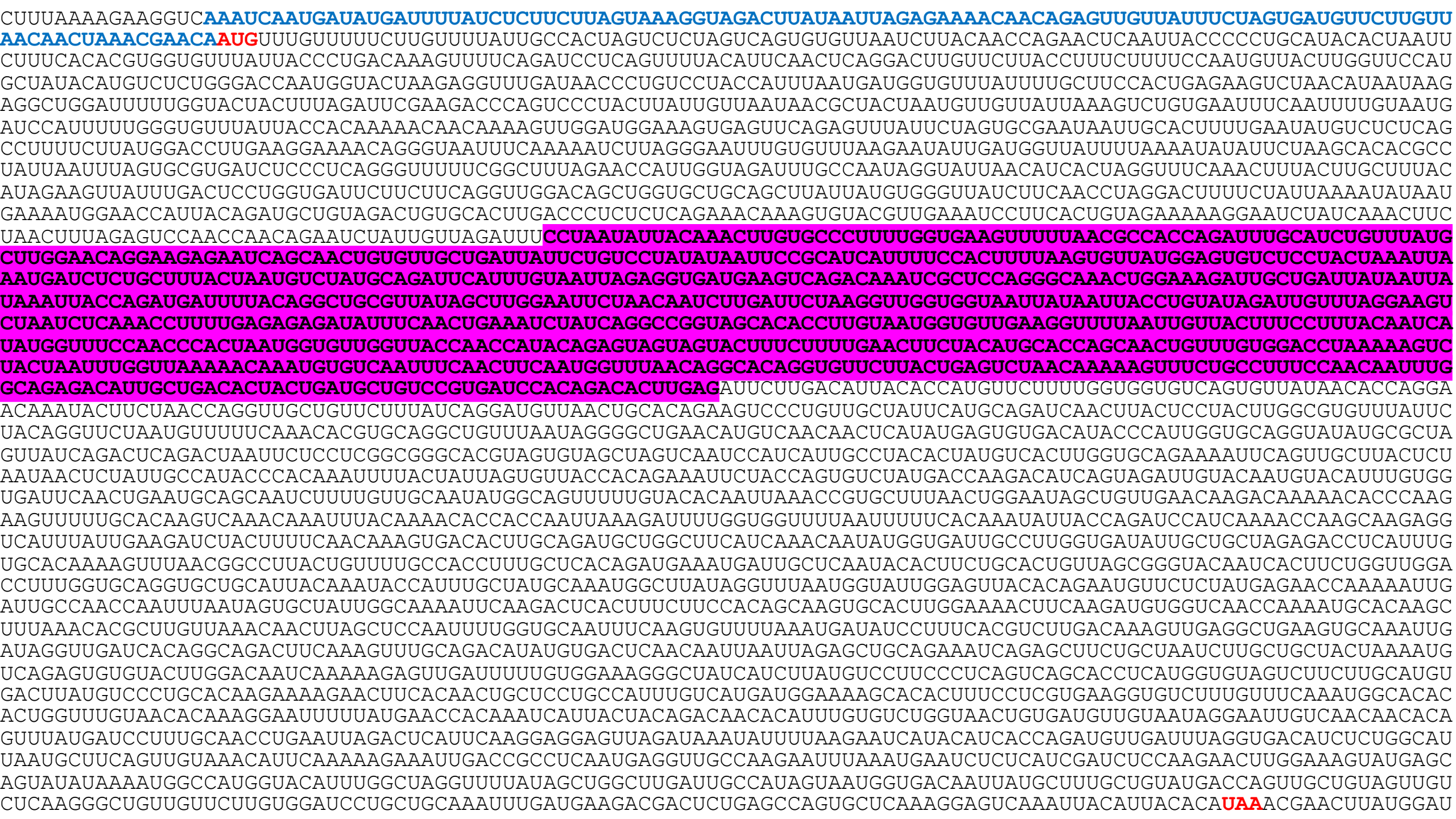



SARS-CoV-2
S - Spike glycoprotein precursor
Spike Glycoprotein

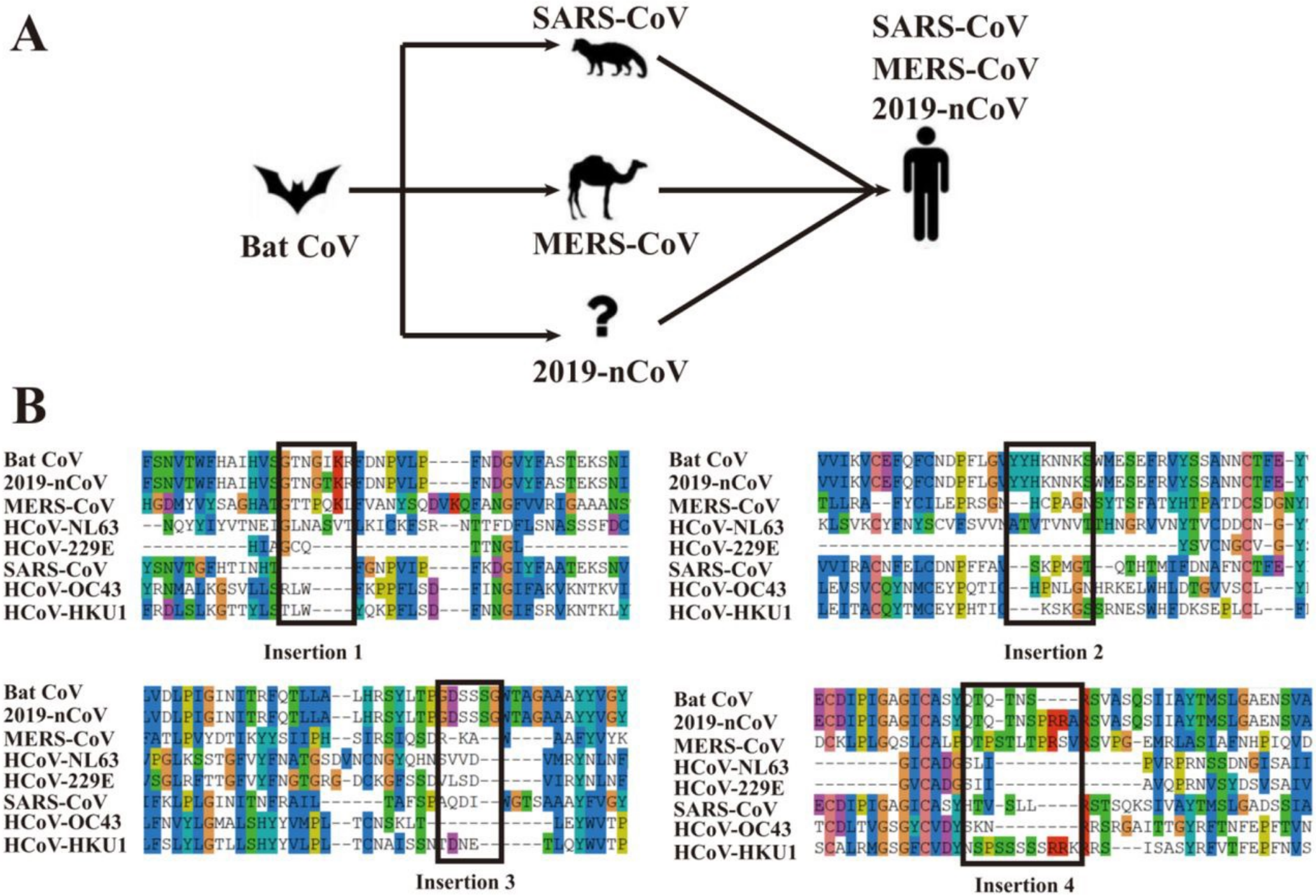


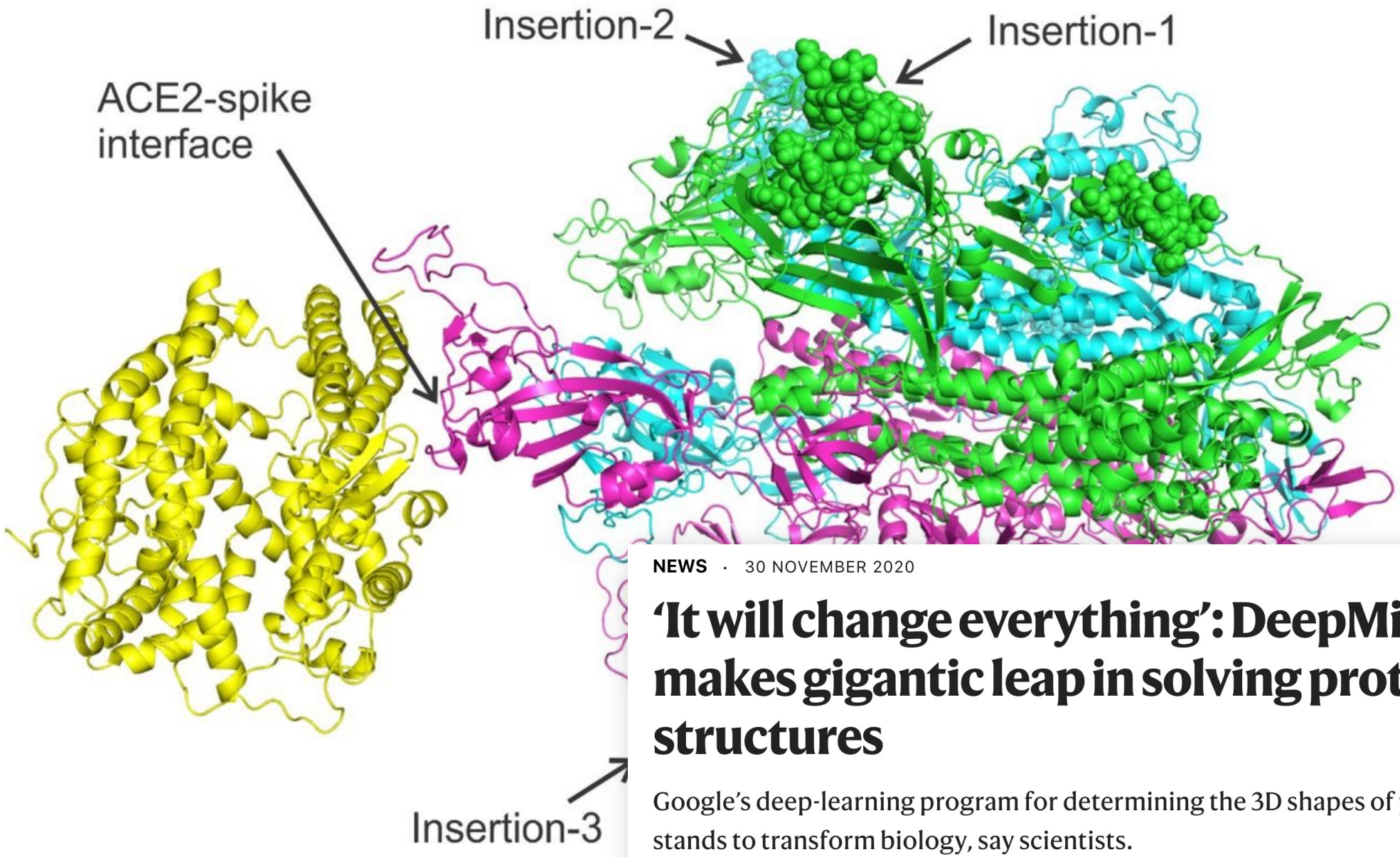
CUUUAAAAGAAGGUCAAAUCAAUGAUAUGAUUUUAUCUCUUCUUAGUAAAGGUAGACUUAUAAUUAGAGAAAACAACAGAGUUGUUAUUUCUAGUGAUGUUCUUGUU
AACAACUAAACGAACA**AUG**UUUGUUUUUCUUGUUUUAUUGCCACUAGUCUCUAGUCAGUGUGUUAAUCUUACAACCAGAACUCAAUUACCCCCUGCAUACACUAAUU
CUUUCACACGUGGUGUUUAUUACCCUGACAAAGUUUUCAGAUCCUCAGUUUUACAUUCAACUCAGGACUUGUUCUUACCUUUCUUUUCCAAUGUUACUUGGUUCCAU
GCUAUACAUGUCUCUGGGACCAAUGGUACUAAGAGGUUUGAUAACCCUGUCCUACCAUUUAAUGAUGGUGUUUAUUUUGCUUCCACUGAGAAGUCUAACAUAAUAAG
AGGCUGGAUUUUUGGUACUACUUUAGAUUCGAAGACCCAGUCCCUACUUAUUGUUAAUAACGCUACUAAUGUUGUUAUUAAAGUCUGUGAAUUUCAAUUUUGUAAUG
AUCCAUUUUUGGGUGUUUAUUACCACAAAAACAACAAAAGUUGGAUGGAAAGUGAGUUCAGAGUUUAUUCUAGUGCGAAUAAUUGCACUUUUGAAUAUGUCUCUCAG
CCUUUUCUUAUGGACCUUGAAGGAAAACAGGGUAAUUUCAAAAAUCUUAGGGAAUUUGUGUUUAAGAAUAUUGAUGGUUAUUUUAAAAUAUAUUCUAAGCACACGCC
UAUUAAUUUAGUGCGUGAUCUCCCCUCAGGGUUUUUCGGCUUUAGAACCAUUGGUAGAUUUGCCAAUAGGUAUUAACAUCACUAGGUUUCAAACUUUACUUGCUUUAC
AUAGAAGUUAUUUGACUCCUGGUGAUUCUUCUUCAGGUUGGACAGCUGGUGCUGCAGCUUAUUAUGUGGGUUAUCUUCAACCUAGGACUUUUCUAUUAAAAUAUAAU
GAAAAUGGAACCAUUACAGAUGCUGUAGACUGUGCACUUGACCCUCUCUCAGAAACAAAGUGUACGUUGAAAUCCUUCACUGUAGAAAAAGGAAUCUAUCAAACUUC
UAACUUUAGAGUCCAACCAACAGAAUCUAUUGUUAGAUUUCCUAAUAUUACAAACUUGUGCCCUUUUGGUGAAGUUUUUAACGCCACCAGAUUUGCAUCUGUUUAUG
CUUGGAACAGGAAGAGAAUCAGCAACUGUGUUGCUGAUUAUUCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAGUGUUAUGGAGUGUCUCCUACUAAAUUA
AAUGAUCUCUGCUUUACUAAUGUCUAUGCAGAUUCAUUUGUAAUUAGAGGUGAUGAAGUCAGACAAAUCGCUCCAGGGCAAACUGGAAAGAUUGCUGAUUAUAAUUA
UAAAUUACCAGAUGAUUUUACAGGCUGCGUUAUAGCUUGGAAUUCUAACAAUCUUGAUUCUAAGGUUGGUGGUAAUUAUAAUUACCUGUAUAGAUUGUUUAGGAAGU
CUAAUCUCAAACCUUUUGAGAGAGAUAUUUCAACUGAAAUCUAUCAGGCCGGUAGCACACCUUGUAAUGGUGUUGAAGGUUUUAAUUGUUACUUUCCUUUACAAUCA
UAUGGUUUCCAACCCACUAAUGGUGUUGGUUACCAACCAUACAGAGUAGUAGUACUUUCUUUUGAACUUCUACAUGCACCAGCAACUGUUUGUGGACCUAAAAAGUC
UACUAAUUUGGUUAAAAACAAAUGUGUCAAUUUCAACUUCAAUGGUUUAACAGGCACAGGUGUUCUUACUGAGUCUAACAAAAAGUUUCUGCCUUUCCAACAAUUUG
GCAGAGACAUUGCUGACACUACUGAUGCUGUCCGUGAUCCACAGACACUUGAGAUUCUUGACAUUACACCAUGUUCUUUUGGUGGUGUCAGUGUUAUAACACCAGGA
ACAAAUACUUCUAACCAGGUUGCUGUUCUUUAUCAGGAUGUUAACUGCACAGAAGUCCCUGUUGCUAUUCAUGCAGAUCAACUUACUCCUACUUGGCGUGUUUAUUC
UACAGGUUCUAAUGUUUUUCAAACACGUGCAGGCUGUUUAAUAGGGGCUGAACAUGUCAACAACUCAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUA
GUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGCUAGUCAAUCCAUCAUUGCCUACACUAUGUCACUUGGUGCAGAAAAUUCAGUUGCUUACUCU
AAUAACUCUAUUGCCAUACCCACAAAUUUUACUAUUAGUGUUACCACAGAAAUUCUACCAGUGUCUAUGACCAAGACAUCAGUAGAUUGUACAAUGUACAUUUGUGG
UGAUUCAACUGAAUGCAGCAAUCUUUUGUUGCAAUAUGGCAGUUUUUGUACACAAUUAAACCGUGCUUUAACUGGAAUAGCUGUUGAACAAGACAAAAACACCCAAG
AAGUUUUUGCACAAGUCAAACAAAUUUACAAAACACCACCAAUUAAAGAUUUUGGUGGUUUUAAUUUUUCACAAAUAUUACCAGAUCCAUCAAAACCAAGCAAGAGG
UCAUUUAUUGAAGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGGCUUCAUCAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCUGCUAGAGACCUCAUUUG
UGCACAAAAGUUUAACGGCCUUACUGUUUUGCCACCUUUGCUCACAGAUGAAAUGAUUGCUCAAUACACUUCUGCACUGUUAGCGGGUACAAUCACUUCUGGUUGGA
CCUUUGGUGCAGGUGCUGCAUUACAAAUACCAUUUGCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAGUUACACAGAAUGUUCUCUAUGAGAACCAAAAAUUG
AUUGCCAACCAAUUUAAUAGUGCUAUUGGCAAAAUUCAAGACUCACUUUCUUCCACAGCAAGUGCACUUGGAAAACUUCAAGAUGUGGUCAACCAAAAUGCACAAGC
UUUAAACACGCUUGUUAAACAACUUAGCUCCAAUUUUGGUGCAAUUUCAAGUGUUUUAAAUGAUAUCCUUUCACGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUG
AUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAGAGCUUCUGCUAAUCUUGCUGCUACUAAAAUG
UCAGAGUGUGUACUUGGACAAUCAAAAAGAGUUGAUUUUUGUGGAAAGGGCUAUCAUCUUAUGUCCUUCCCUCAGUCAGCACCUCAUGGUGUAGUCUUCUUGCAUGU
GACUUAUGUCCCUGCACAAGAAAAGAACUUCACAACUGCUCCUGCCAUUUGUCAUGAUGGAAAAGCACACUUUCCUCGUGAAGGUGUCUUUGUUUCAAAUGGCACAC
ACUGGUUUGUAACACAAAGGAAUUUUUAUGAACCACAAAUCAUUACUACAGACAACACAUUUGUGUCUGGUAACUGUGAUGUUGUAAUAGGAAUUGUCAACAACACA
GUUUAUGAUCCUUUGCAACCUGAAUUAGACUCAUUCAAGGAGGAGUUAGAUAAAUAUUUUAAGAAUCAUACAUCACCAGAUGUUGAUUUAGGUGACAUCUCUGGCAU
UAAUGCUUCAGUUGUAAACAUUCAAAAAGAAAUUGACCGCCUCAAUGAGGUUGCCAAGAAUUUAAAUGAAUCUCUCAUCGAUCUCCAAGAACUUGGAAAGUAUGAGC
AGUAUAUAAAAUGGCCAUGGUACAUUUGGCUAGGUUUUAUAGCUGGCUUGAUUGCCAUAGUAAUGGUGACAAUUAUGCUUUGCUGUAUGACCAGUUGCUGUAGUUGU
CUCAAGGGCUGUUGUUCUUGUGGAUCCUGCUGCAAAUUUGAUGAAGACGACUCUGAGCCAGUGCUCAAAGGAGUCAAAUUACAUUACACAUAAACGAACUUAUGGAU

CUUUAAAAGAAGGUCAAAUCAAUGAUAUGAUUUUAUCUCUUCUUAGUAAAGGUAGACUUAUAAUUAGAGAAAACAACAGAGUUGUUAUUUCUAGUGAUGUUCUUGUU
AACAACUAAACGAACAAUGUUUGUUUUUCUUGUUUUAUUGCCACUAGUCUCUAGUCAGUGUGUUAAUCUUACAACCAGAACUCAAUUACCCCCUGCAUACACUAAUU
CUUUCACACGUGGUGUUUAUUACCCUGACAAAGUUUUCAGAUCCUCAGUUUUACAUUCAACUCAGGACUUGUUCUUACCUUUCUUUUCCAAUGUUACUUGGUUCCAU
GCUAUACAUGUCUCUGGGACCAAUGGUACUAAGAGGUUUGAUAACCCUGUCCUACCAUUUAAUGAUGGUGUUUAUUUUGCUUCCACUGAGAAGUCUAACAUAAUAAG
AGGCUGGAUUUUUGGUACUACUUUAGAUUCGAAGACCCAGUCCCUACUUAUUGUUAAUAACGCUACUAAUGUUGUUAUUAAAGUCUGUGAAUUUCAAUUUUGUAAUG
AUCCAUUUUUGGGUGUUUAUUACCACAAAAACAACAAAAGUUGGAUGGAAAGUGAGUUCAGAGUUUAUUCUAGUGCGAAUAAUUGCACUUUUGAAUAUGUCUCUCAG
CCUUUUCUUAUGGACCUUGAAGGAAAACAGGGUAAUUUCAAAAAUCUUAGGGAAUUUGUGUUUAAGAAUAUUGAUGGUUAUUUUAAAAUAUAUUCUAAGCACACGCC
UAUUAAUUUAGUGCGUGAUCUCCCCUCAGGGUUUUUCGGCUUUAGAACCAUUGGUAGAUUUGCCAAUAGGUAUUAACAUCACUAGGUUUCAAACUUUACUUGCUUUAC
AUAGAAGUUAUUUGACUCCUGGUGAUUCUUCUUCAGGUUGGACAGCUGGUGCUGCAGCUUAUUAUGUGGGUUAUCUUCAACCUAGGACUUUUCUAUUAAAAUAUAAU
GAAAAUGGAACCAUUACAGAUGCUGUAGACUGUGCACUUGACCCUCUCUCAGAAACAAAGUGUACGUUGAAAUCCUUCACUGUAGAAAAAGGAAUCUAUCAAACUUC
UAACUUUAGAGUCCAACCAACAGAAUCUAUUGUUAGAUUUCCUAAUAUUACAAACUUGUGCCCUUUUGGUGAAGUUUUUAACGCCACCAGAUUUGCAUCUGUUUAUG
CUUGGAACAGGAAGAGAAUCAGCAACUGUGUUGCUGAUUAUUCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAGUGUUAUGGAGUGUCUCCUACUAAAUUA
AAUGAUCUCUGCUUUACUAAUGUCUAUGCAGAUUCAUUUGUAAUUAGAGGUGAUGAAGUCAGACAAAUCGCUCCAGGGCAAACUGGAAAGAUUGCUGAUUAUAAUUA
UAAAUUACCAGAUGAUUUUACAGGCUGCGUUAUAGCUUGGAAUUCUAACAAUCUUGAUUCUAAGGUUGGUGGUAAUUAUAAUUACCUGUAUAGAUUGUUUAGGAAGU
CUAAUCUCAAACCUUUUGAGAGAGAUAUUUCAACUGAAAUCUAUCAGGCCGGUAGCACACCUUGUAAUGGUGUUGAAGGUUUUAAUUGUUACUUUCCUUUACAAUCA
UAUGGUUUCCAACCCACUAAUGGUGUUGGUUACCAACCAUACAGAGUAGUAGUACUUUCUUUUGAACUUCUACAUGCACCAGCAACUGUUUGUGGACCUAAAAAGUC
UACUAAUUUGGUUAAAAACAAAUGUGUCAAUUUCAACUUCAAUGGUUUAACAGGCACAGGUGUUCUUACUGAGUCUAACAAAAAGUUUCUGCCUUUCCAACAAUUUG
GCAGAGACAUUGCUGACACUACUGAUGCUGUCCGUGAUCCACAGACACUUGAGAUUCUUGACAUUACACCAUGUUCUUUUGGUGGGUGUCAGUGUUAUAACACCAGGA
ACAAAUACUUCUAACCAGGUUGCUGUUCUUUAUCAGGAUGUUAACUGCACAGAAGUCCCUGUUGCUAUUCAUGCAGAUCAACUUACUCCUACUUGGCGUGUUUAUUC
UACAGGUUCUAAUGUUUUUCAAACACGUGCAGGCUGUUUAAUAGGGGCUGAACAUGUCAACAACUCAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUA
GUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGCUAGUCAAUCCAUCAUUGCCUACACUAUGUCACUUGGUGCAGAAAAUUCAGUUGCUUACUCU
AAUAACUCUAUUGCCAUACCCACAAAUUUUACUAUUAGUGUUACCACAGAAAUUCUACCAGUGUCUAUGACCAAGACAUCAGUAGAUUGUACAAUGUACAUUUGUGG
UGAUUCAACUGAAUGCAGCAAUCUUUUGUUGCAAUAUGGCAGUUUUUGUACACAAUUAAACCGUGCUUUAACUGGAAUAGCUGUUGAACAAGACAAAAACACCCAAG
AAGUUUUUGCACAAGUCAAACAAAUUUACAAAACACCACCAAUUAAAGAUUUUGGUGGUUUUAAUUUUUCACAAAUAUUACCAGAUCCAUCAAAACCAAGCAAGAGG
UCAUUUAUUGAAGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGGCUUCAUCAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCUGCUAGAGACCUCAUUUG
UGCACAAAAGUUUAACGGCCUUACUGUUUUGCCACCUUUGCUCACAGAUGAAAUGAUUGCUCAAUACACUUCUGCACUGUUAGCGGGUACAAUCACUUCUGGUUGGA
CCUUUGGUGCAGGUGCUGCAUUACAAAUACCAUUUGCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAGUUACACAGAAUGUUCUCUAUGAGAACCAAAAAUUG
AUUGCCAACCAAUUUAAUAGUGCUAUUGGCAAAAUUCAAGACUCACUUUCUUCCACAGCAAGUGCACUUGGAAAACUUCAAGAUGUGGUCAACCAAAAUGCACAAGC
UUUAAACACGCUUGUUAAACAACUUAGCUCCAAUUUUGGUGCAAUUUCAAGUGUUUUAAAUGAUAUCCUUUCACGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUG
AUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAGAGCUUCUGCUAAUCUUGCUGCUACUAAAAUG
UCAGAGUGUGUACUUGGACAAUCAAAAAGAGUUGAUUUUUGUGGAAAGGGCUAUCAUCUUAUGUCCUUCCCUCAGUCAGCACCUCAUGGUGUAGUCUUCUUGCAUGU
GACUUAUGUCCCUGCACAAGAAAAGAACUUCACAACUGCUCCUGCCAUUUGUCAUGAUGGAAAAGCACACUUUCCUCGUGAAGGUGUCUUUGUUUCAAAUGGCACAC
ACUGGUUUGUAACACAAAGGAAUUUUUAUGAACCACAAAUCAUUACUACAGACAACACAUUUGUGUCUGGUAACUGUGAUGUUGUAAUAGGAAUUGUCAACAACACA
GUUUAUGAUCCUUUGCAACCUGAAUUAGACUCAUUCAAGGAGGAGUUAGAUAAAUAUUUUAAGAAUCAUACAUCACCAGAUGUUGAUUUAGGUGACAUCUCUGGCAU
UAAUGCUUCAGUUGUAAACAUUCAAAAAGAAAUUGACCGCCUCAAUGAGGUUGCCAAGAAUUUAAAUGAAUCUCUCAUCGAUCUCCAAGAACUUGGAAAGUAUGAGC
AGUAUAUAAAAUGGCCAUGGUACAUUUGGCUAGGUUUUAUAGCUGGCUUGAUUGCCAUAGUAAUGGUGACAAUUAUGCUUUGCUGUAUGACCAGUUGCUGUAGUUGU
CUCAAGGGCUGUUGUUCUUGUGGAUCCUGCUGCAAAUUUGAUGAAGACGACUCUGAGCCAGUGCUCAAAGGAGUCAAAUUACAUUACACAUAAACGAACUUAUGGAU


CUUUAAAAGAAGGUCAAAUCAAUGAUAUGAUUUUAUCUCUUCUUAGUAAAGGUAGACUUAUAAUUAGAGAAAACAACAGAGUUGUUAUUUCUAGUGAUGUUCUUGUU
AACAACUAAACGAACAAUGUUUGUUUUUCUUGUUUUAUUGCCACUAGUCUCUAGUCAGUGUGUUAAUCUUACAACCAGAACUCAAUUACCCCCUGCAUACACUAAUU
CUUUCACACGUGGUGUUUAUUACCCUGACAAAGUUUUCAGAUCCUCAGUUUUACAUUCAACUCAGGACUUGUUCUUACCUUUCUUUUCCAAUGUUACUUGGUUCCAU
GCUAUACAUGUCUCUGGGACCAAUGGUACUAAGAGGUUUGAUAACCCUGUCCUACCAUUUAAUGAUGGUGUUUAUUUUGCUUCCACUGAGAAGUCUAACAUAAUAAG
AGGCUGGAUUUUUGGUACUACUUUAGAUUCGAAGACCCAGUCCCUACUUAUUGUUAAUAACGCUACUAAUGUUGUUAUUAAAGUCUGUGAAUUUCAAUUUUGUAAUG
AUCCAUUUUUGGGUGUUUAUUACCACAAAAACAACAAAAGUUGGAUGGAAAGUGAGUUCAGAGUUUAUUCUAGUGCGAAUAAUUGCACUUUUGAAUAUGUCUCUCAG
CCUUUUCUUAUGGACCUUGAAGGAAAACAGGGUAAUUUCAAAAAUCUUAGGGAAUUUGUGUUUAAGAAUAUUGAUGGUUAUUUUAAAAUAUAUUCUAAGCACACGCC
UAUUAAUUUAGUGCGUGAUCUCCCCUCAGGGUUUUUCGGCUUUAGAACCAUUGGUAGAUUUGCCAAUAGGUAUUAACAUCACUAGGUUUCAAACUUUACUUGCUUUAC
AUAGAAGUUAUUUGACUCCUGGUGAUUCUUCUUCAGGUUGGACAGCUGGUGCUGCAGCUUAUUAUGUGGGUUAUCUUCAACCUAGGACUUUUCUAUUAAAAUAUAAU
GAAAAUGGAACCAUUACAGAUGCUGUAGACUGUGCACUUGACCCUCUCUCAGAAACAAAGUGUACGUUGAAAUCCUUCACUGUAGAAAAAGGAAUCUAUCAAACUUC
UAACUUUAGAGUCCAACCAACAGAAUCUAUUGUUAGAUUUCCUAAUAUUACAAACUUGUGCCCUUUUGGUGAAGUUUUUAACGCCACCAGAUUUGCAUCUGUUUAUG
CUUGGAACAGGAAGAGAAUCAGCAACUGUGUUGCUGAUUAUUCUGUCCUAUAUAAUUCCGCAUCAUUUUCCACUUUUAAGUGUUAUGGAGUGUCUCCUACUAAAUUA
AAUGAUCUCUGCUUUACUAAUGUCUAUGCAGAUUCAUUUGUAAUUAGAGGUGAUGAAGUCAGACAAAUCGCUCCAGGGCAAACUGGAAAGAUUGCUGAUUAUAAUUA
UAAAUUACCAGAUGAUUUUACAGGCUGCGUUAUAGCUUGGAAUUCUAACAAUCUUGAUUCUAAGGUUGGUGGUAAUUAUAAUUACCUGUAUAGAUUGUUUAGGAAGU
CUAAUCUCAAACCUUUUGAGAGAGAUAUUUCAACUGAAAUCUAUCAGGCCGGUAGCACACCUUGUAAUGGUGUUGAAGGUUUUAAUUGUUACUUUCCUUUACAAUCA
UAUGGUUUCCAACCCACUAAUGGUGUUGGUUACCAACCAUACAGAGUAGUAGUACUUUCUUUUGAACUUCUACAUGCACCAGCAACUGUUUGUGGACCUAAAAAGUC
UACUAAUUUGGUUAAAAACAAAUGUGUCAAUUUCAACUUCAAUGGUUUAACAGGCACAGGUGUUCUUACUGAGUCUAACAAAAAGUUUCUGCCUUUCCAACAAUUUG
GCAGAGACAUUGCUGACACUACUGAUGCUGUCCGUGAUCCACAGACACUUGAGAUUCUUGACAUUACACCAUGUUCUUUUGGUGGUGUCAGUGUUAUAACACCAGGA
ACAAAUACUUCUAACCAGGUUGCUGUUCUUUAUCAGGAUGUUAACUGCACAGAAGUCCCUGUUGCUAUUCAUGCAGAUCAACUUACUCCUACUUGGCGUGUUUAUUC
UACAGGUUCUAAUGUUUUUCAAACACGUGCAGGCUGUUUAAUAGGGGCUGAACAUGUCAACAACUCAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCUA
GUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGCUAGUCAAUCCAUCAUUGCCUACACUAUGUCACUUGGUGCAGAAAAUUCAGUUGCUUACUCU
AAUAACUCUAUUGCCAUACCCACAAAUUUUACUAUUAGUGUUACCACAGAAAUUCUACCAGUGUCUAUGACCAAGACAUCAGUAGAUUGUACAAUGUACAUUUGUGG
UGAUUCAACUGAAUGCAGCAAUCUUUUGUUGCAAUAUGGCAGUUUUUGUACACAAUUAAACCGUGCUUUAACUGGAAUAGCUGUUGAACAAGACAAAAACACCCAAG
AAGUUUUUGCACAAGUCAAACAAAUUUACAAAACACCACCAAUUAAAGAUUUUGGUGGUUUUAAUUUUUCACAAAUAUUACCAGAUCCAUCAAAACCAAGCAAGAGG
UCAUUUAUUGAAGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGGCUUCAUCAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCUGCUAGAGACCUCAUUUG
UGCACAAAAGUUUAACGGCCUUACUGUUUUGCCACCUUUGCUCACAGAUGAAAUGAUUGCUCAAUACACUUCUGCACUGUUAGCGGGUACAAUCACUUCUGGUUGGA
CCUUUGGUGCAGGUGCUGCAUUACAAAUACCAUUUGCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAGUUACACAGAAUGUUCUCUAUGAGAACCAAAAAUUG
AUUGCCAACCAAUUUAAUAGUGCUAUUGGCAAAAUUCAAGACUCACUUUCUUCCACAGCAAGUGCACUUGGAAAACUUCAAGAUGUGGUCAACCAAAAUGCACAAGC
UUUAAACACGCUUGUUAAACAACUUAGCUCCAAUUUUGGUGCAAUUUCAAGUGUUUUAAAUGAUAUCCUUUCACGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUG
AUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAGAGCUUCUGCUAAUCUUGCUGCUACUAAAAUG
UCAGAGUGUGUACUUGGACAAUCAAAAAGAGUUGAUUUUUGUGGAAAGGGCUAUCAUCUUAUGUCCUUCCCUCAGUCAGCACCUCAUGGUGUAGUCUUCUUGCAUGU
GACUUAUGUCCCUGCACAAGAAAAGAACUUCACAACUGCUCCUGCCAUUUGUCAUGAUGGAAAAGCACACUUUCCUCGUGAAGGUGUCUUUGUUUCAAAUGGCACAC
ACUGGUUUGUAACACAAAGGAAUUUUUAUGAACCACAAAUCAUUACUACAGACAACACAUUUGUGUCUGGUAACUGUGAUGUUGUAAUAGGAAUUGUCAACAACACA
GUUUAUGAUCCUUUGCAACCUGAAUUAGACUCAUUCAAGGAGGAGUUAGAUAAAUAUUUUAAGAAUCAUACAUCACCAGAUGUUGAUUUAGGUGACAUCUCUGGCAU
UAAUGCUUCAGUUGUAAACAUUCAAAAAGAAAUUGACCGCCUCAAUGAGGUUGCCAAGAAUUUAAAUGAAUCUCUCAUCGAUCUCCAAGAACUUGGAAAGUAUGAGC
AGUAUAUAAAAUGGCCAUGGUACAUUUGGCUAGGUUUUAUAGCUGGCUUGAUUGCCAUAGUAAUGGUGACAAUUAUGCUUUGCUGUAUGACCAGUUGCUGUAGUUGU
CUCAAGGGCUGUUGUUCUUGUGGAUCCUGCUGCAAAUUUGAUGAAGACGACUCUGAGCCAGUGCUCAAAGGAGUCAAAUUACAUUACACAUAAACGAACUUAUGGAU

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNP
VLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSA
NNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRS
YLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFP
NITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPG
QTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSY
GFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAV
RDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVN
NSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYI
CGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKV
TLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFN
GIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEA
EVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPA
QEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEE
LDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIML
CCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

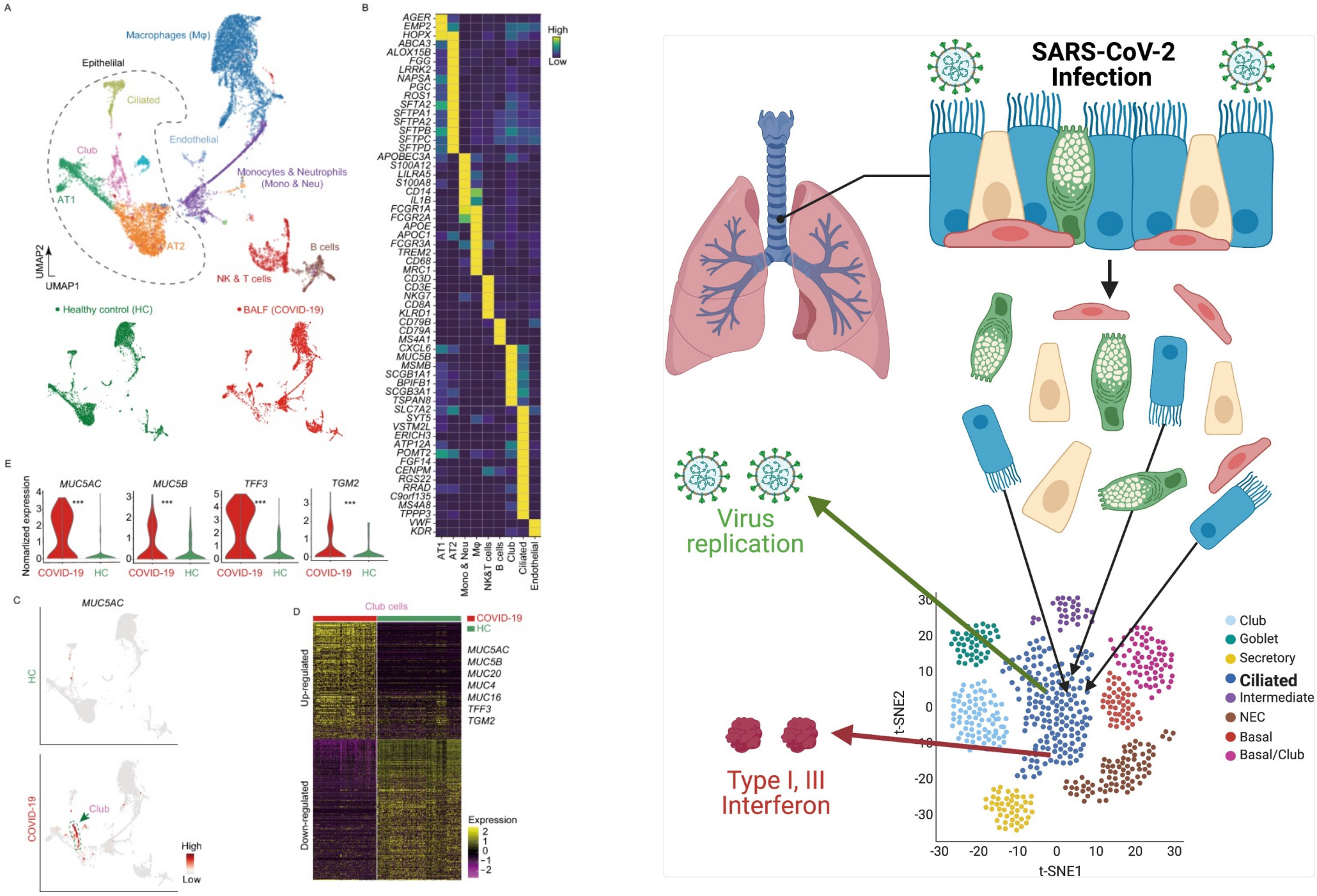

A
SARS-CoV
SARS-CoV
MERS-CoV
2019-nCoV
Bat CoV
MERS-CoV
?
2019-nCoV
B
Bat CoV
2019-nCoV
MERS-CoV
HCoV-NL63
HCoV-229E
SARS-CoV
HCoV-OC43
HCoV-HKU1
Insertion 1
Insertion 2
Insertion 3
Insertion 4



NEWS · 30 NOVEMBER 2020

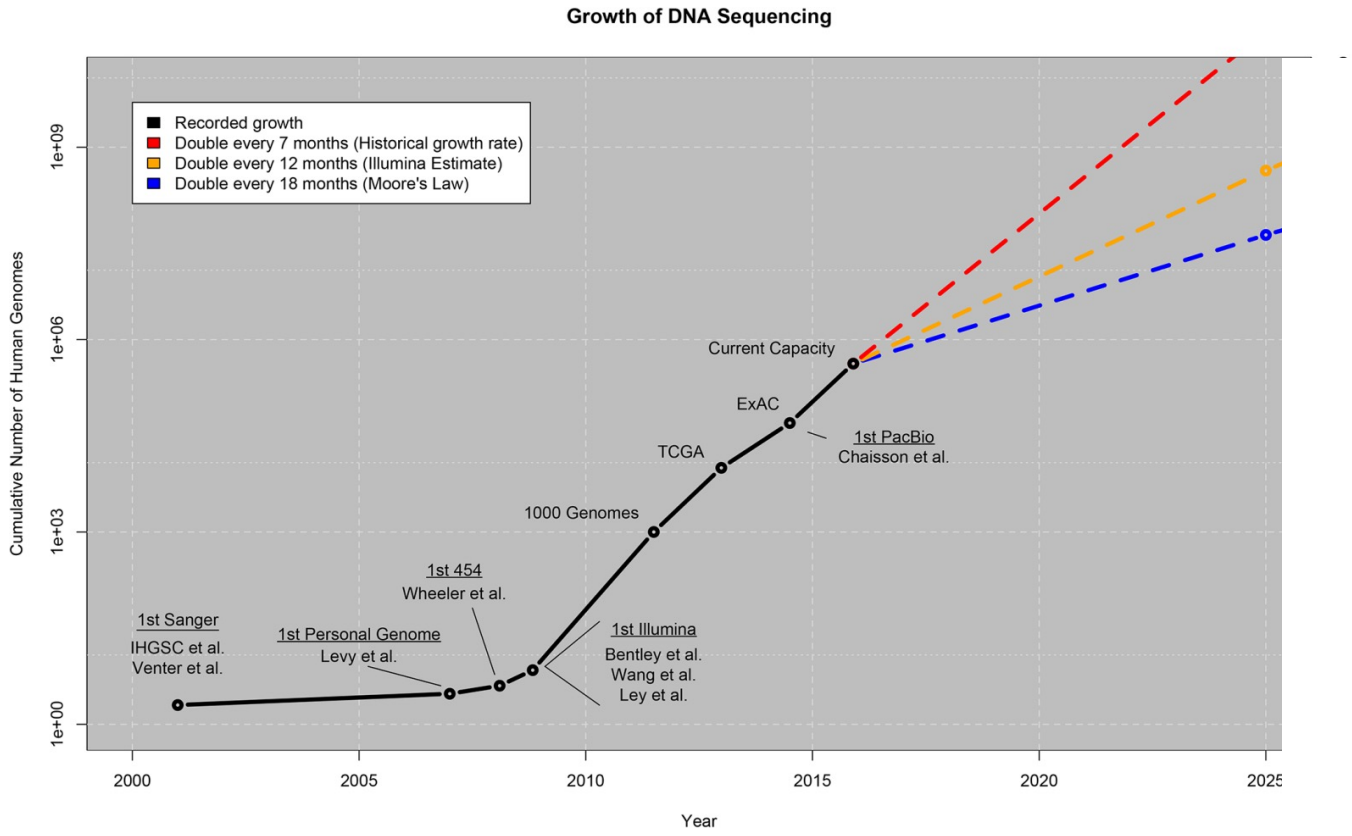
‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

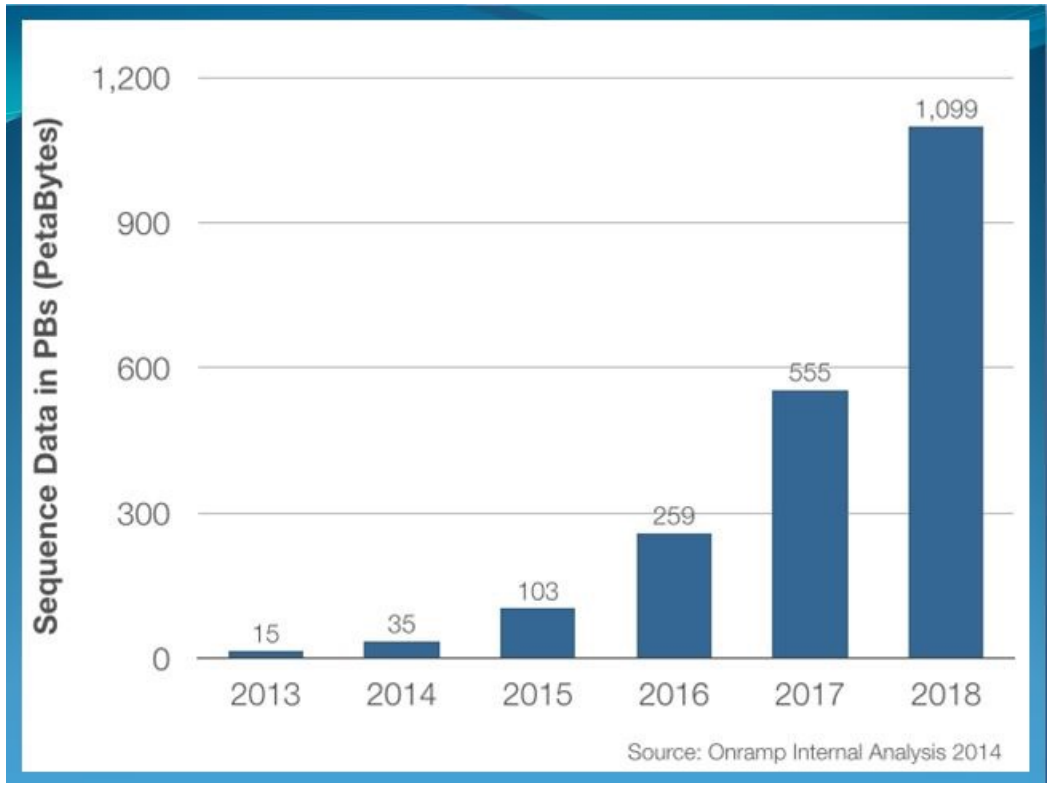
Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.


A
Macrophages (Mφ)
Epithelilal
Ciliated
Endothelial
Club
Monocytes & Neutrophils
(Mono & Neu)
AT1
AT2
B cells
NK & T cells
UMAP2
UMAP1
Healthy control (HC)
BALF (COVID-19)
B
High
Low
AGER
EMP2
HOPX
ABCA3
ALOX15B
FGG
LRRK2
NAPSA
PGC
ROS1
SFTA2
SFTPA1
SFTPA2
SFTPB
SFTPC
SFTPD
APOBEC3A
S100A12
LILRA5
S100A8
CD14
IL1B
FCGR1A
FCGR2A
APOE
APOC1
FCGR3A
TREM2
CD68
MRC1
CD3D
CD3E
NKG7
CD8A
KLRD1
CD79B
CD79A
MS4A1
CXCL6
MUC5B
MSMB
SCGB1A1
BPIFB1
SCGB3A1
TSPAN8
SLC7A2
SYT5
VSTM2L
ERICH3
ATP12A
POMT2
FGF14
CENPM
RGS22
RRAD
C9orf135
MS4A8
TPPP3
VWF
KDR
AT1
AT2
Mono & Neu
Mφ
NK&T cells
B cells
Club
Ciliated
Endothelial
E
MUC5AC
MUC5B
TFF3
TGM2
Nomarlized expression
***
COVID-19
HC
C
MUC5AC
HC
COVID-19
Club
High
Low
D
Club cells
COVID-19
HC
Up-regulated
Down-regulated
MUC5AC
MUC5B
MUC20
MUC4
MUC16
TFF3
TGM2
Expression
SARS-CoV-2
Infection
Virus
replication
Type I, III
Interferon
t-SNE2
t-SNE1
Club
Goblet
Secretory
Ciliated
Intermediate
NEC
Basal
Basal/Club

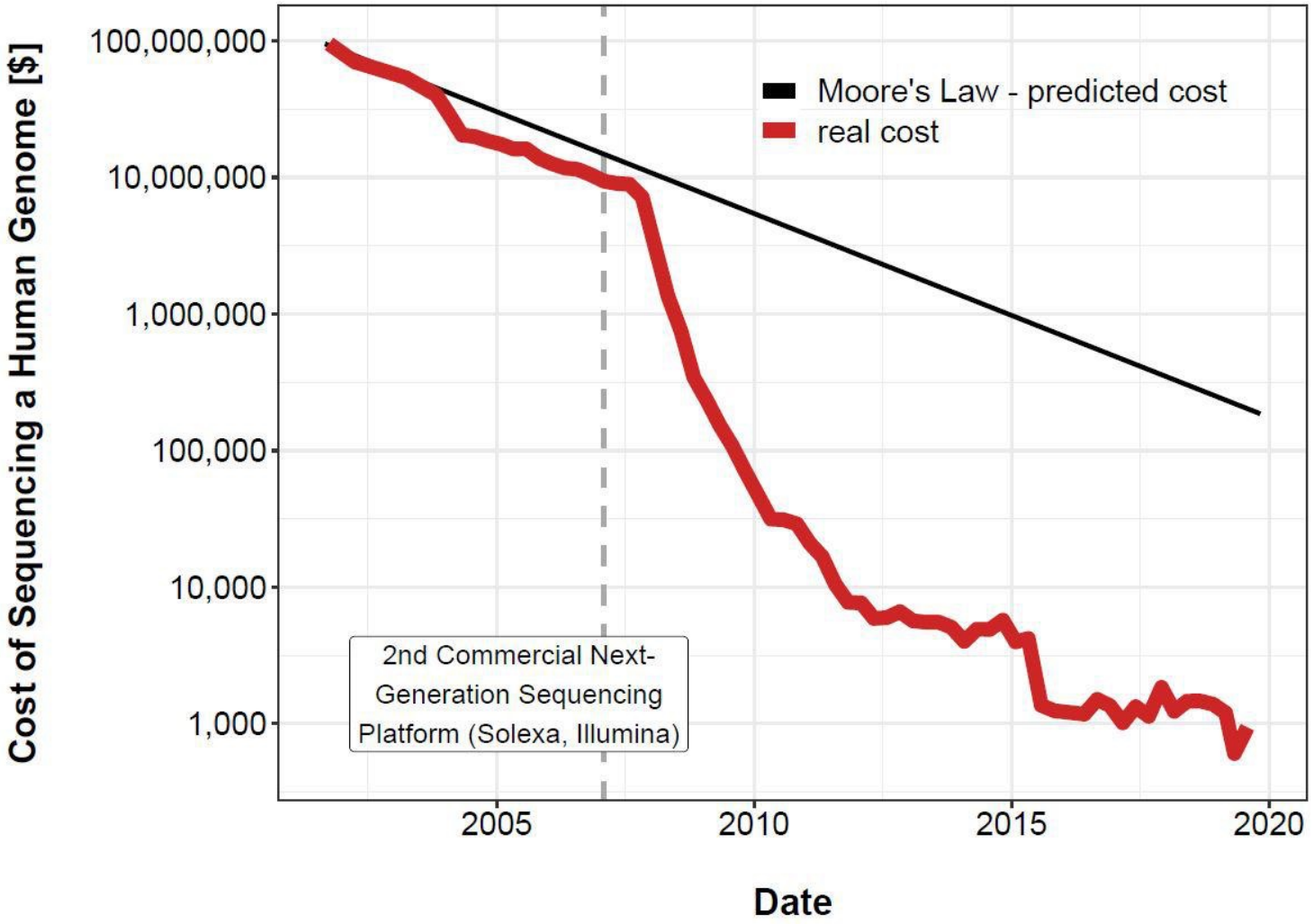

Experimental Biologist
Computational Biologist
NO EATING
NO DRINKING
NO BREATHING
NO OPEN TOE SHOES
NO SHORTS
NO MINI SKIRTS
ALWAYS LAB COATS ON
LIFE IS BEAUTIFUL
www.biocomicals.com

Information explosion in biology





Cost of sequencing the human genome



A Deluge of Data

adsp
Alzheimer's Disease Sequencing Project

1000 Plant Genomes

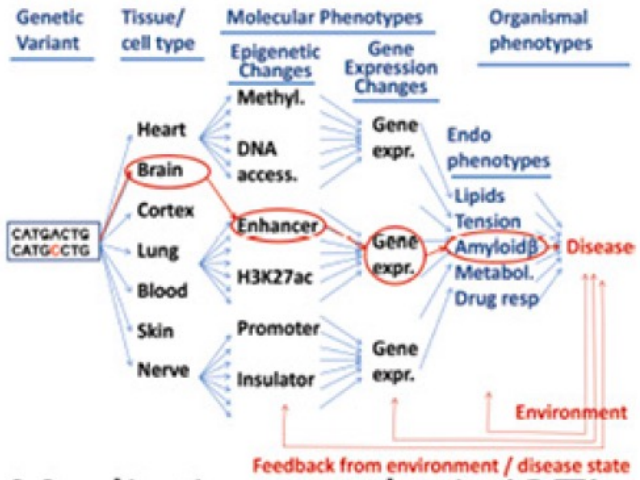
Fish-T1K
Transcriptomes of 1000 Fishes

International Cancer Genome Consortium

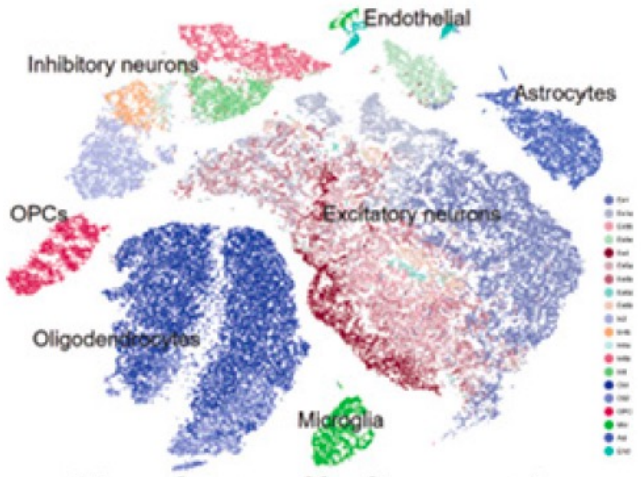
THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute

NIH HUMAN MICROBIOME PROJECT

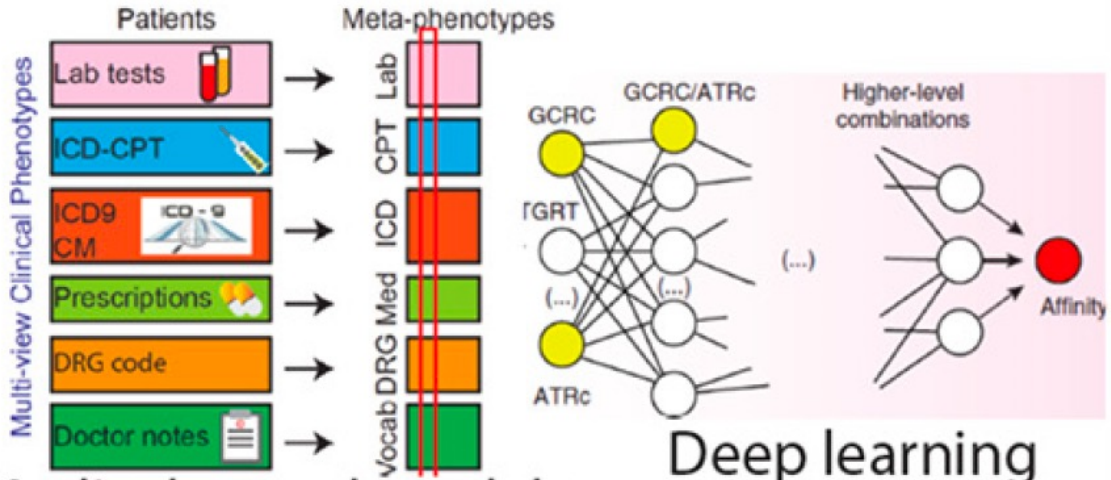
Many different applications



Mediation analysis/QTLs



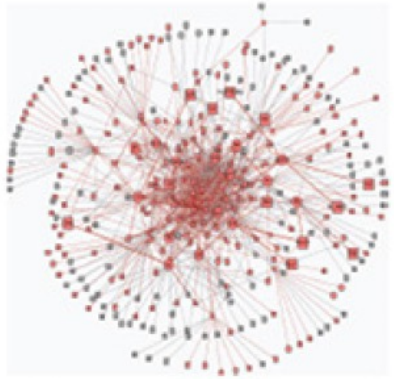
Single-cell dissection



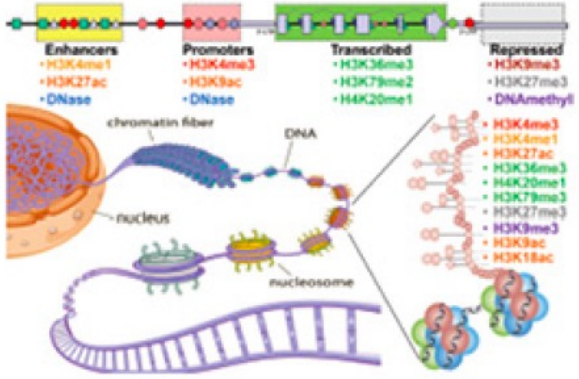
Medical record models

Deep learning

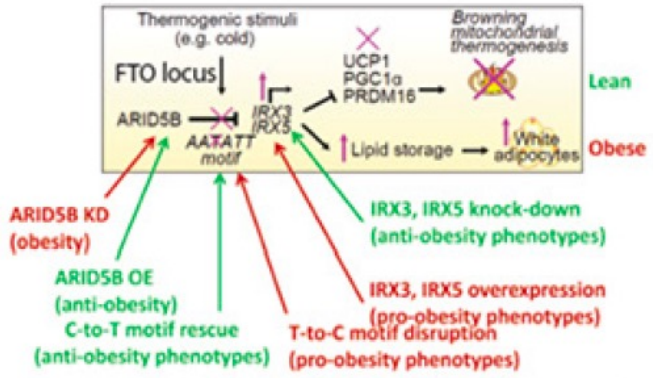
DNA motifs



Gene networks



Epigenomics



Manipulate disease circuitry

Multi-phenotype

mRNA vaccinces

- The full genetic sequence of SARS-CoV-2 was deposited on January 5, 2020 (and made public on January 10)

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

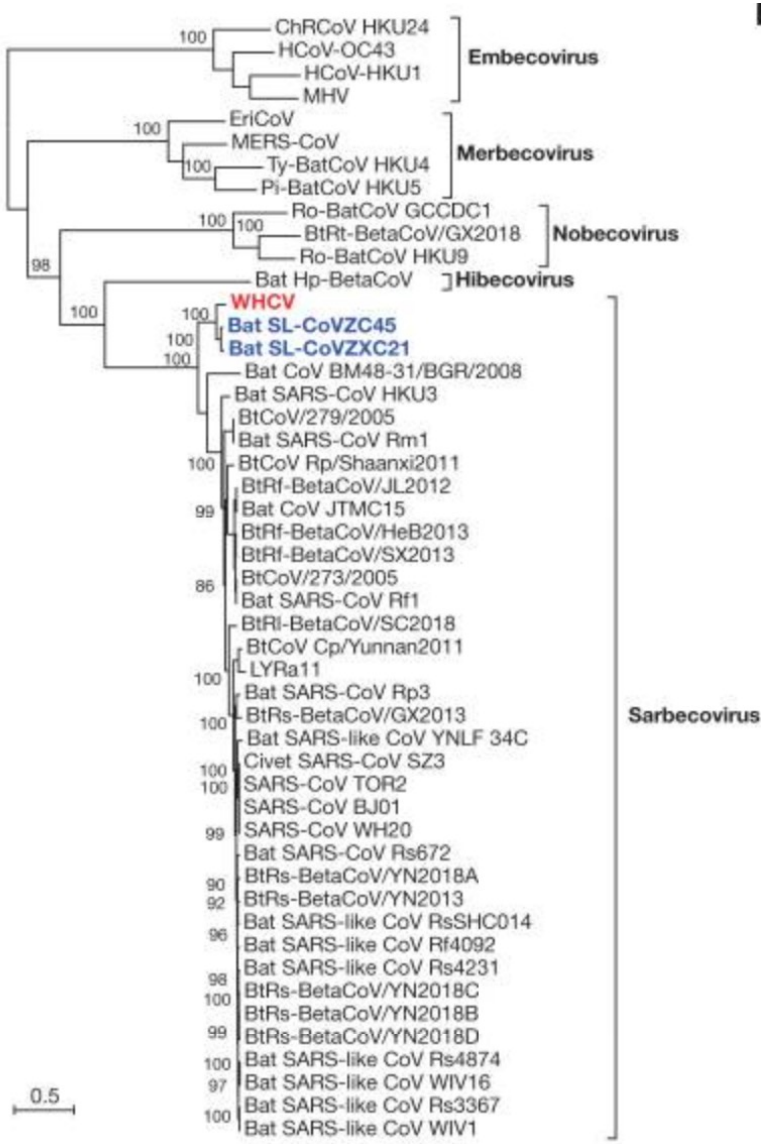
GenBank: MN908947.3

FASTA Graphics

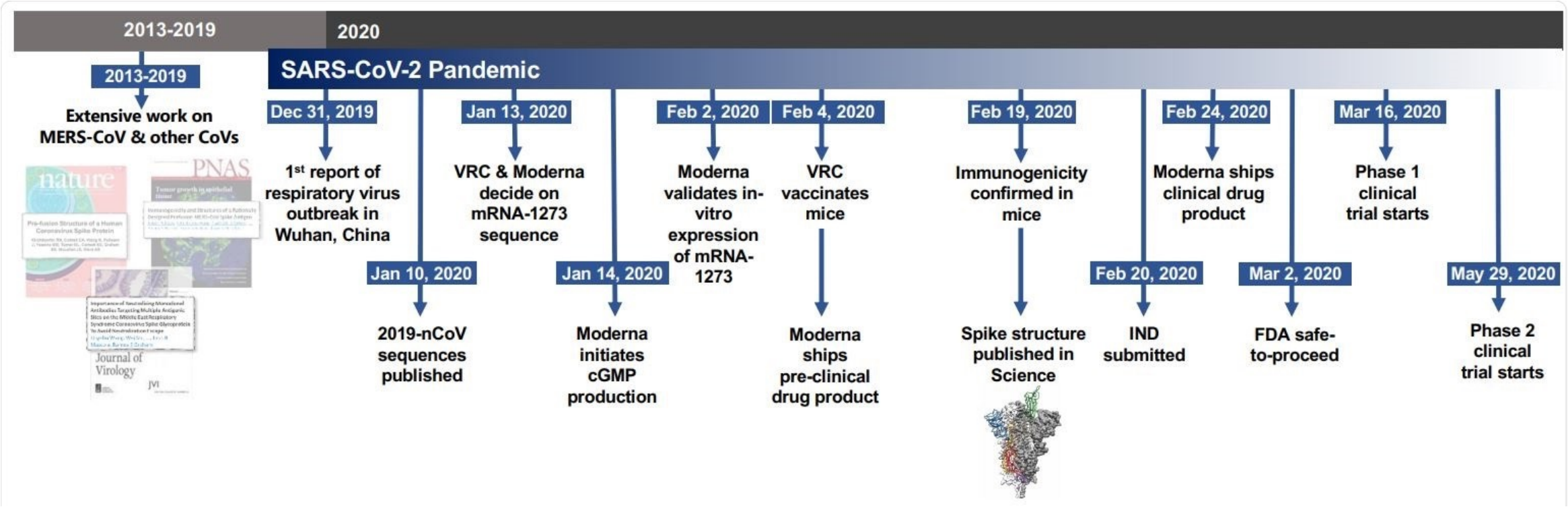
Go to:

```
LOCUS       MN908947              29903 bp ss-RNA     linear   VRL 18-MAR-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,
            complete genome.
ACCESSION   MN908947
VERSION     MN908947.3
KEYWORDS    .
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 29903)
  AUTHORS   Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,
            Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,
            Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.
  TITLE     A new coronavirus associated with human respiratory disease in
            China
  JOURNAL   Nature 579 (7798), 265-269 (2020)
   PUBMED   32015508
REFERENCE   2  (bases 1 to 29903)
  AUTHORS   Wu,F., Zhao,S., Yu,B., Chen,Y.-M., Wang,W., Hu,Y., Song,Z.-G.,
            Tao,Z.-W., Tian,J.-H., Pei,Y.-Y., Yuan,M.L., Zhang,Y.-L.,
```

https://www.ncbi.nlm.nih.gov/nuccore/MN908947

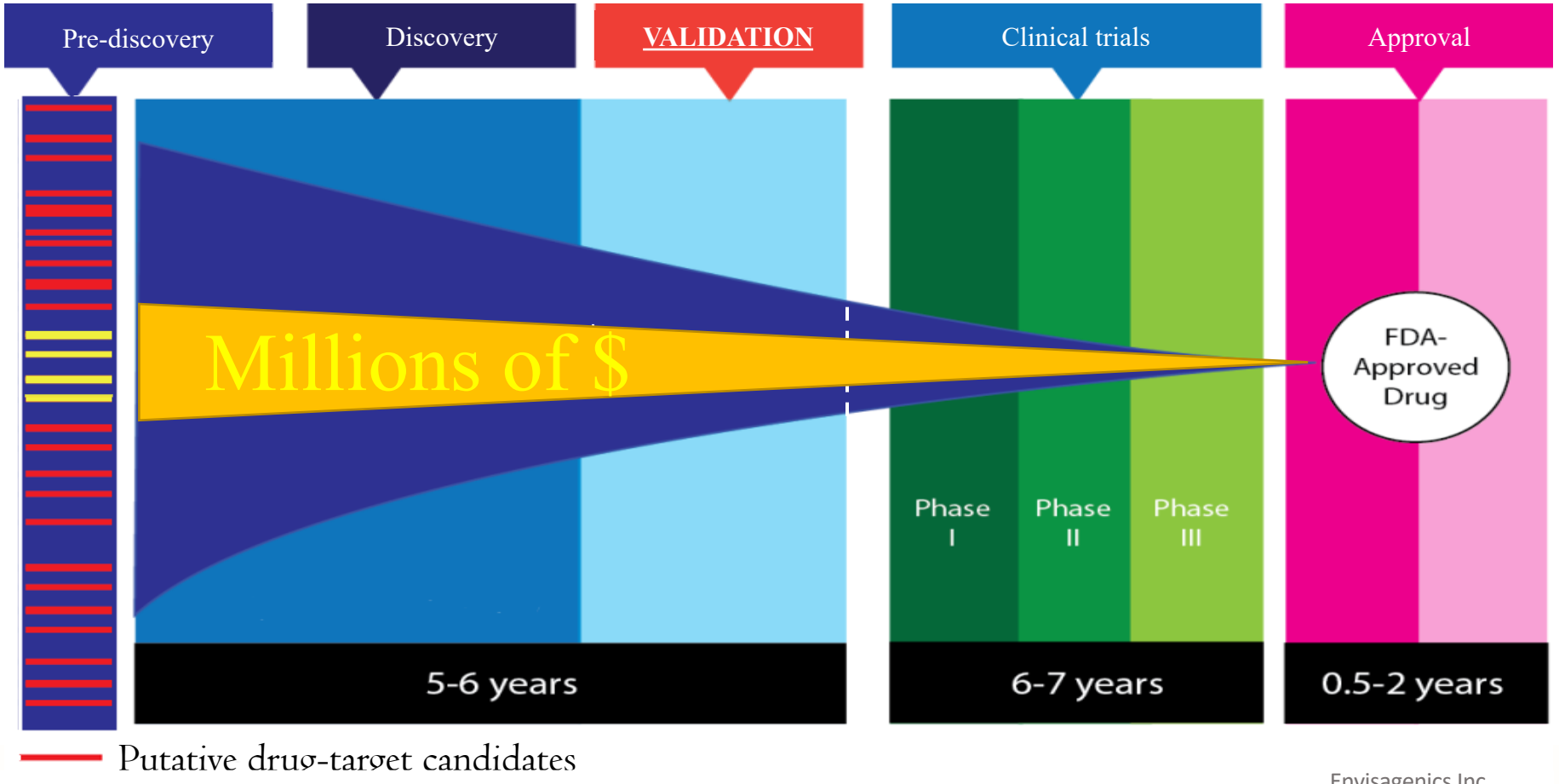


Moderna vaccine development timeline



2 days from sequence to vaccine candidate!

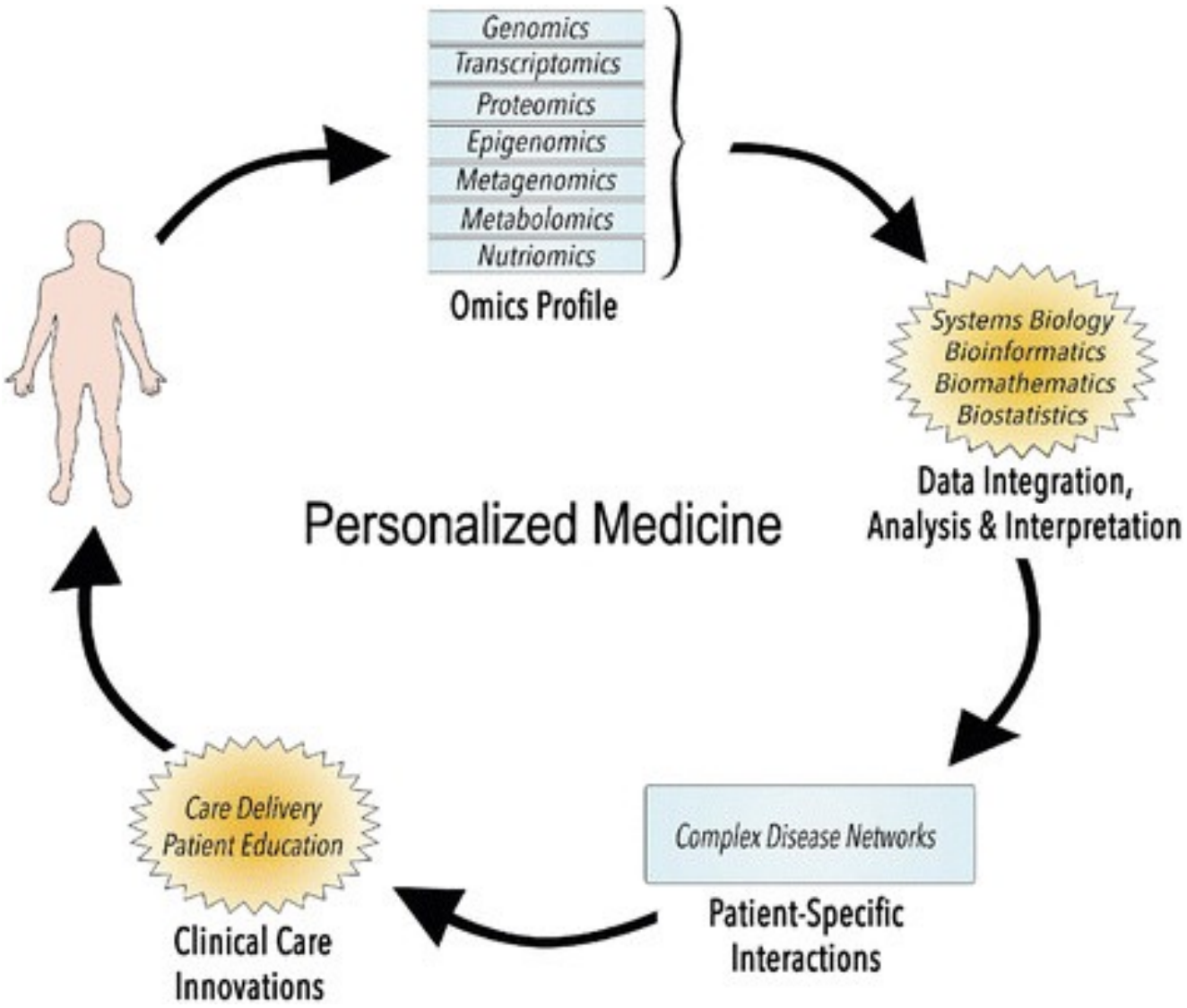
Bioinformatics can dramatically reduce the cost and time for developing a new drug

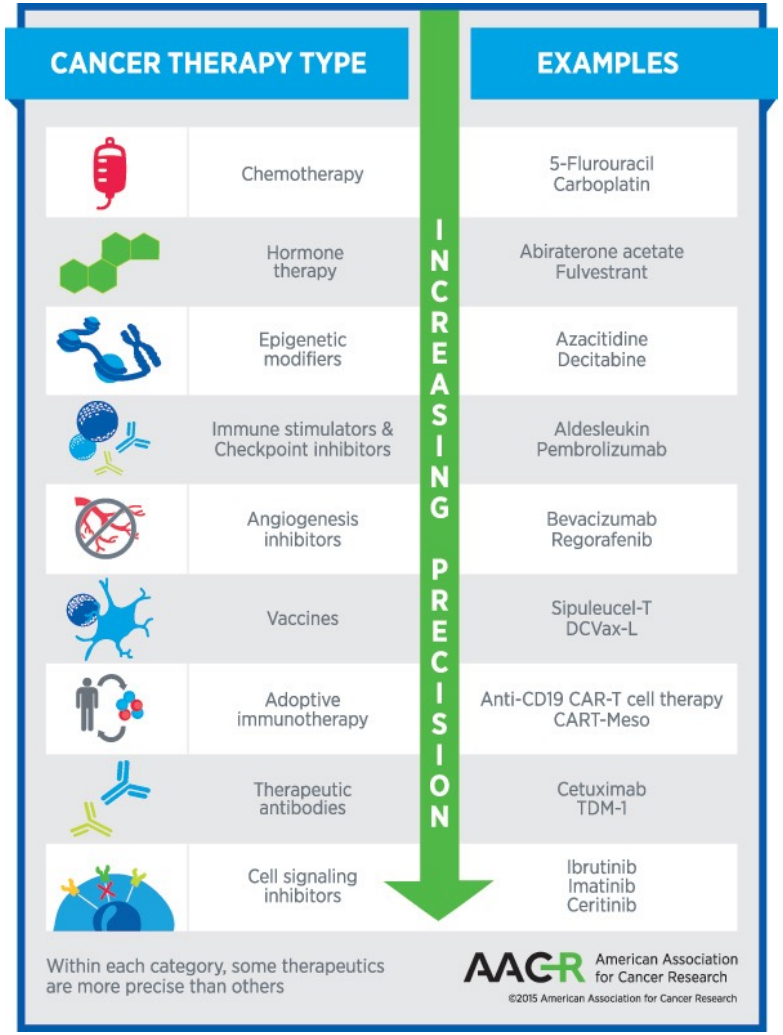


Envisagenics Inc.

From drug discovery through FDA approval, developing a new medicine takes at least 10 years on average and costs an average of $2.6 billion.* Less than 12% of the candidate medicines that make it into Phase I clinical trials will be approved by the FDA.

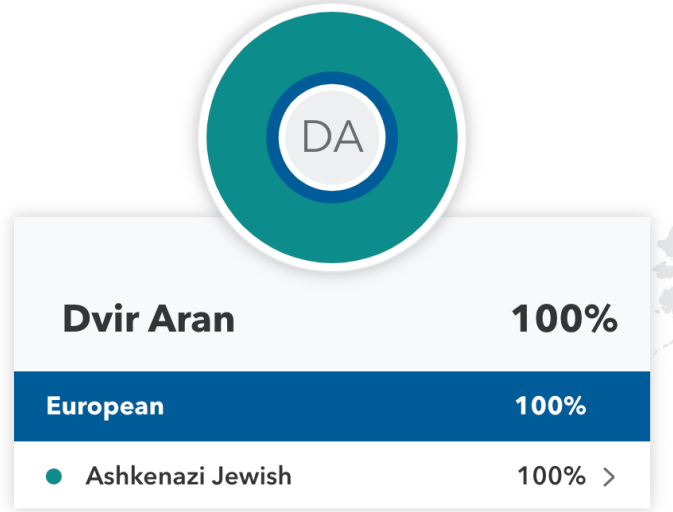
Personalized/precision medicine





Ancestry composition and identifying far relatives

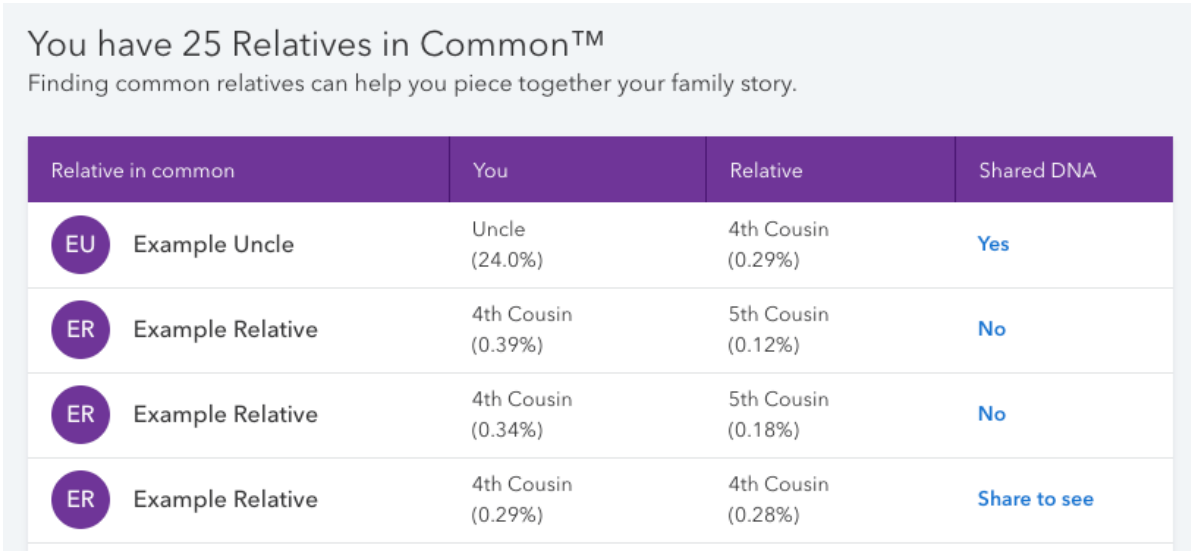
- ~50 million people in the world deposited their genetic information in public databases.
- Bioinformatics methods allow to reconstruct family trees.



23andMe

ancestry

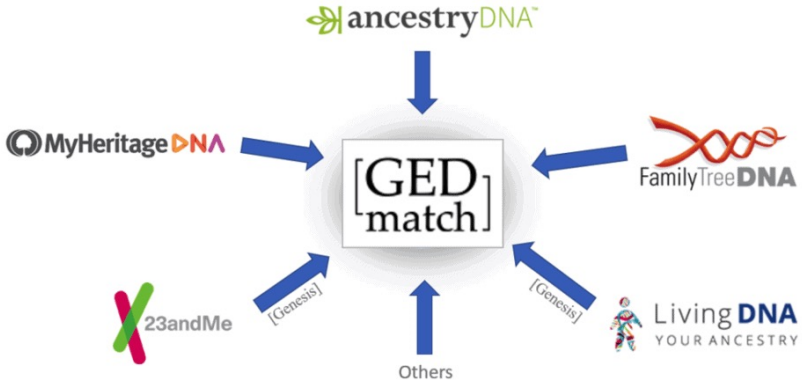
MyHeritage

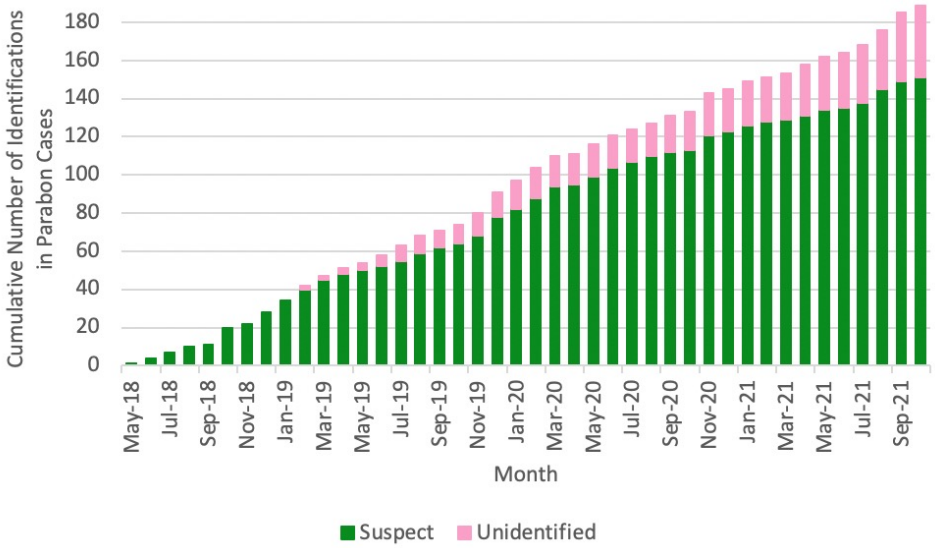


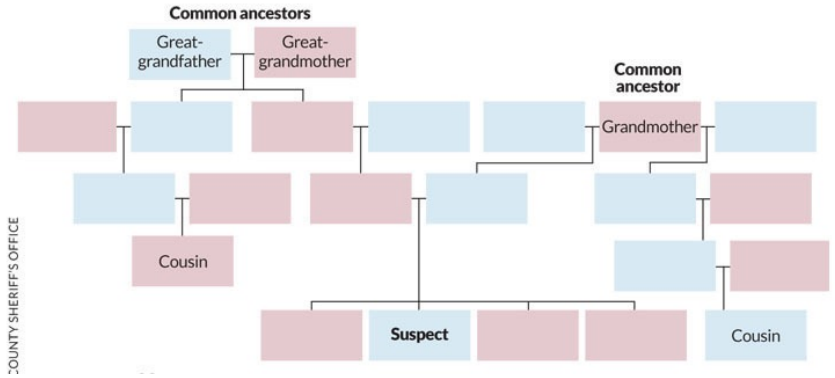
Identify far relatives… and criminals!

Long-range familial search

Golden State Killer

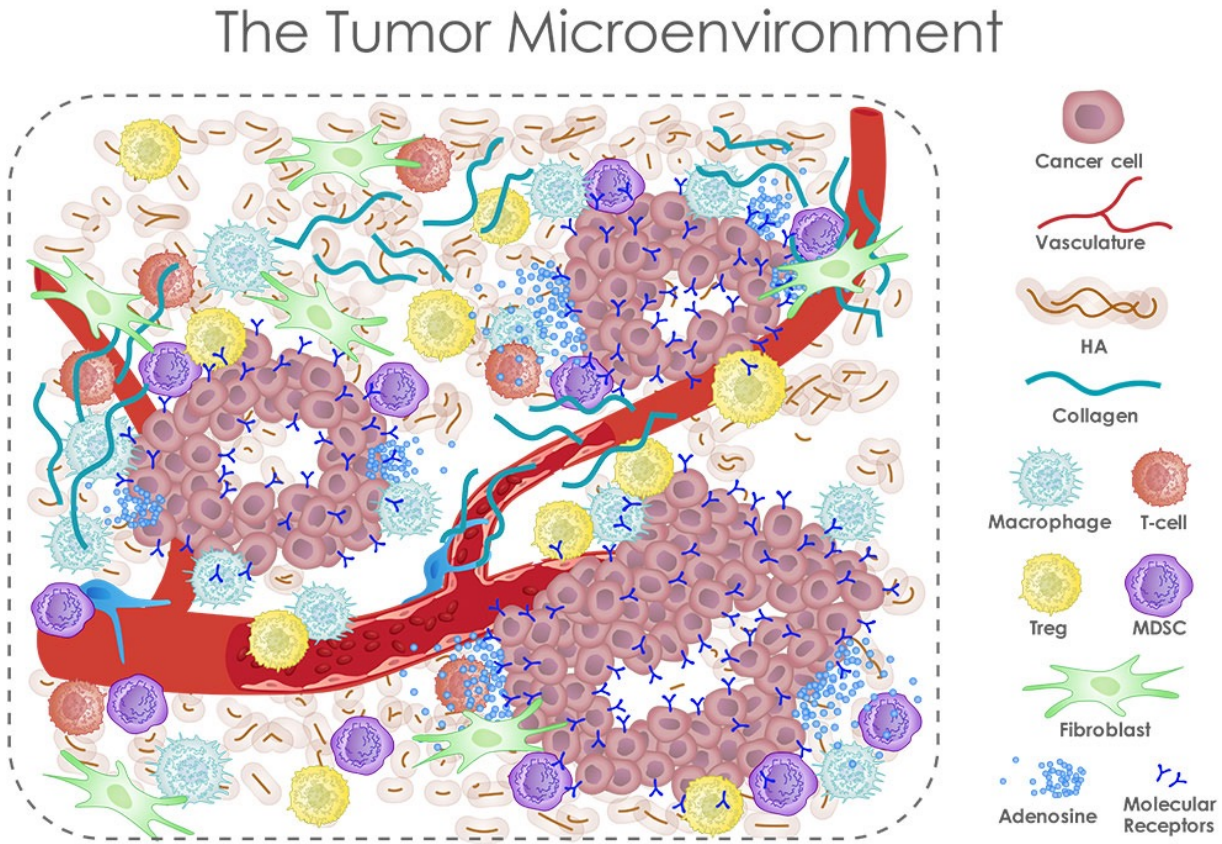




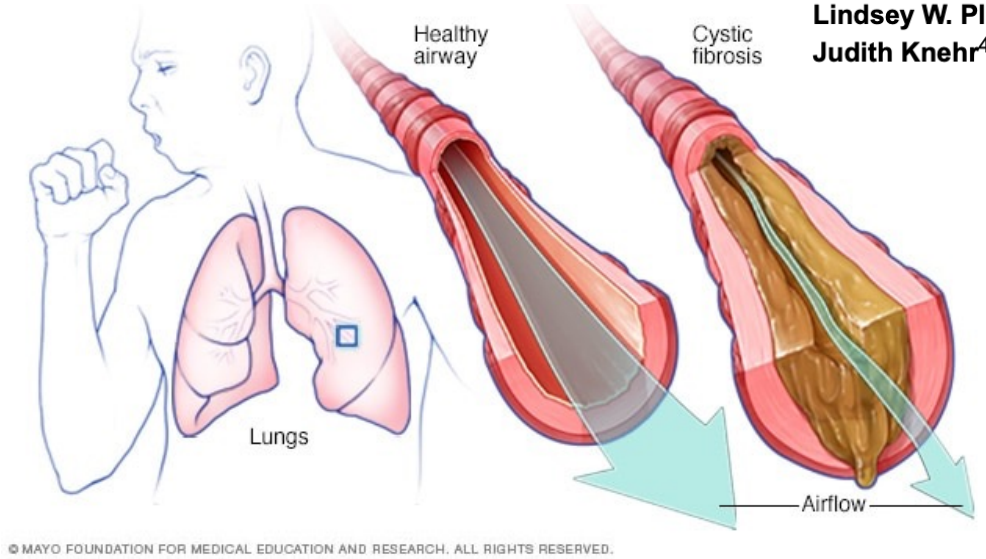


Single-cell biology

- How a single cell develops into an adult animal with multiple organs and billions of cells?
- How a mutation causes a specific disease?
- How is the immune system shaped in the tumor microenvironment?



Cystic Fibrosis



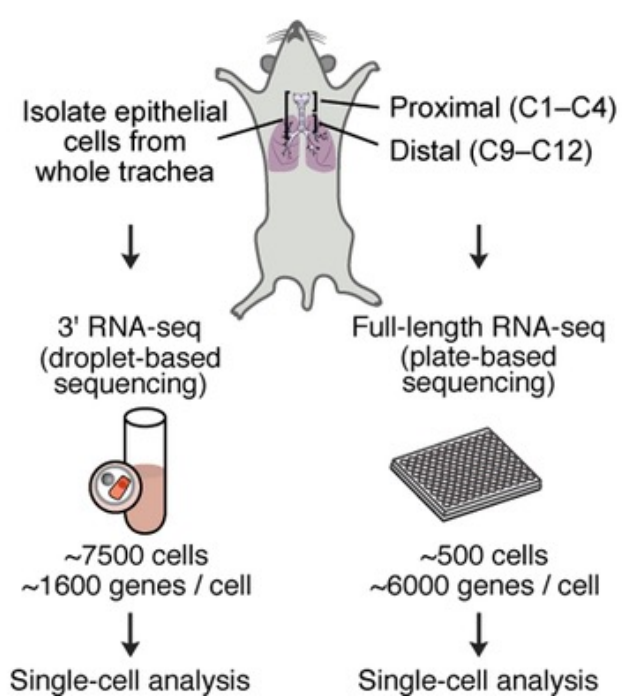
- Accumulation of thick, sticky mucus
- Coughing or shortness of breath
- Frequent chest infections
- Salty-tasting skin
- Poor growth and poor weight gain despite normal food intake
- **Mutation in the CFTR gene.**
- 1/2500 babies
- 1/25 every AZ is a carrier

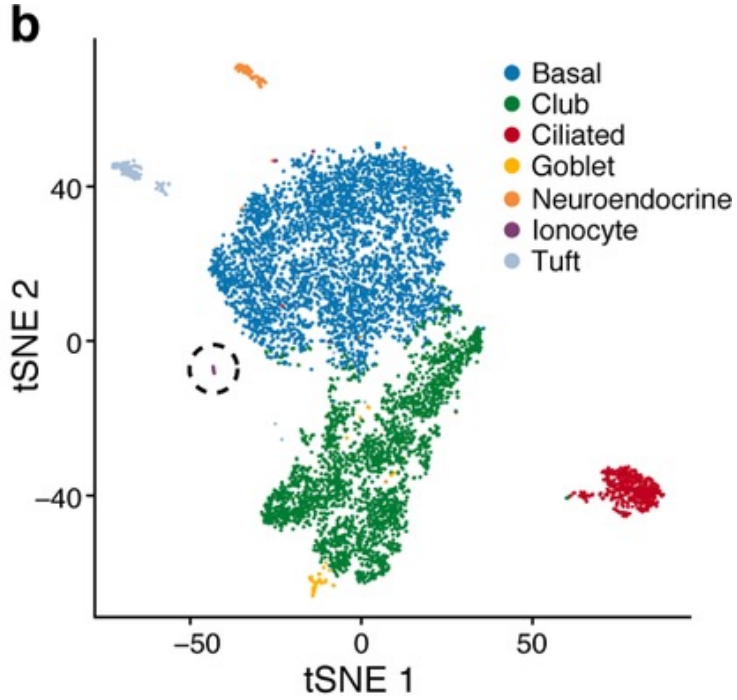
A single cell atlas of the tracheal epithelium reveals the CFTR-rich pulmonary ionocyte

Lindsey W. Plasschaert[#1], Rapolas Žilionis[#2,3], Rayman Choo-Wing[1], Virginia Savova[2], Judith Knehr[4], Guglielmo Roma[4], Allon M. Klein[2,†], and Aron B. Jaffe[1,†]

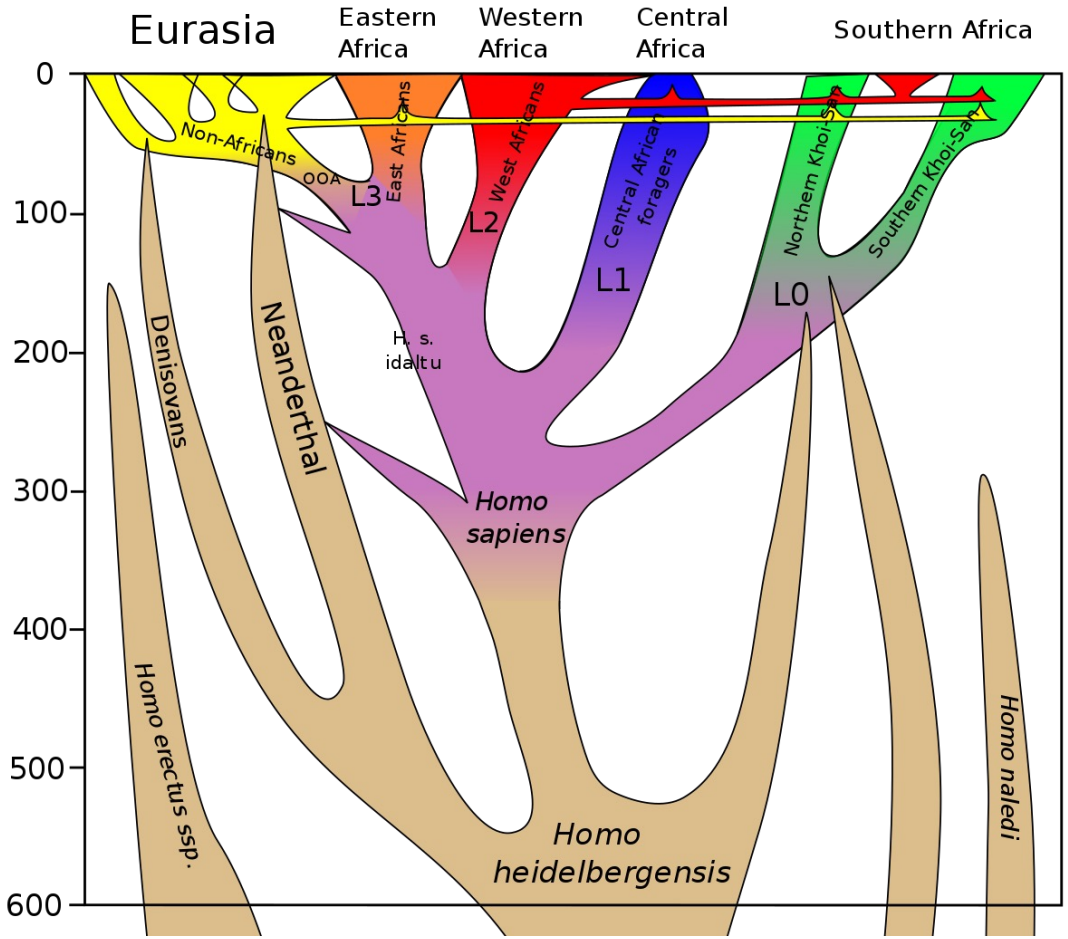
A revised airway epithelial hierarchy includes CFTR-expressing ionocytes

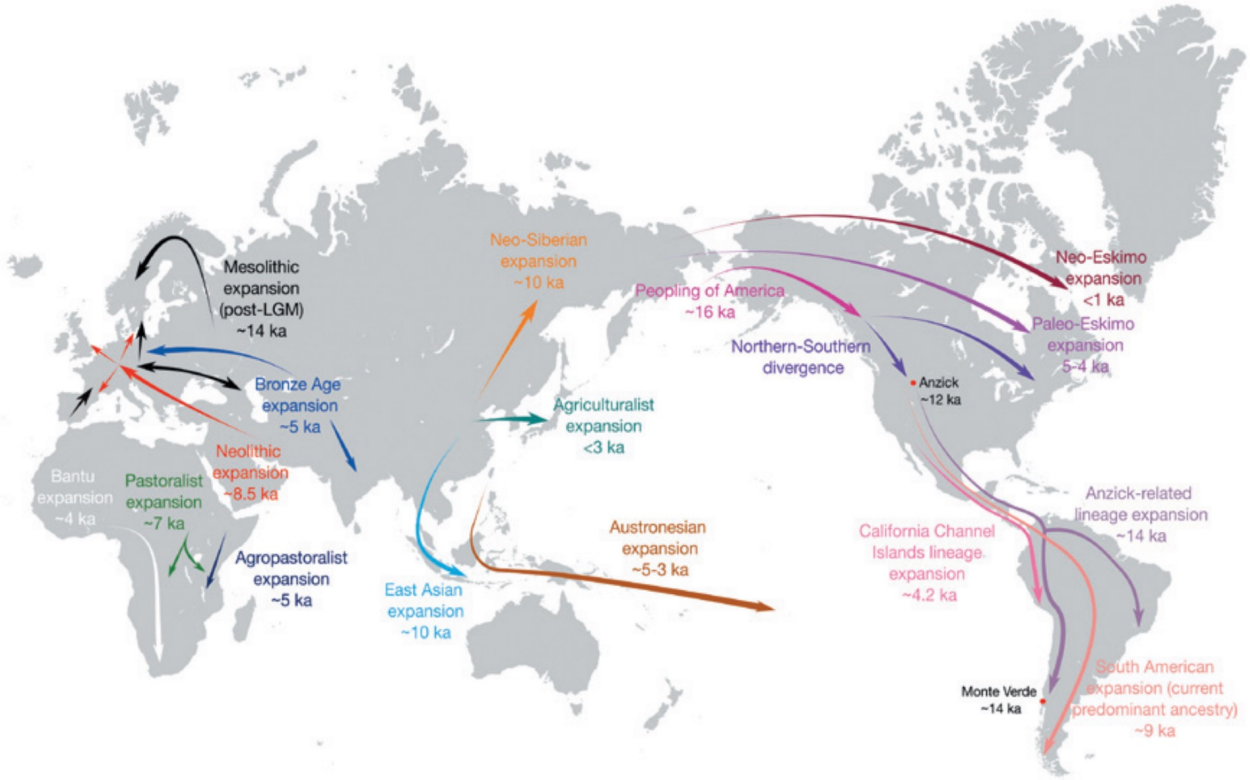
Daniel T. Montoro[#1,2,3], Adam L. Haber[#4], Moshe Biton[#4,5], Vladimir Vinarsky[1,2,3], Brian Lin[1,2,3], Susan Birket[6,7], Feng Yuan[8], Sijia Chen[9], Hui Min Leung[10,11], Jorge Villoria[1,2,3], Noga Rogel[4], Grace Burgin[4], Alexander Tsankov[4], Avinash Waghray[1,2,3], Michal Slyper[4], Julia Waldmann[4], Lan Nguyen[4], Danielle Dionne[4], Orit Rozenblatt-Rosen[4], Purushothama Rao Tata[12,13,14,15], Hongmei Mou[16,17], Manjunatha Shivaraju[1,2,3], Hermann Bihler[18], Martin Mense[18], Guillermo J. Tearney[10,11], Steven M. Rowe[6,7], John F. Engelhardt[8], Aviv Regev[4,19,§], and Jayaraj Rajagopal[1,2,3,§]





Human history



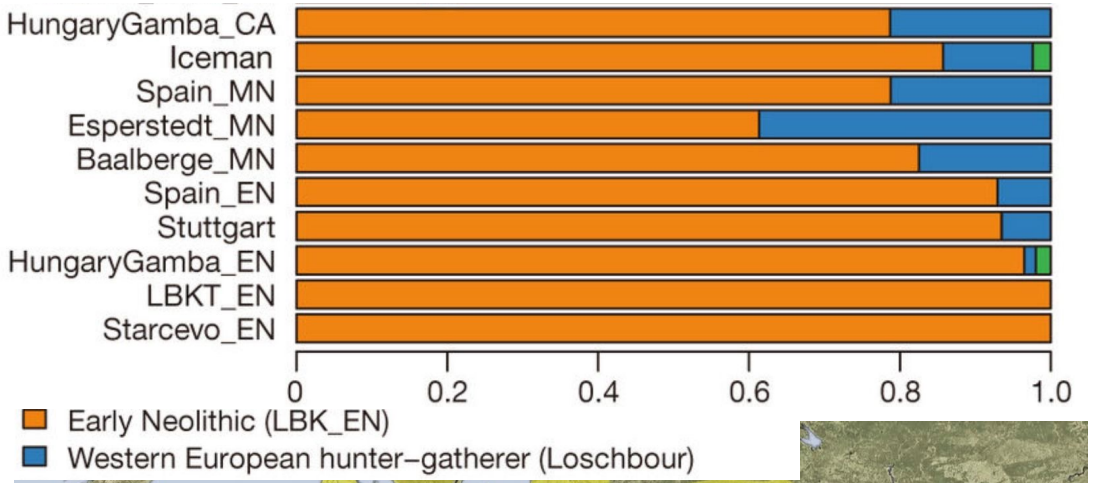


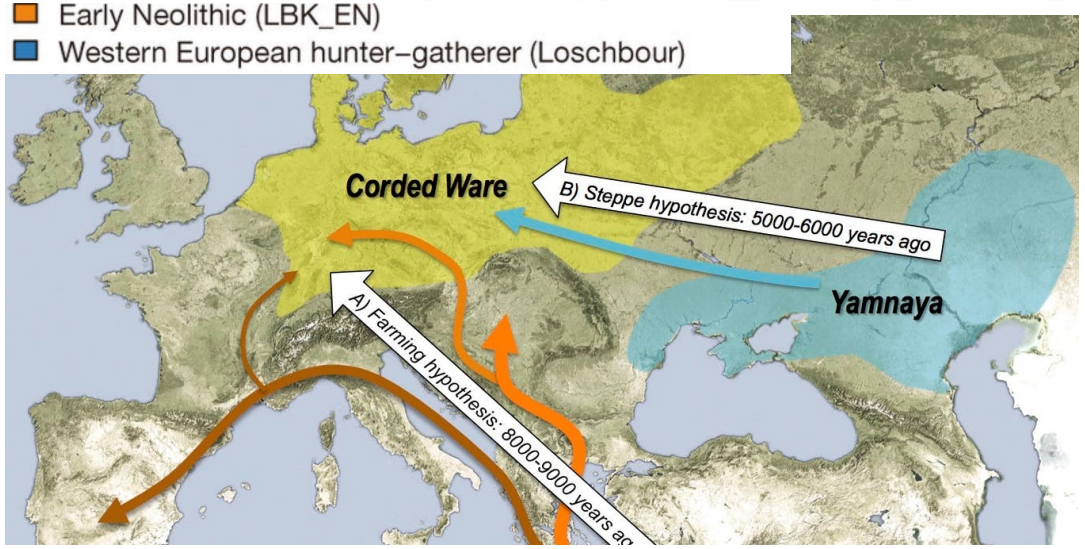
Ancient DNA

Svaante Paabo

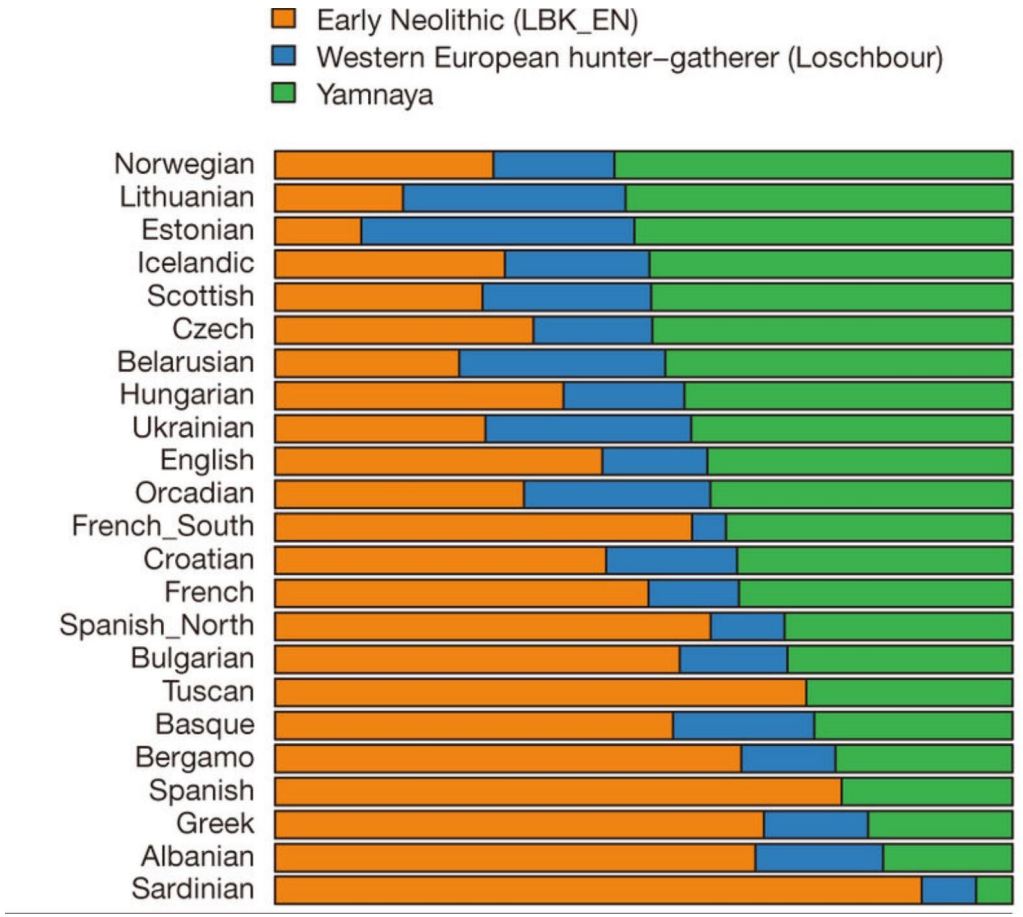
Ancient DNA

>6K years ago

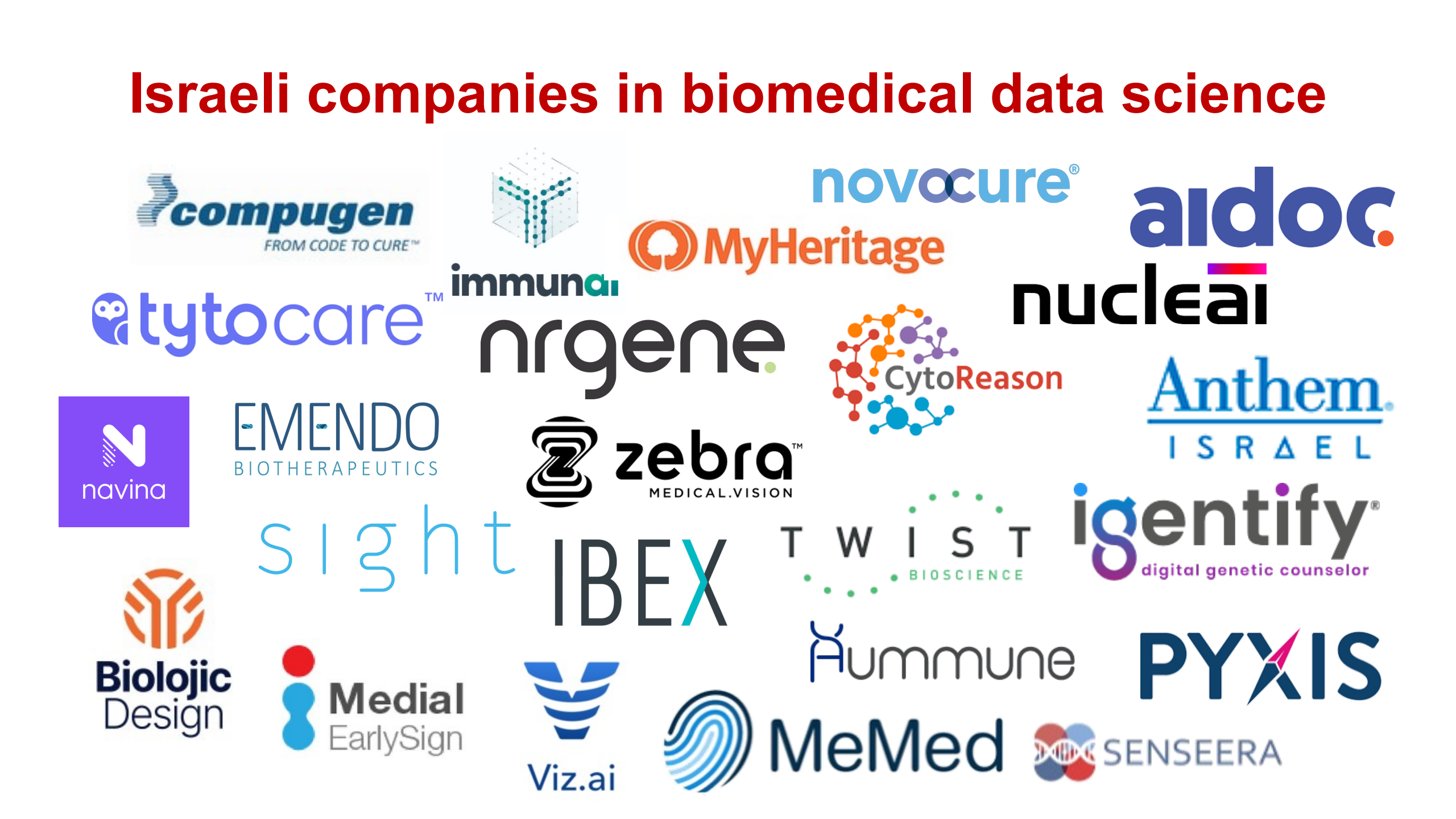




Today



Israeli companies in biomedical data science

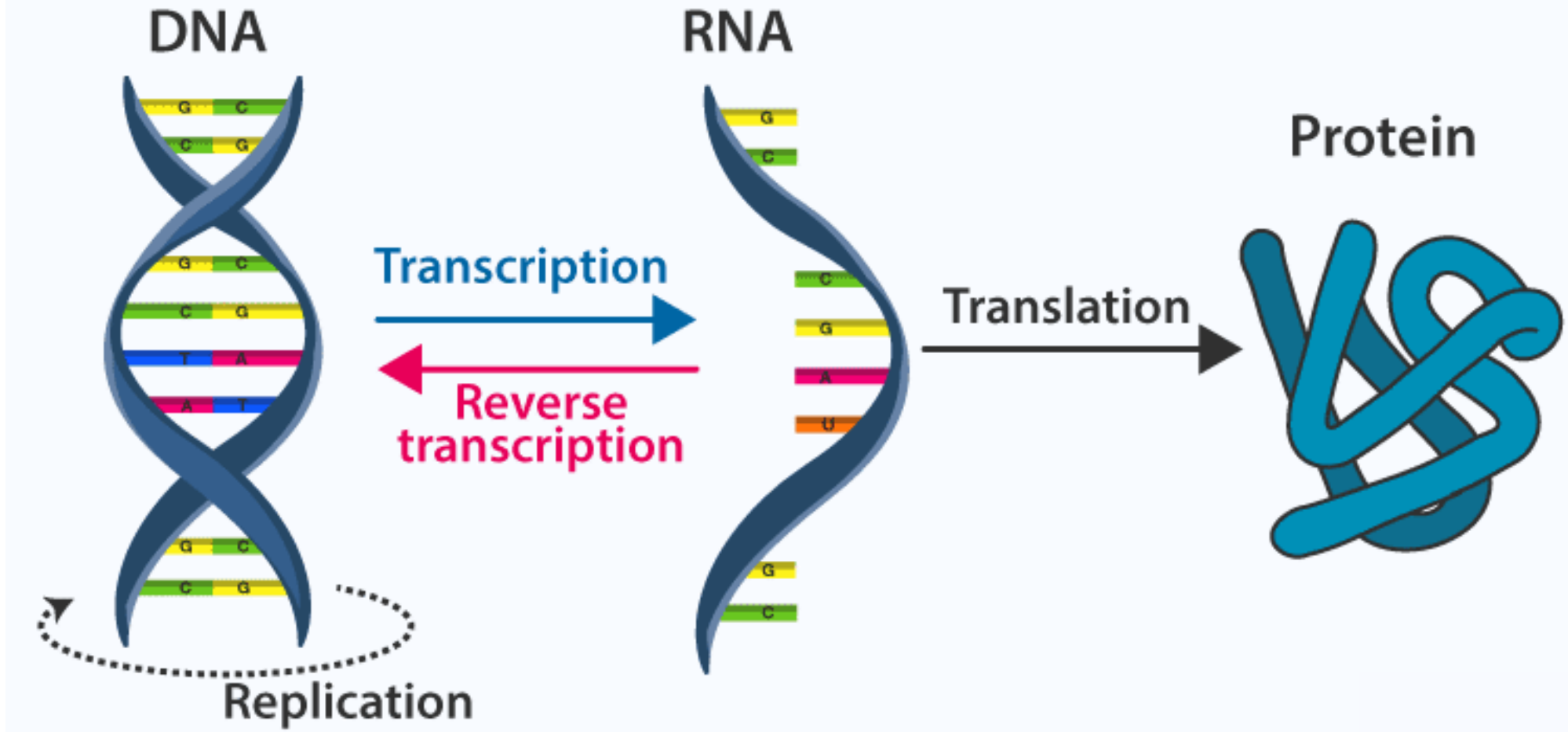



AI
AI

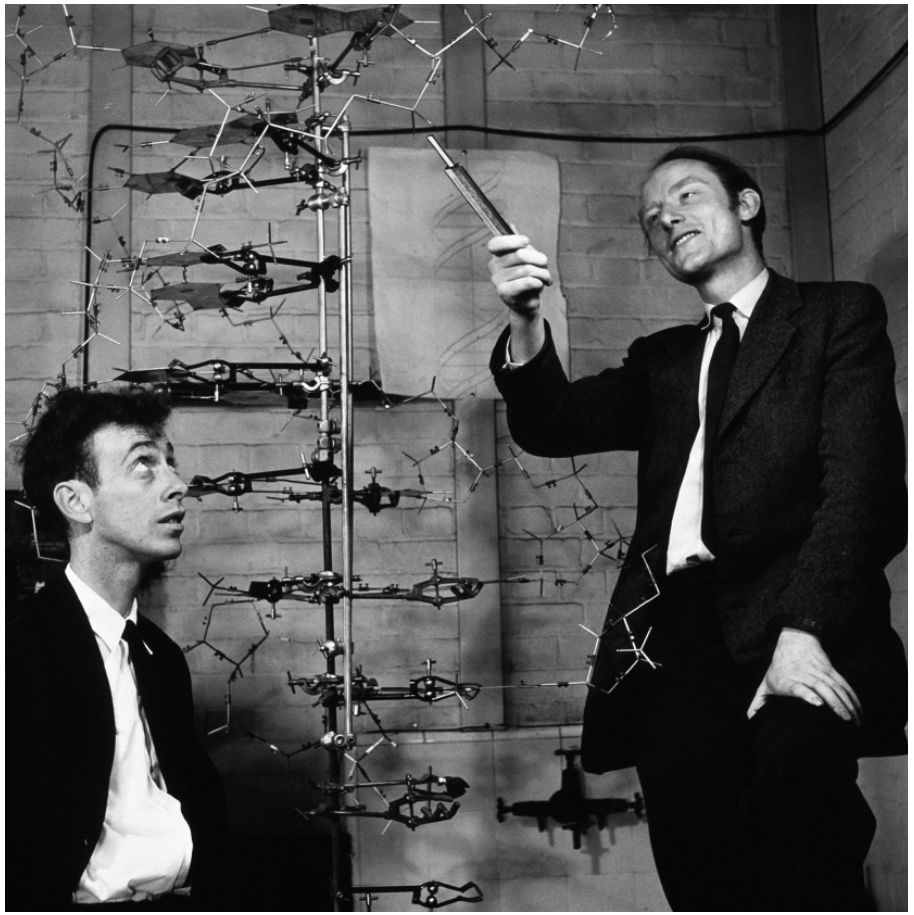
Biology primer

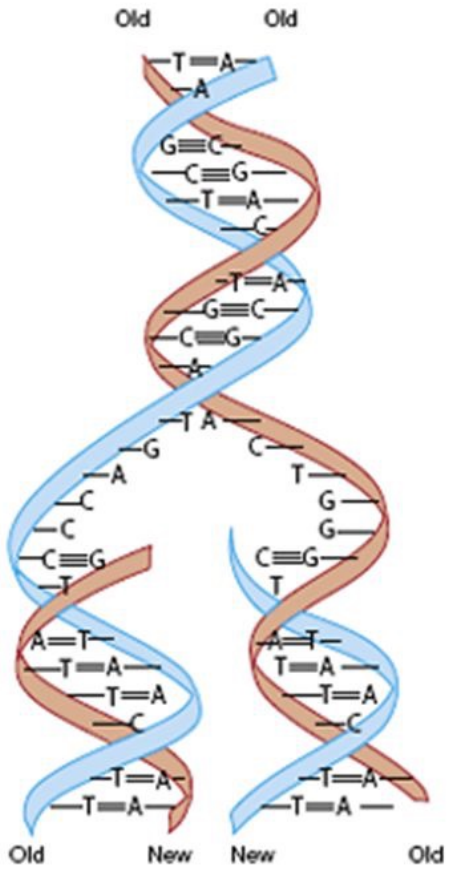
Quick introduction to molecular biology
and information transfer within the cell

CENTRAL DOGMA : DNA TO RNA TO PROTEIN



The Central Dogma of Molecular Biology





"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

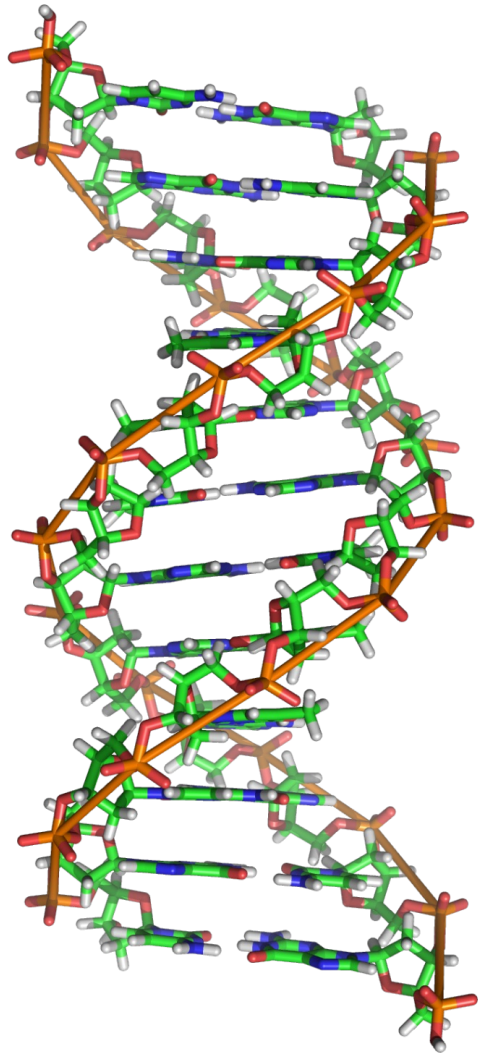
-Watson and Crick

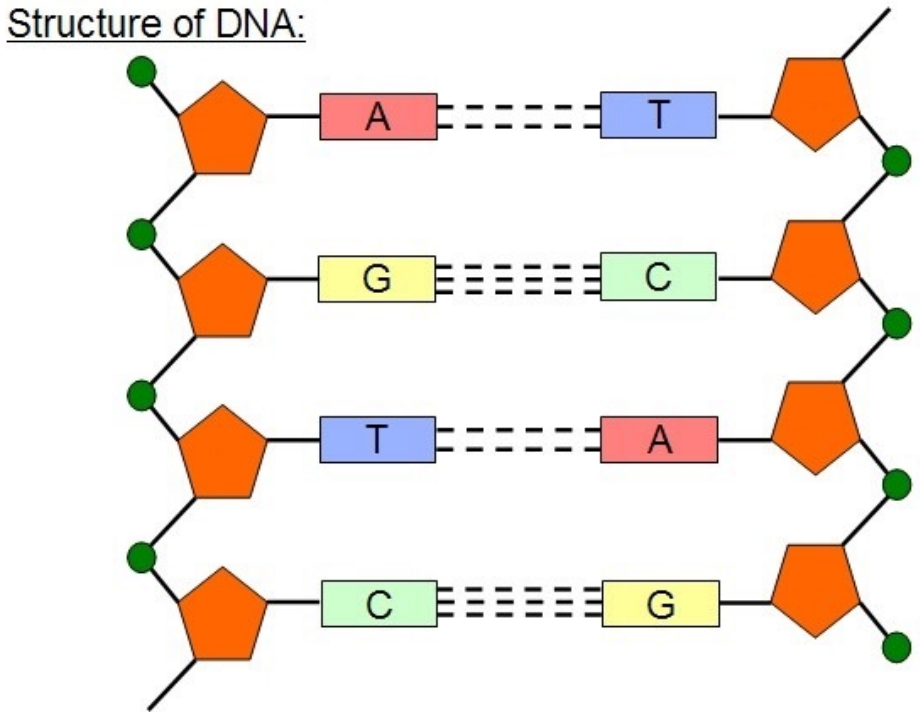
Cellular molecules:

1. DNA
2. RNA
3. Protein

DNA

Each strand composed of sequence of covalently bonded **nucleotides** (**bases**).





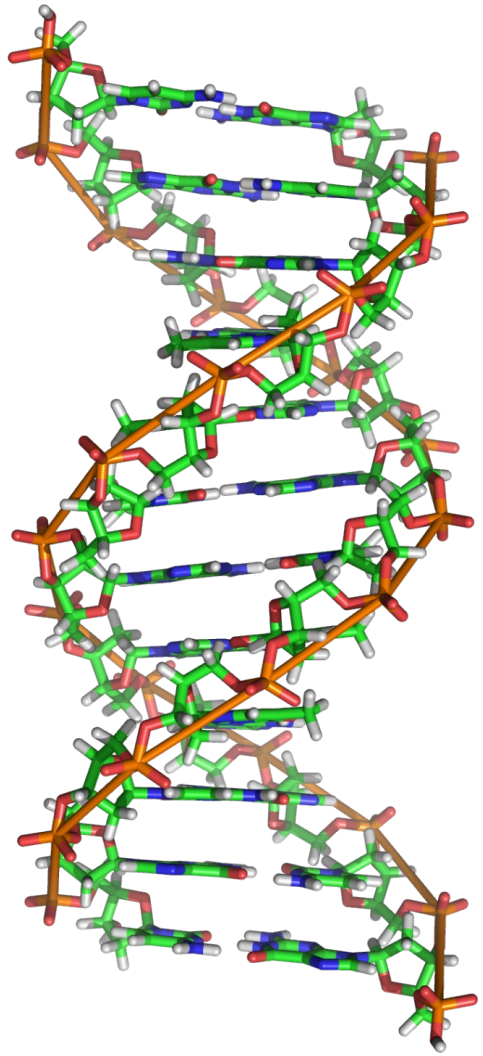
Four nucleotides:

A (adenine)
C (cytosine)
T (thymine)
G (guanine)

A←→T, C←→G Watson-Crick base-pairing

DNA

Each strand composed of sequence of covalently bonded **nucleotides** (**bases**).



5’ …ACGTGACTGAGGACCGTG… 3’

… | | | | | | | | | | | | | | | | | | …

3’ …TGCACTGACTCCTGGCAC… 5’

Pair of strings from 4-character alphabet

5’ …ACGTGACTGAGGACCGTG

CGACTGAGACTGACTGGGT

CTAGCTAGACTACGTTTTA

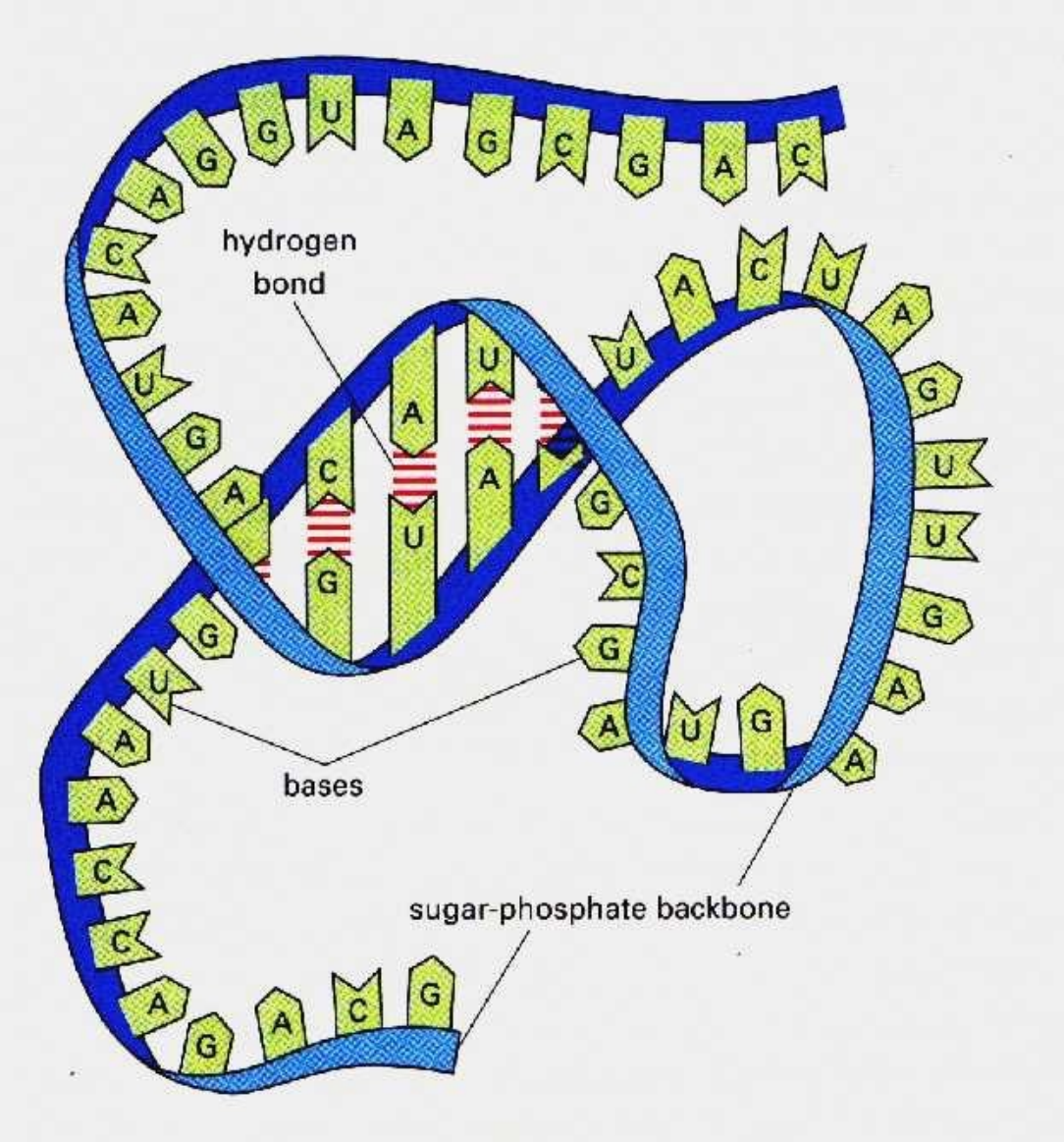
TATATATATACGTCGTCGT

ACTGATGACTAGATTACAG

TGATTTTAAAAAAATATT… 3’

Single string from 4-character alphabet

RNA



- **Single-stranded**
  - A (adenine)
  - C (cytosine)
  - U (uracil)
  - G (guanine)
- Can fold into **structures** due to base complementarity.

  A ← → U, C ← → G
- Comes in many flavors:

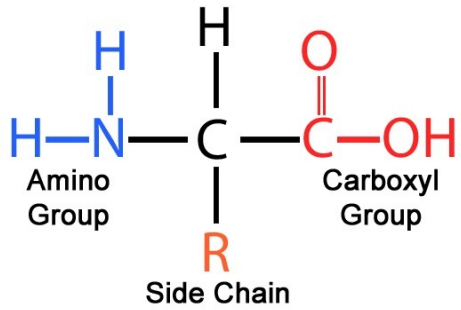
  mRNA, rRNA, tRNA, tmRNA, snRNA, snoRNA, scaRNA, aRNA, asRNA, piwiRNA, etc.

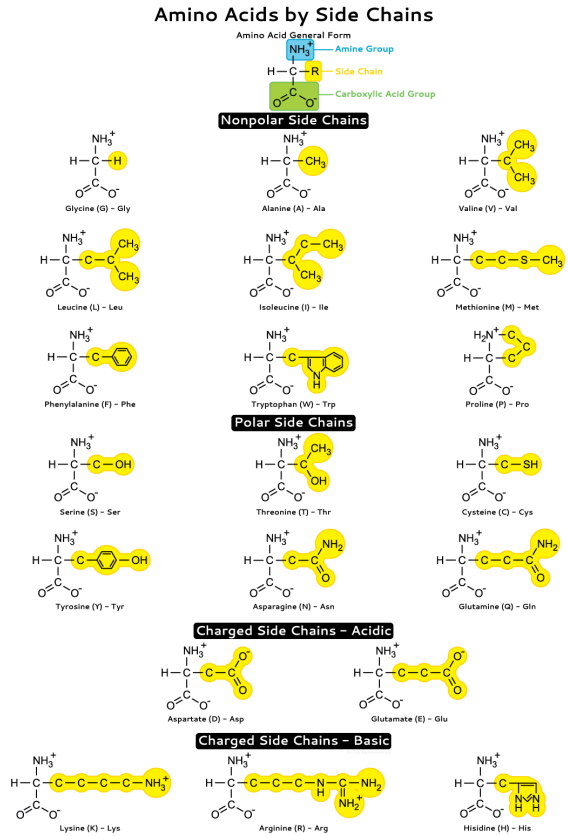
Protein

- String of amino acids: 20 letter alphabet

...DTIGDWNSPSFFGIQLVSSVHT
TLWYRENAFPVLGGFSWLSWFNW
HNMGYYYPVYHIGYPMIRCGTHL
VPMQFAFQSIARSFALVHWNAPM
VLKINPHERQDPVFWPCLYYSVD
IRSMHIGYPMIRCYQA...

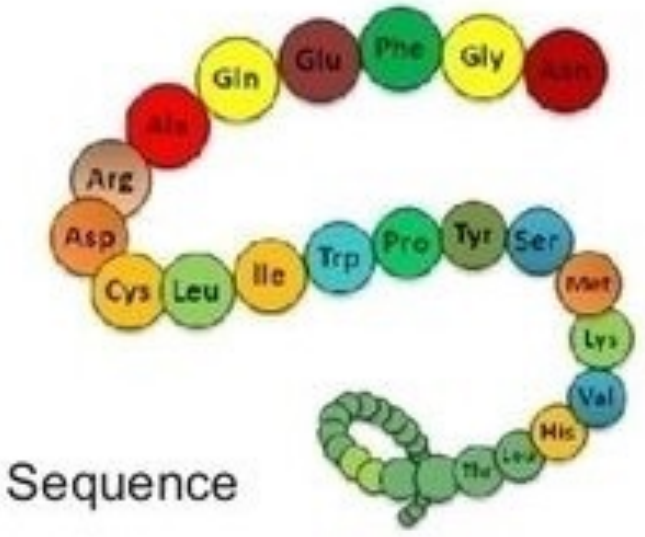
Amino Acid	3-Letters	1-Letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

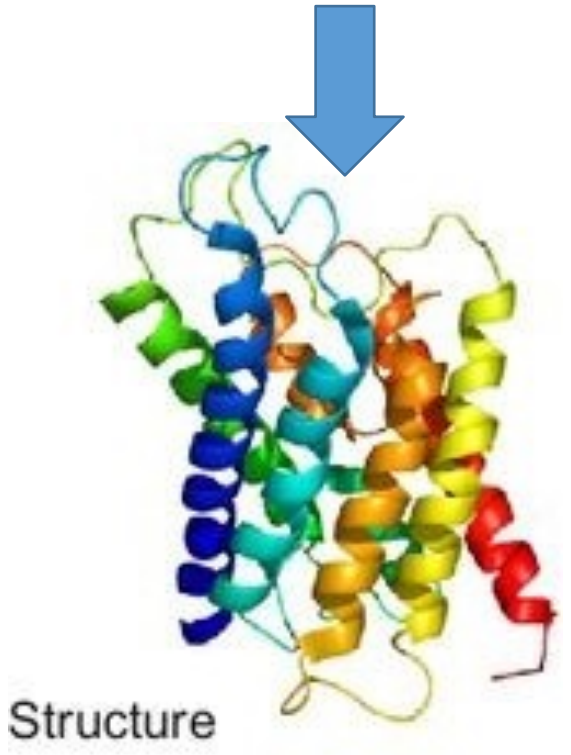




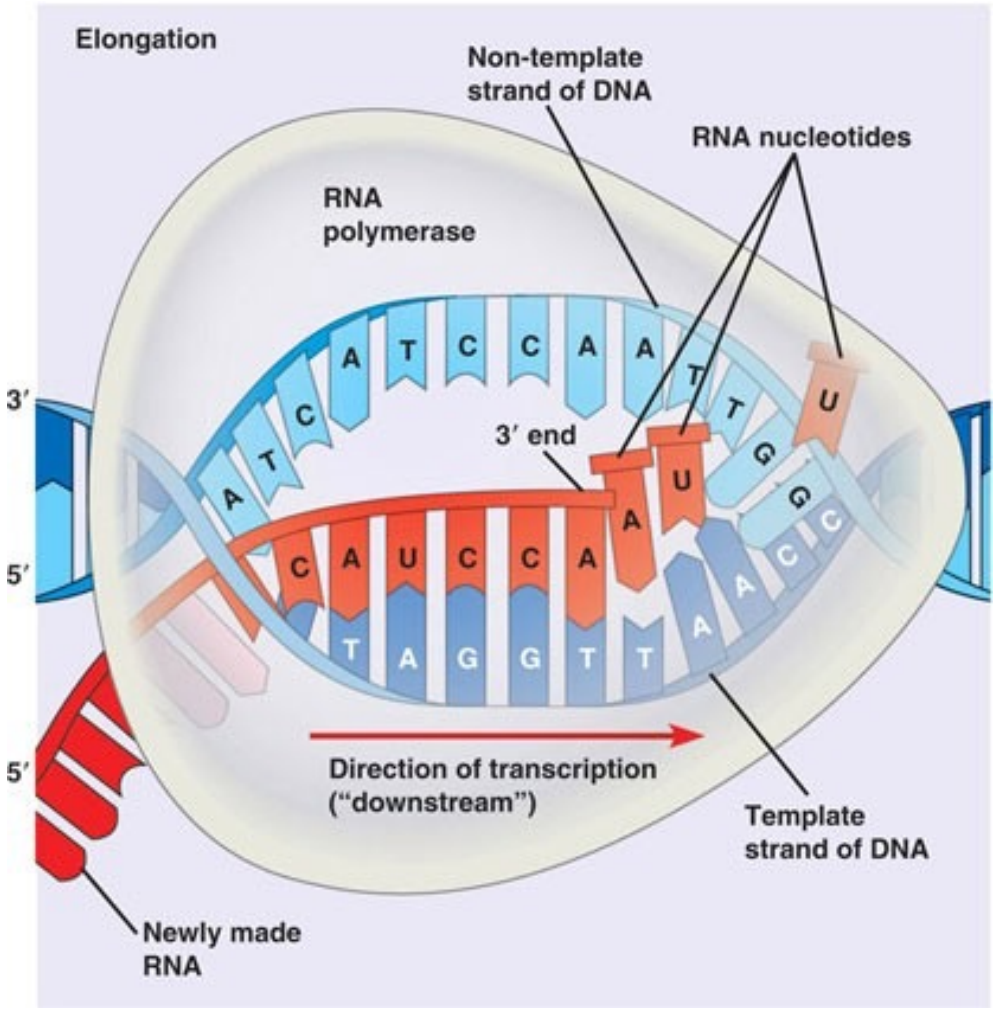
Protein

- String of amino acids: 20 letter alphabet
- Folds into 3D structures to perform various functions in cells

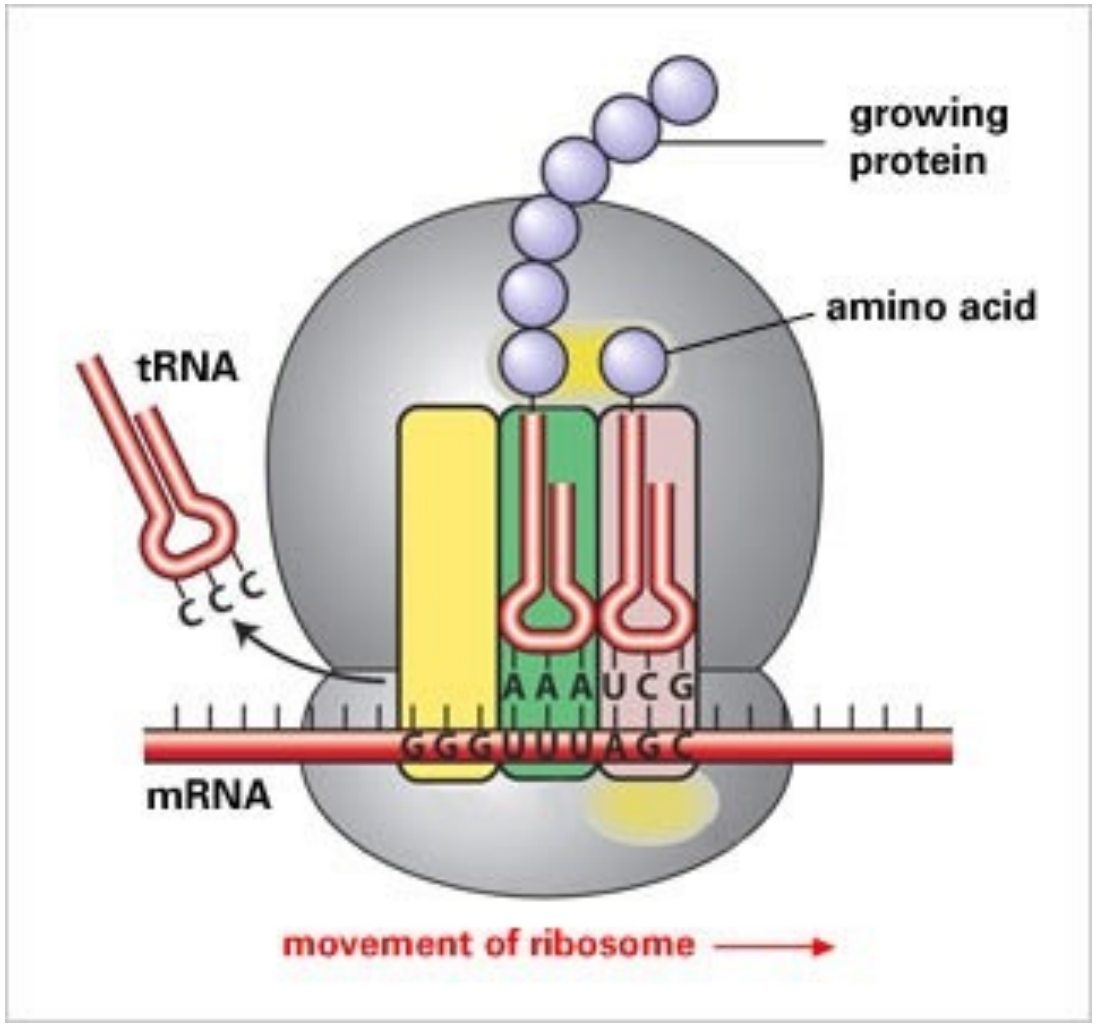




Transcription and Translation

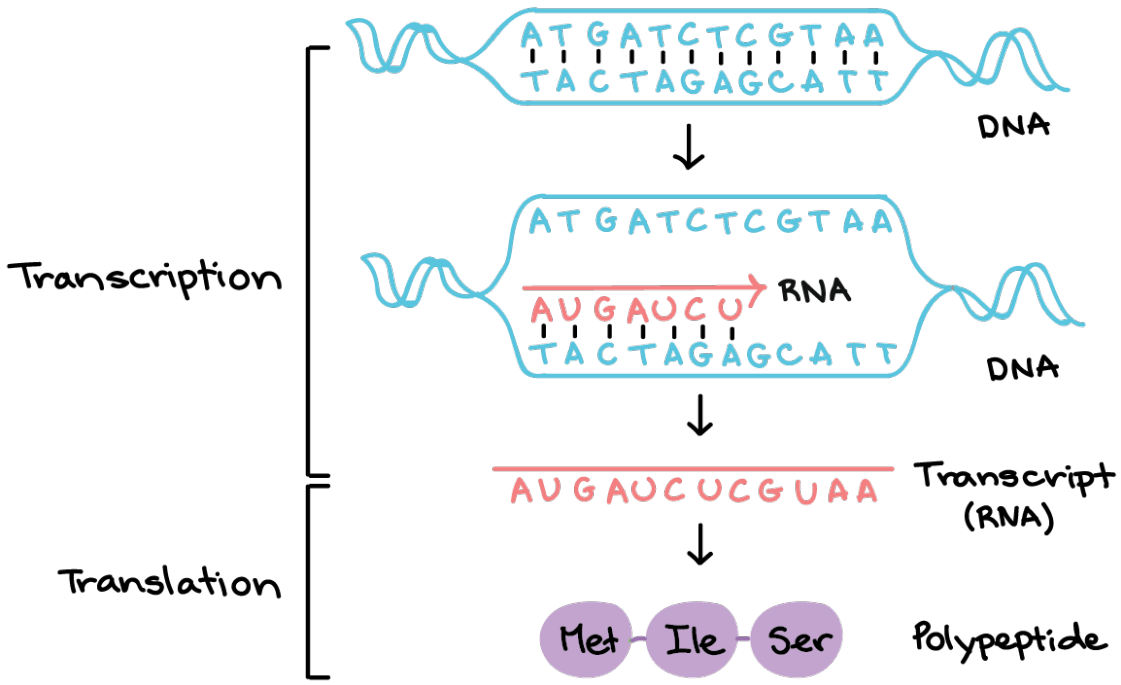


http://dna-rna.net/wp-content/uploads/2011/08/rna-transcription2.jpg



http://www.frontiers-in-genetics.org/en/pictures/translation_1.jpg

Transcription and Translation



https://www.khanacademy.org/science/biology/gene-expression-central-dogma/transcription-of-dna-into-rna/a/overview-of-transcription

Second base

First base	U	C	A	G	Third base
U	UUU, UUC Phenylalanine F	UCU, UCC, UCA, UCG Serine S	UAU, UAC Tyrosine Y	UGU, UGC Cysteine C	U
	UUA, UUG Leucine L		UAA, UAG Stop codon	UGA Stop codon / UGG Tryptophan W	A
					G
C	CUU, CUC, CUA, CUG Leucine L	CCU, CCC, CCA, CCG Proline P	CAU, CAC Histidine H	CGU, CGC, CGA, CGG Arginine R	U
			CAA, CAG Glutamine Q		C
					A
					G
A	AUU, AUC, AUA Isoleucine I	ACU, ACC, ACA, ACG Threonine T	AAU, AAC Asparagine N	AGU, AGC Serine S	U
	AUG Methionine M (start codon)		AAA, AAG Lysine K	AGA, AGG Arginine R	C
					A
					G
G	GUU, GUC, GUA, GUG Valine V	GCU, GCC, GCA, GCG Alanine A	GAU, GAC Aspartic acid D	GGU, GGC, GGA, GGG Glycine G	U
			GAA, GAG Glutamic acid E		C
					A
					G

http://bioinfo.bisr.res.in/project/crat/pictures/codon.jpg

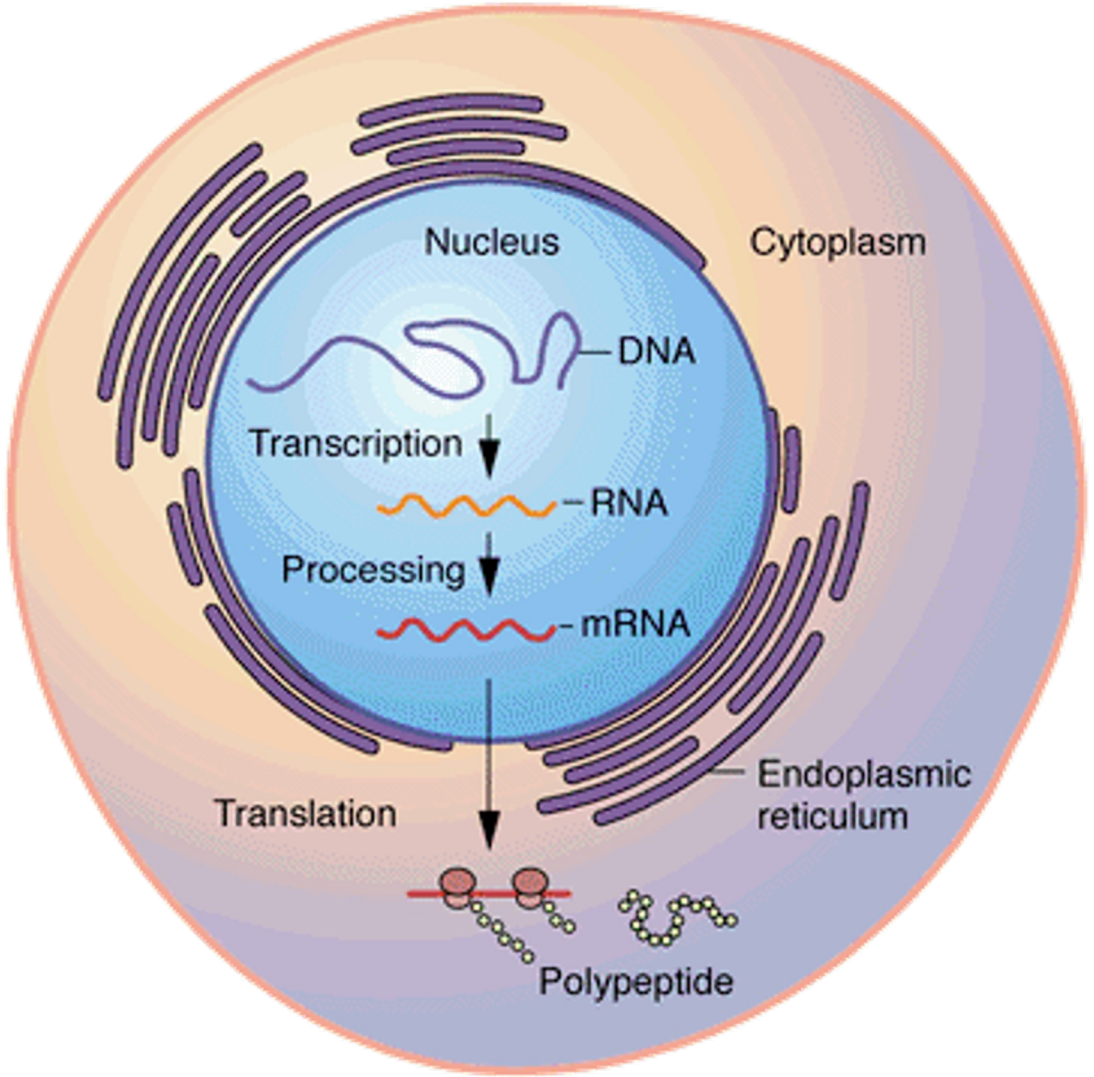
Primer on Molecular Biology

Three fundamental molecules:

1. **DNA**
   Information storage.

2. **RNA**
   Old view: Mostly a "messenger".
   New view: Performs many important functions.

3. **Protein**
   Perform most cellular functions (biochemistry, signaling, control, etc.)

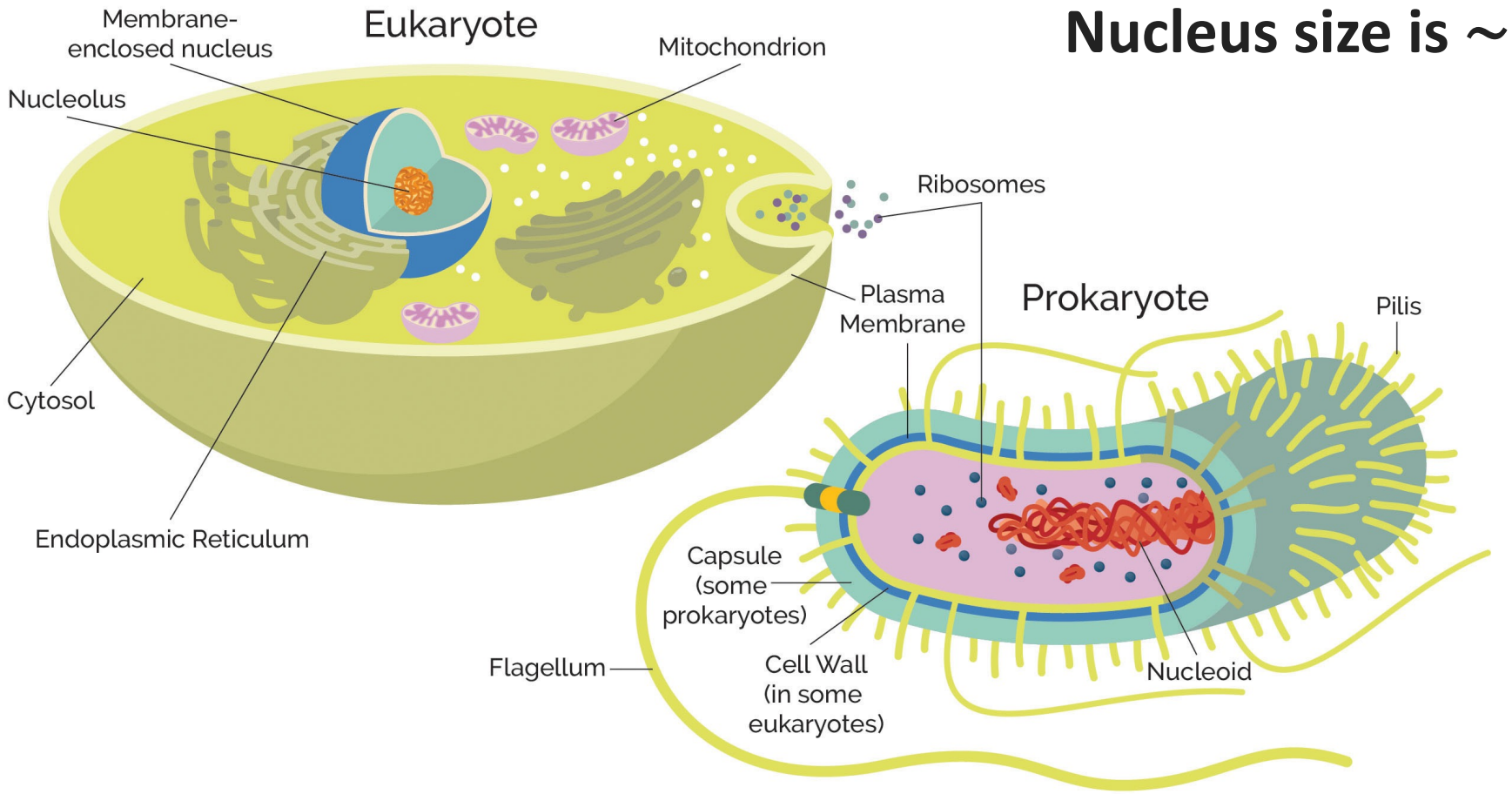


Genomes size

Species	*T2 phage*	*Escherichia coli*	*Drosophila melanogaster*	*Homo sapiens*	*Paris japonica*
Genome Size	170,000 bp	4.6 million bp	130 million bp	3.2 billion bp	150 billion bp
Common Name	 Virus	 Bacteria	 Fruit fly	 Human	 Canopy Plant

Streched out – human genome is 2.13 meters long

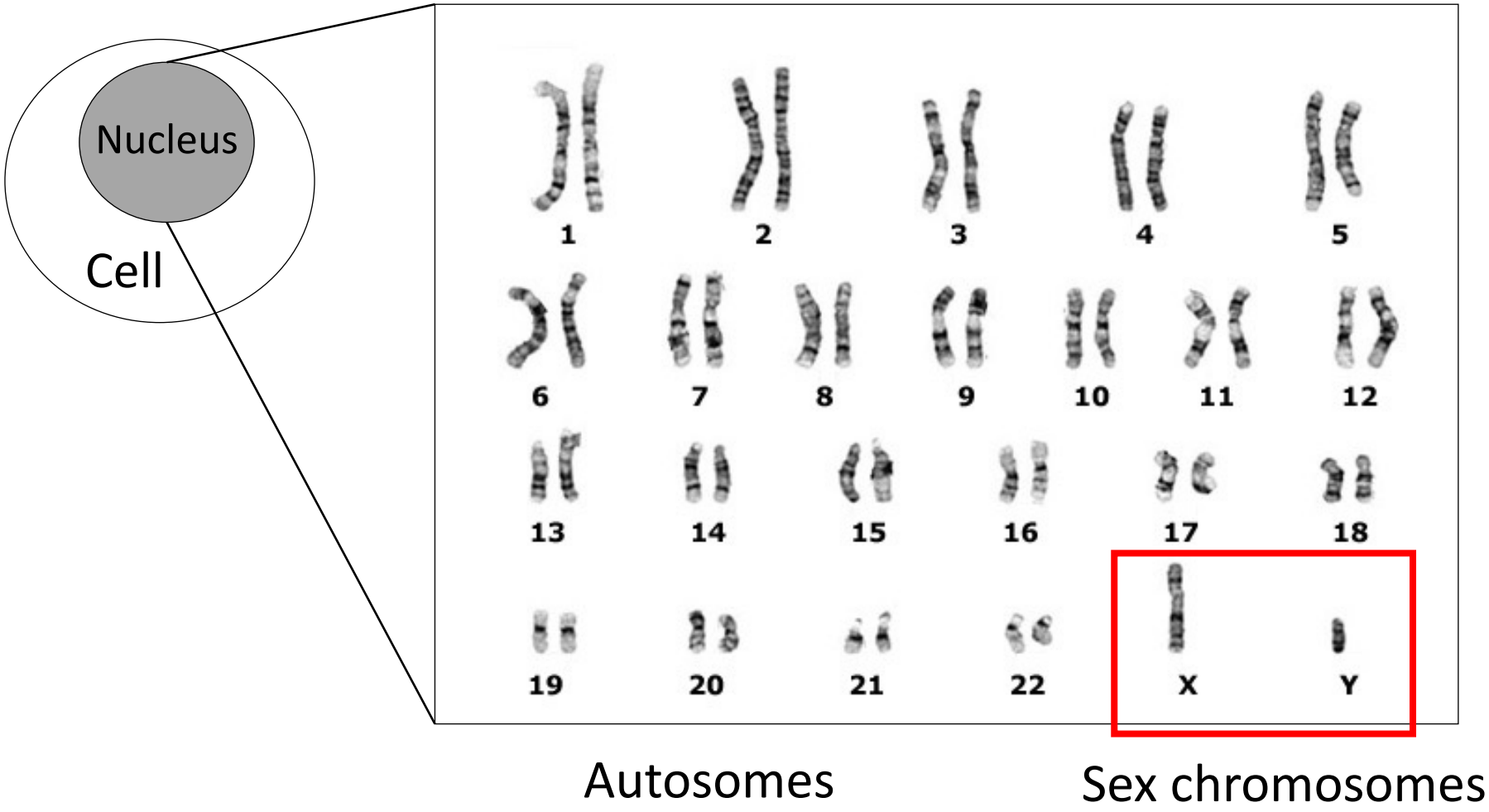
Chromosomes inside the cell



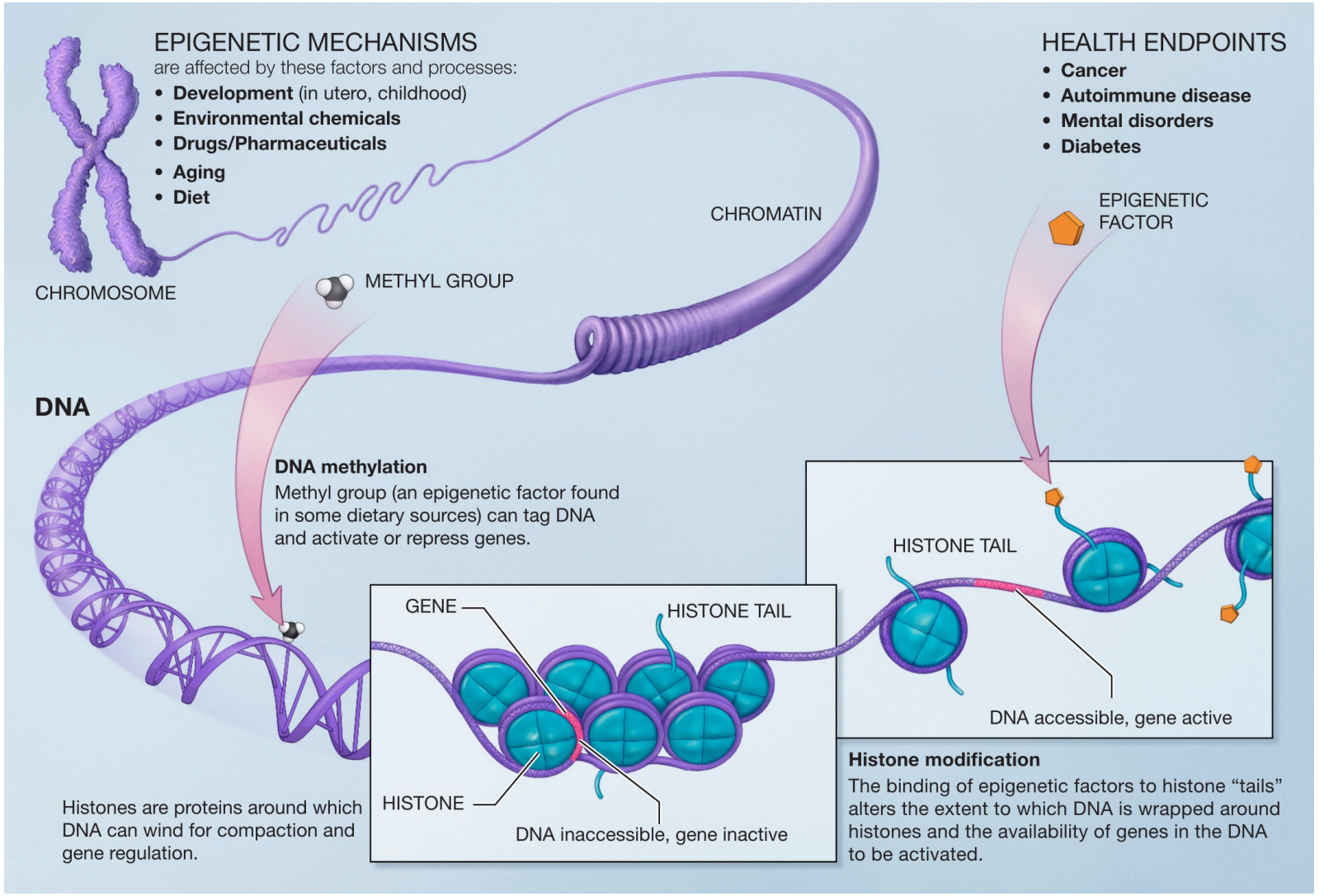
Nucleus size is ~6 µm

30,000,000,000,000 in human body

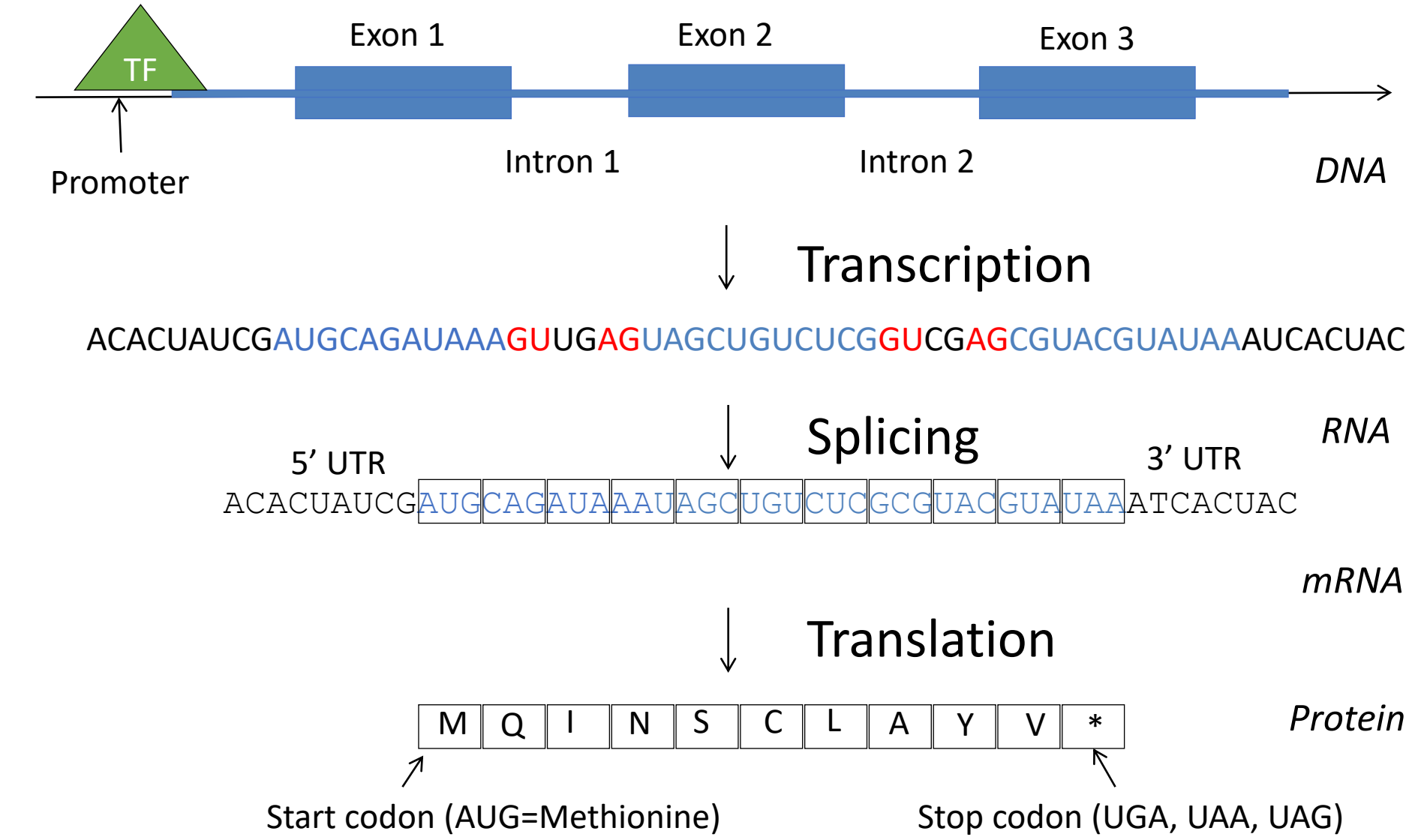
Bird's eye view of the human genome



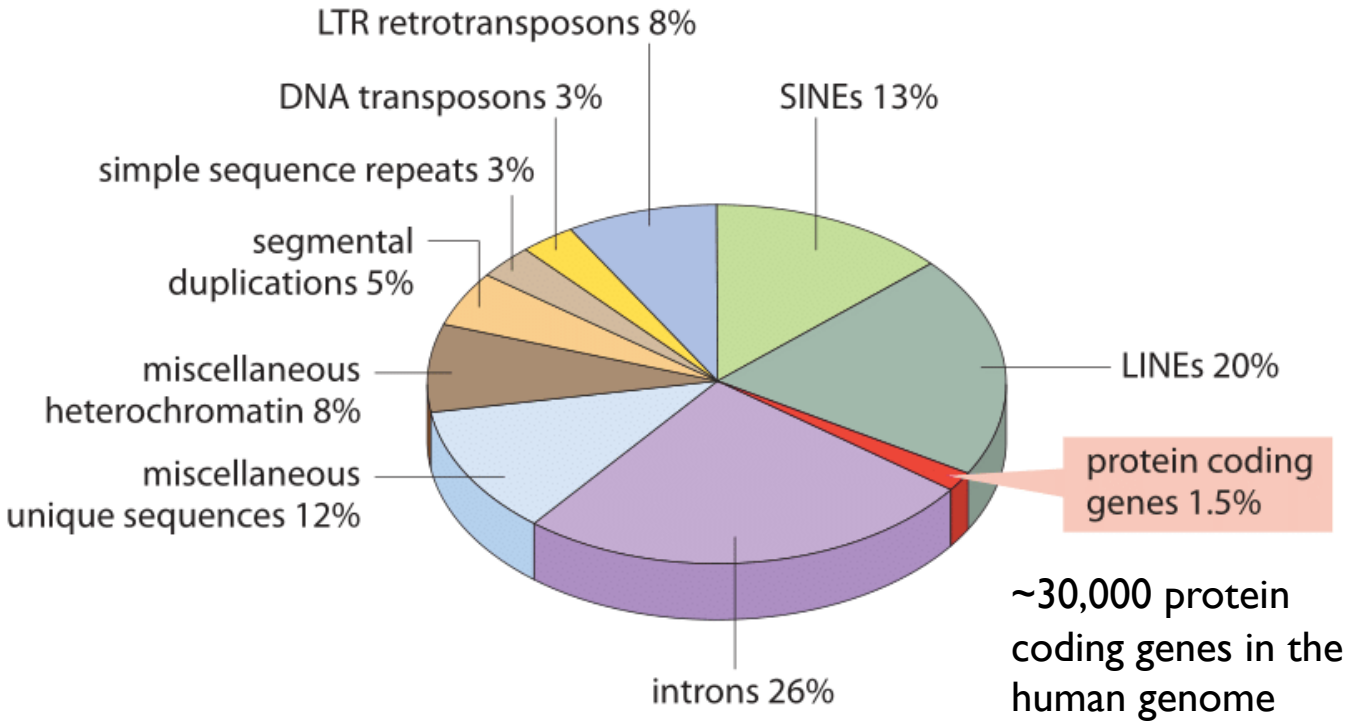
Diverse epigenetic modifications



The structure of a gene



Organization of the human genome

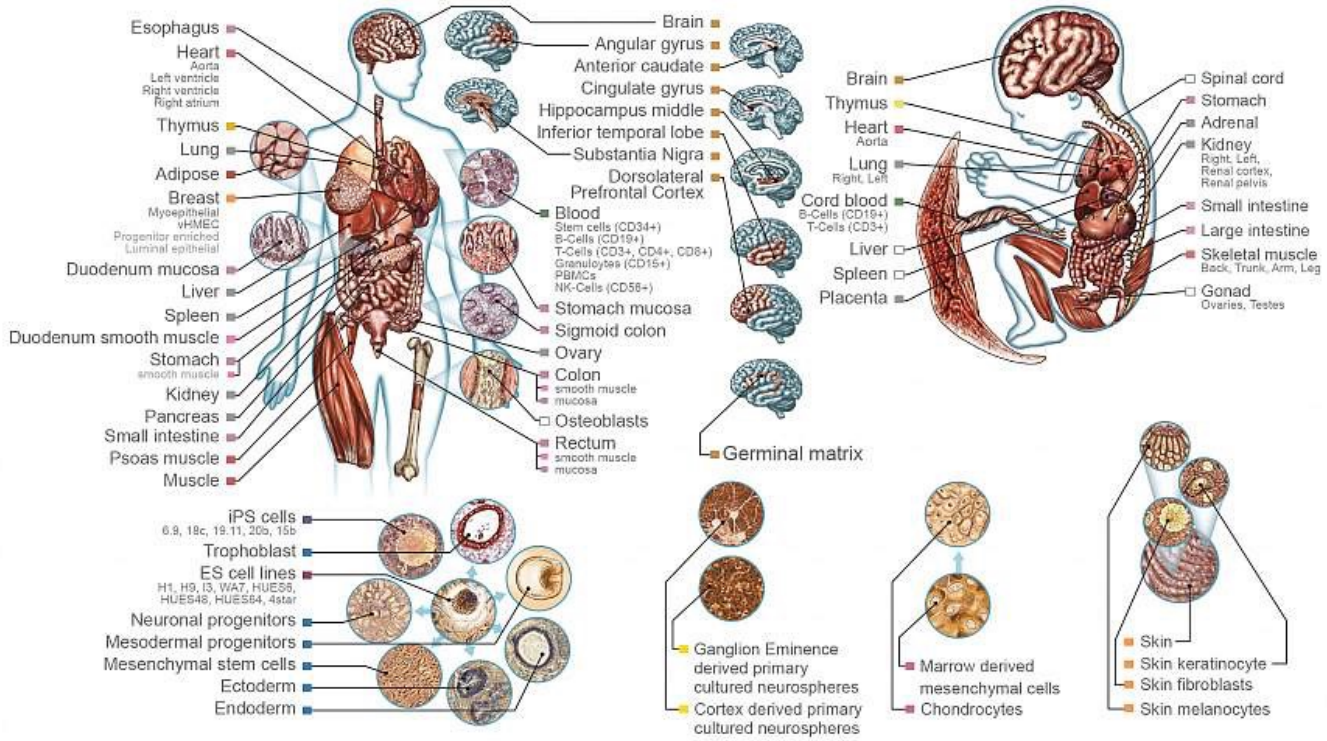


~30,000 protein coding genes in the human genome

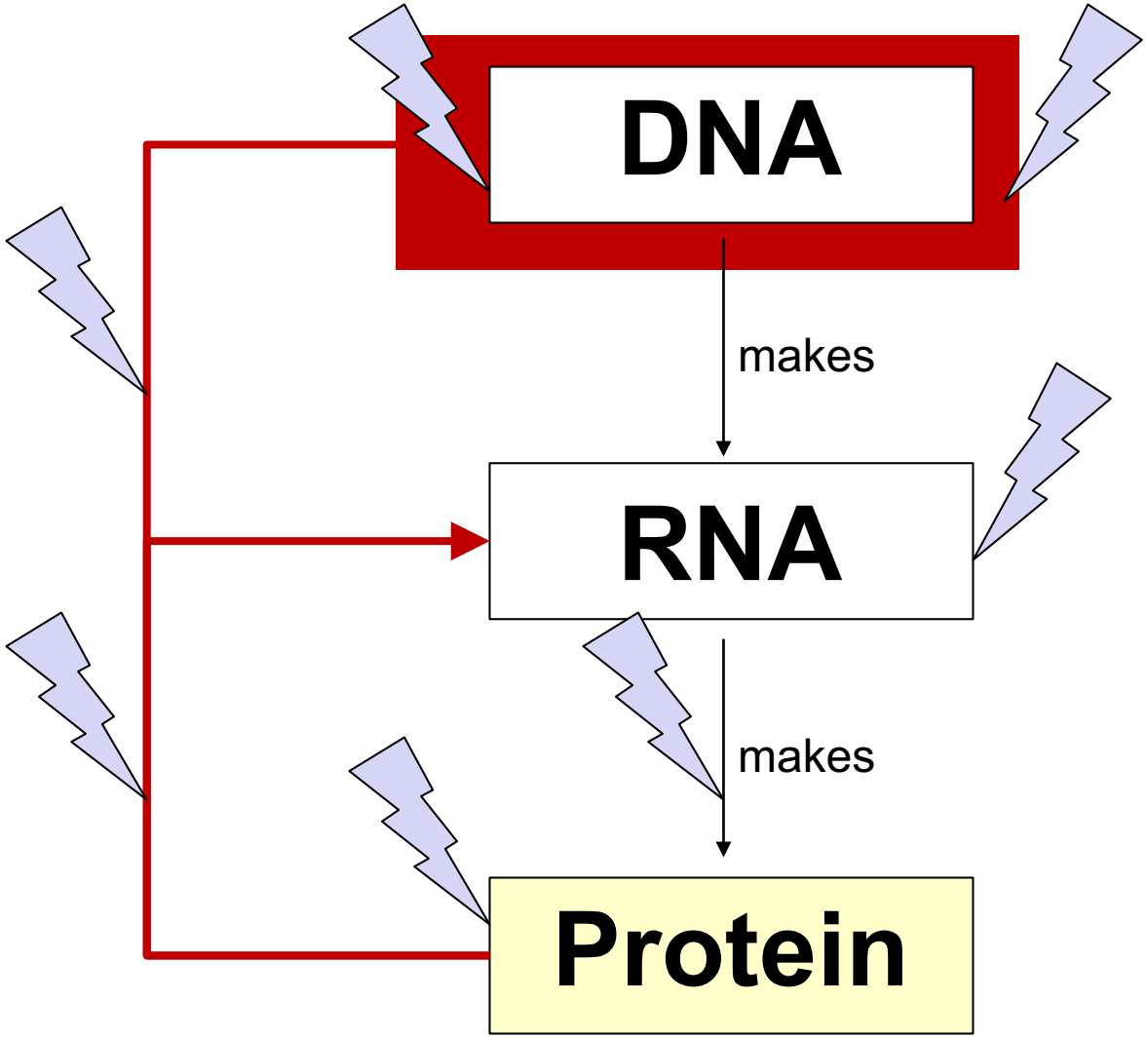
http://book.bionumbers.org/how-many-genes-are-in-a-genome/

One genome ⇔ Many cell types

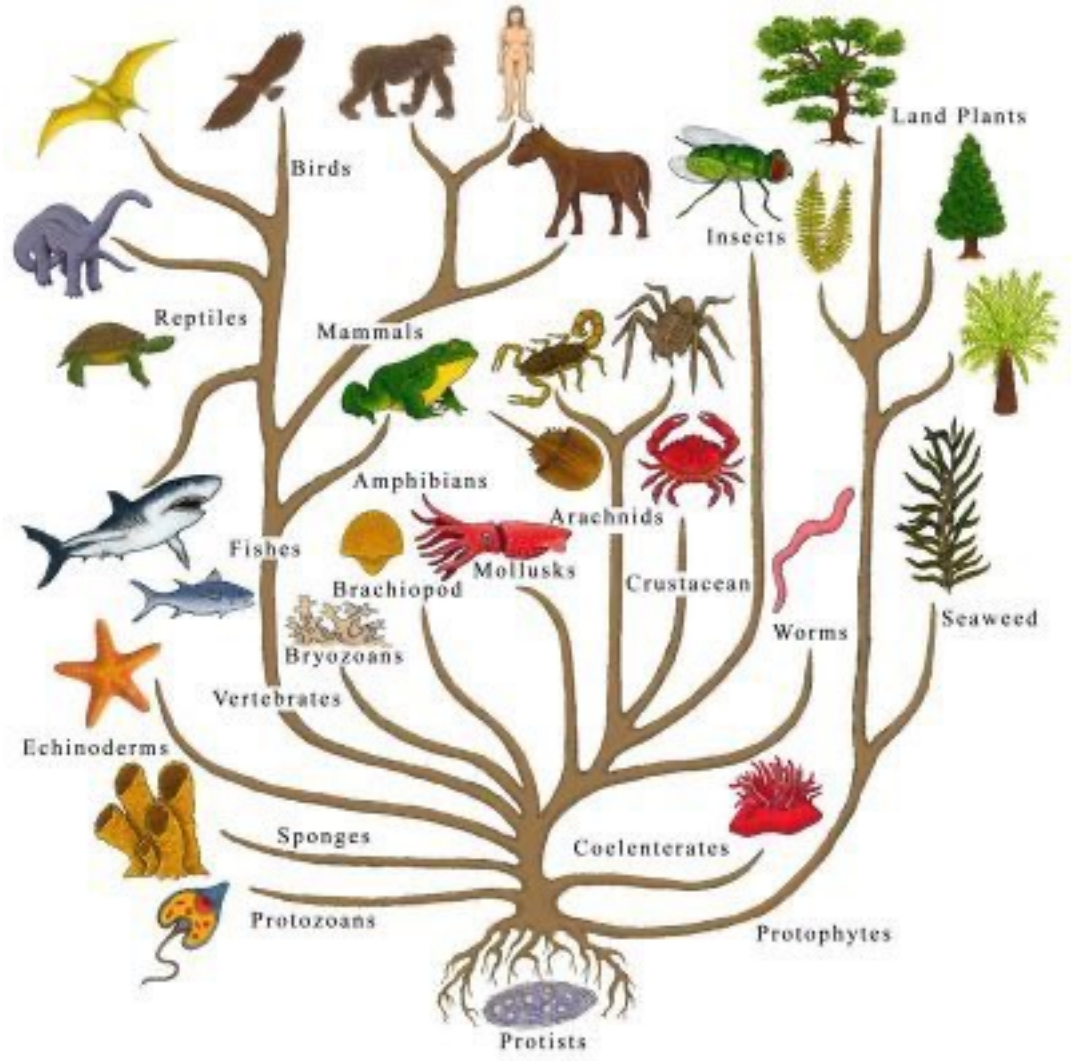
ACCAGTTACGACGGTCA
GGGTACTGATACCCCAA
ACCGTTGACCGCATTTA
CAGACGGGGTTTGGGTT
TTGCCCCACACAGGTAC
GTTAGCTACTGGTTTAG
CAATTTACCGTTACAAC
GTTTACAGGGTTACGGT
TGGGATTTGAAAAAAAG
TTTGAGTTGGTTTTTTC
ACGGTAGAACGTACCGT
TACCAGTA

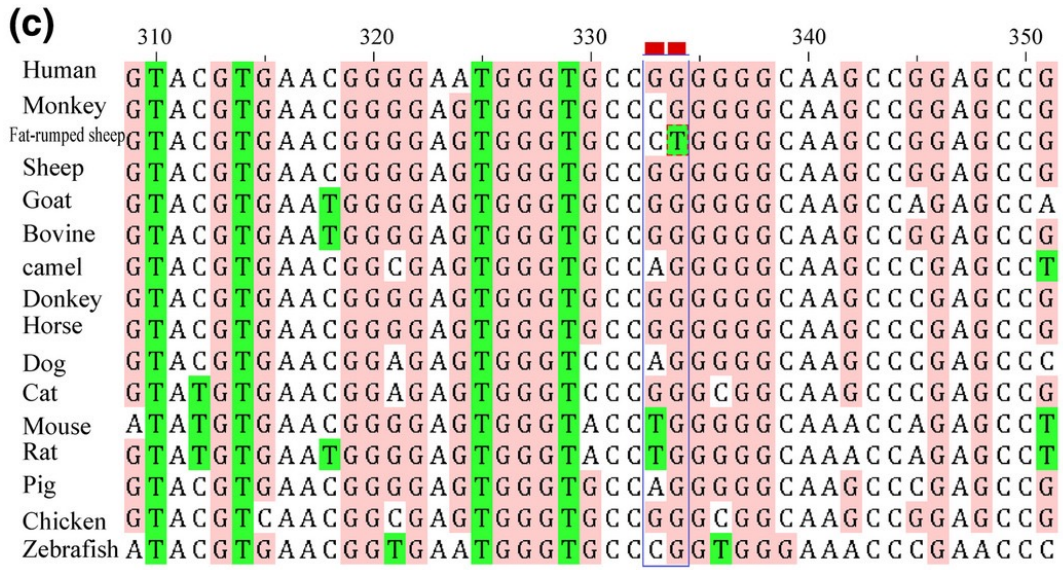


The role of genetic alterations

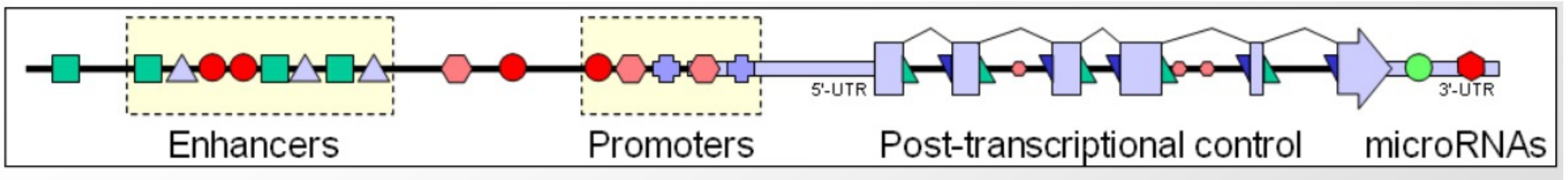


Phylogenetic tree of life





The components of genomes and gene regulation



Goal: A systems-level understanding of genomes and gene regulation:

- The genome: Map reads, align genes/genomes, assembly strategies
- The genes: Protein-coding exons, introns, non-coding RNA, RNA folding
- The control regions: Promoters, enhancers, insulators, chromatin states
- The actual words: Regulatory motifs, high-resolution accessibility maps
- The regulators: Transcription factors, chromatin modifiers, nucleosomes
- The dynamics: Changing maps between cell types, across development
- The networks: regulator→enhancer→target, ChIP-seq, correlated activity
- The grammars: TF/motif/mark combinations, predictive models
- Human variation: Human diversity, population genomics, linkage maps
- Evolution: Phylogenetics, phylogenomics, coalescent, human ancestry
- GWAS/QTLs: Genome variation ⇔organismal/molecular phenotypes
- Disease: Personal (epi)genomics, pharmacogenomics, synthetic biology

Donald Knuth

Professor emeritus of Computer Science
Stanford University
Turing Award winner
"father of the analysis of algorithms."

"I can't be as confident about computer science as I can about biology. Biology easily has 500 years of exciting problems to work on. It's at that level."