

Introduction to Bioinformatics

Saeedeh Akbari & Niloufar Razani

Department of Computer Engineering

Sharif University of Technology

Fall 2023

Outline

- **Course Information**
- **Introduction to bioinformatics**
- **Course Overview**

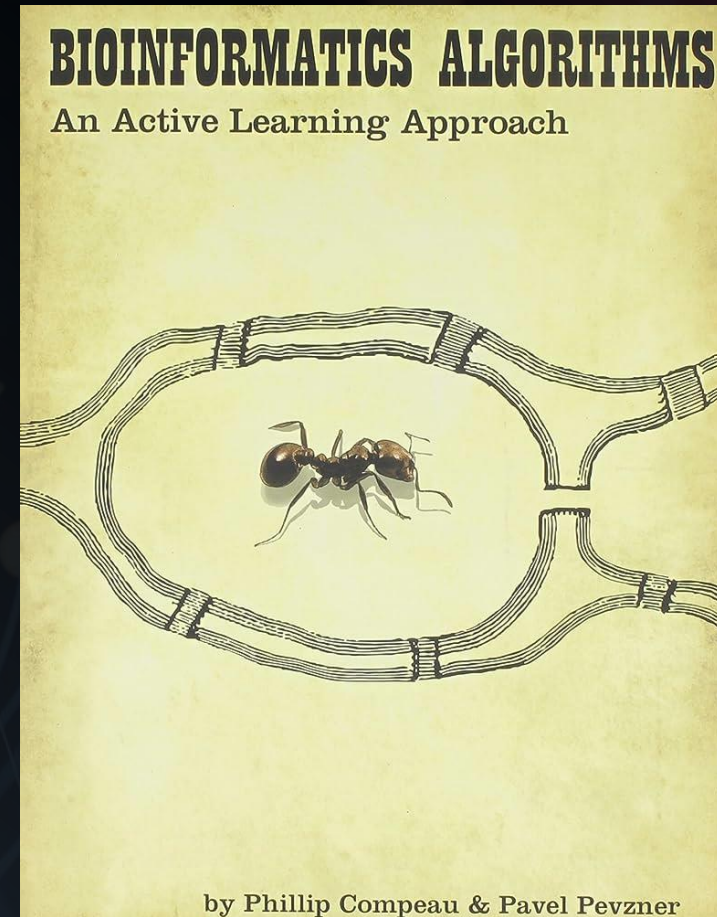
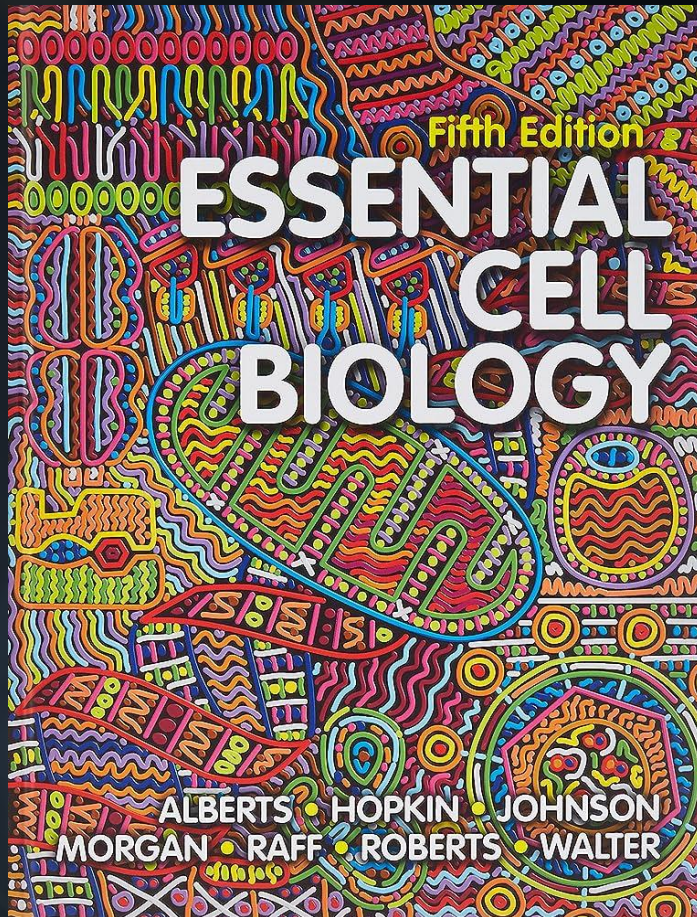


Course Information

Course Info

- Instructors : Saeedeh Akbari
(saeedeh.akbarira@gmail.com)
Niloufar Razani
(razani.niloufar@gmail.com)
- Lectures : Sat & Mon (15 – 16:30) – Class 102
- Head TA : Mohammad Asadi
(asadimohammad331@gmail.com)
- Discussion and Homework : Quera

References



Grading

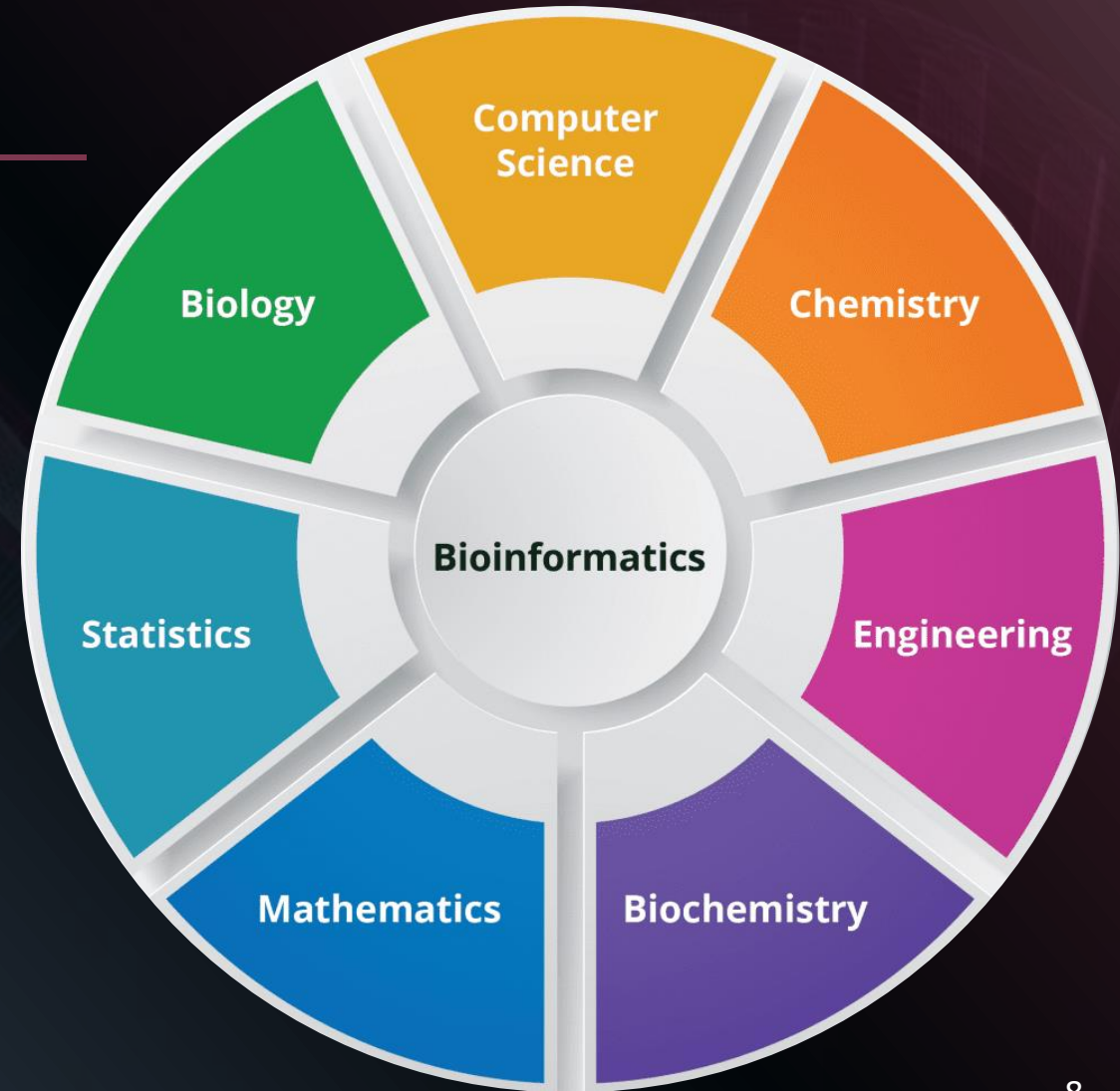
- **Midterm:** 4
- **Final Exam:** 6
- **Assignments (4):** 6
- **Project (2 phases):** 4 + 1.5
- **Quizzes:** +0.5



Introduction to Bioinformatics!

What is Bioinformatics?

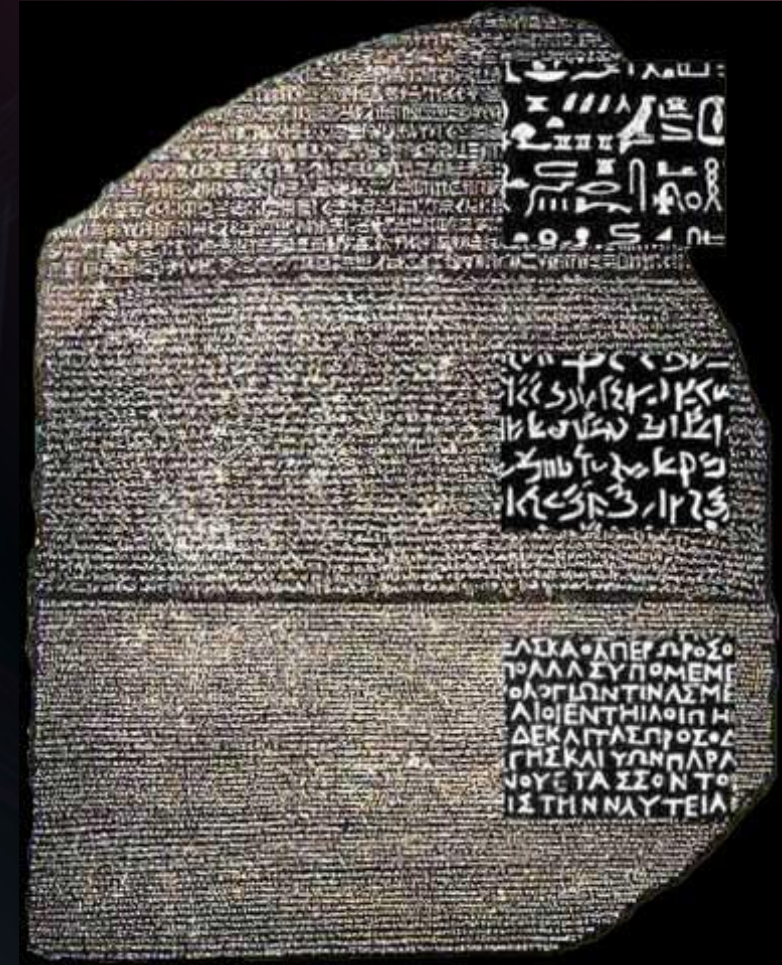
- Bioinformatics is “research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral, or health data, including those to acquire, store, organize, analyze, or visualize such data.”
 - National Institutes of Health (NIH)



History

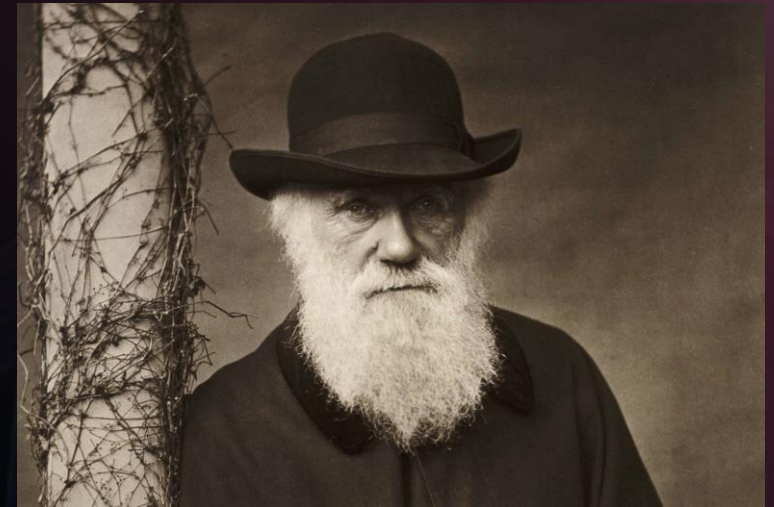
- Rosetta Stone is an inscription stone from ancient Egypt, which has been a useful tool in understanding of hieroglyphic writing.
-
- The top and middle texts are in Ancient Egyptian using hieroglyphic and Demotic scripts respectively, while the bottom is in Ancient Greek.
- All three parts of the inscription relate to the same subject, and the Greek script was the key to understanding the other two parts.

DNA is the Rosetta Stone of Biology!

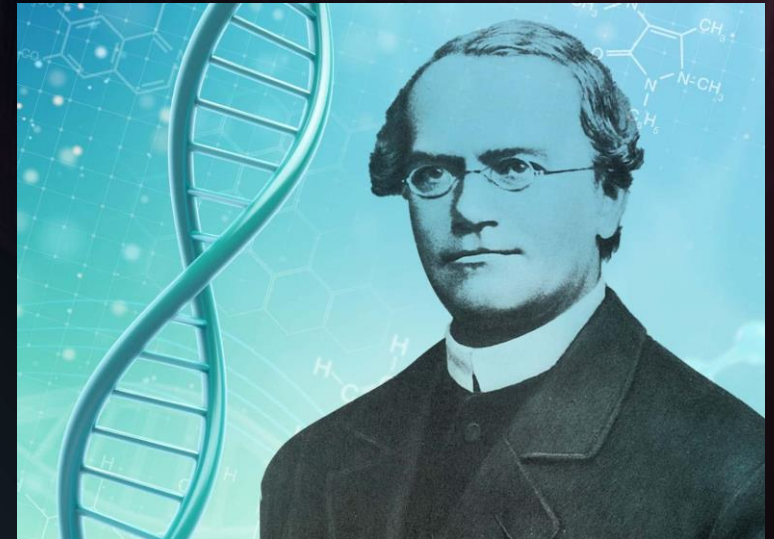


History

- Developing the theory of evolution by natural selection. This explained for the first time how species evolve and adapt through natural selection acting on variation in traits.
- Conducted hybridization experiments in pea plants over many generations. This allowed him to uncover the basic mechanisms of heredity.
- Discovered that traits are inherited as discrete units, which he called "factors" (now known as genes).



Charles Darwin (1809-1882)



Gregor Mendel (1822-1884)

History

- In the 1880s, Walther Flemming observed thread-like structures in cell nuclei during division, which he named chromosomes (colored bodies). He speculated chromosomes carried hereditary material.



Wilhelm Johannsen & William Bateson



Walther Flemming

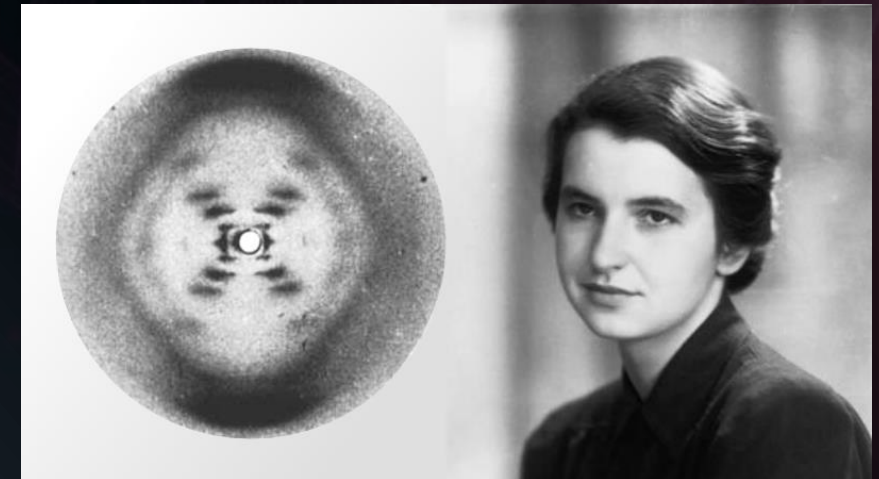
- The discovery of the results of Mendel's research by William Bateson in 1900 made her confident about the results of Mendel's experiments.
- He Coined the term 'genetics' in 1906 to describe the study of biological inheritance and variation.
- Wilhelm Johannsen Coined the word 'gene' in 1909 to describe the fundamental physical and functional units of heredity and inheritance

History

- Thomas Morgan's work with fruit flies (*Drosophila melanogaster*) linked specific genes to specific chromosomes. This suggested chromosomes carried the genetic material.
- In the 1920s, experiments by Nikolai Koltsov and Phoebus Levene revealed DNA was a large molecule with a phosphate-sugar backbone and four different bases (A, T, G, C).
- In the 1940s, Oswald Avery, Colin MacLeod and Maclyn McCarty showed DNA extracted from bacteria could transform harmless bacteria into disease-causing bacteria. This first demonstrated DNA carried genetic information.
- In the 1950, Rosalind Franklin conducted X-ray diffraction studies on DNA, providing critical insights into its helical structure.



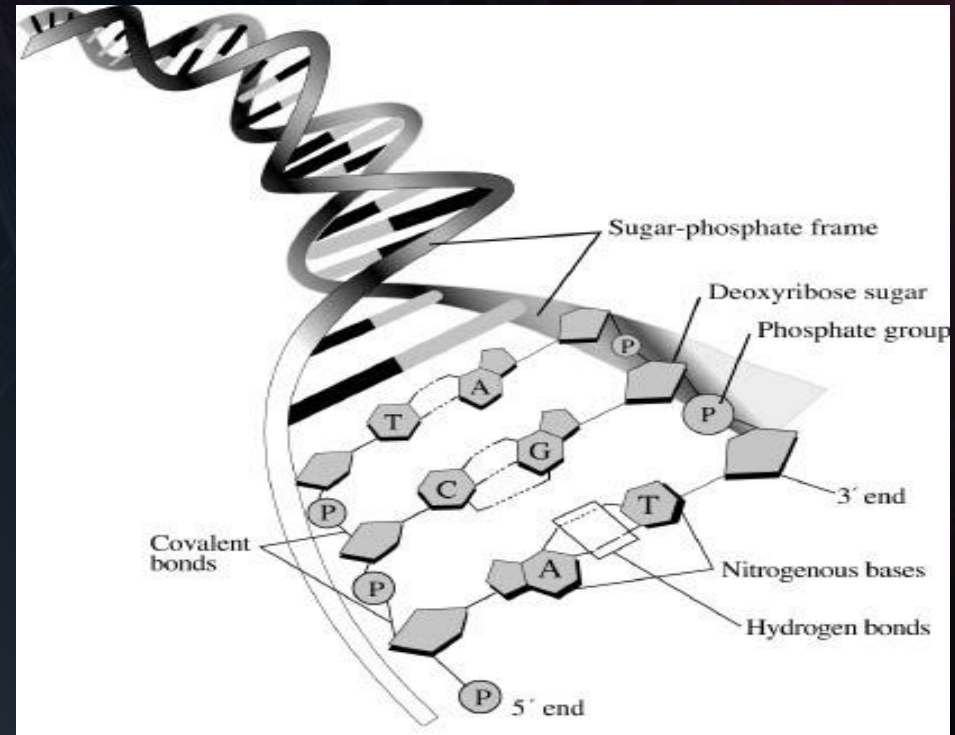
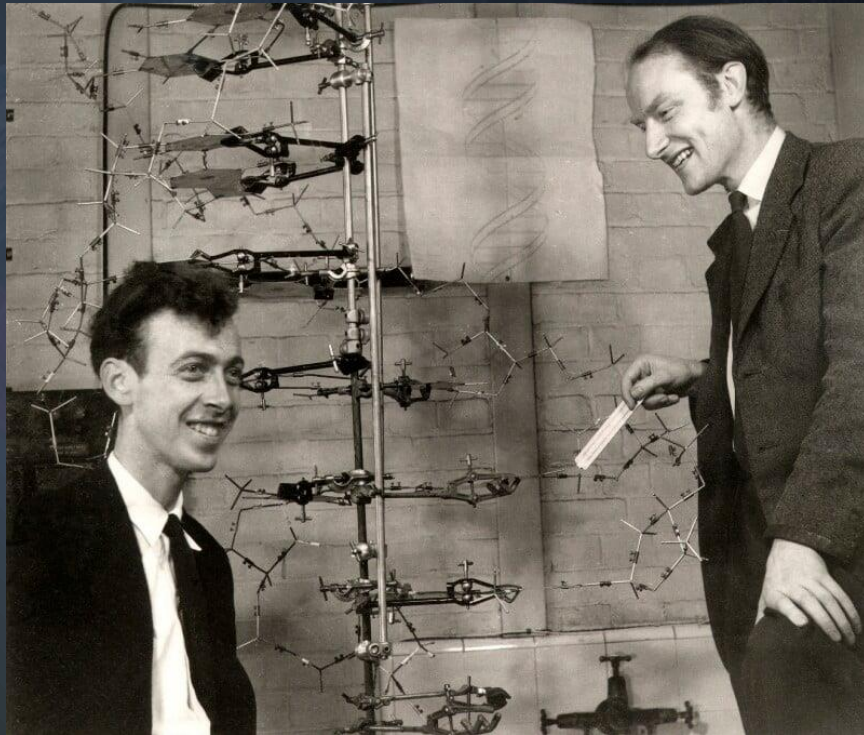
Thomas Morgan



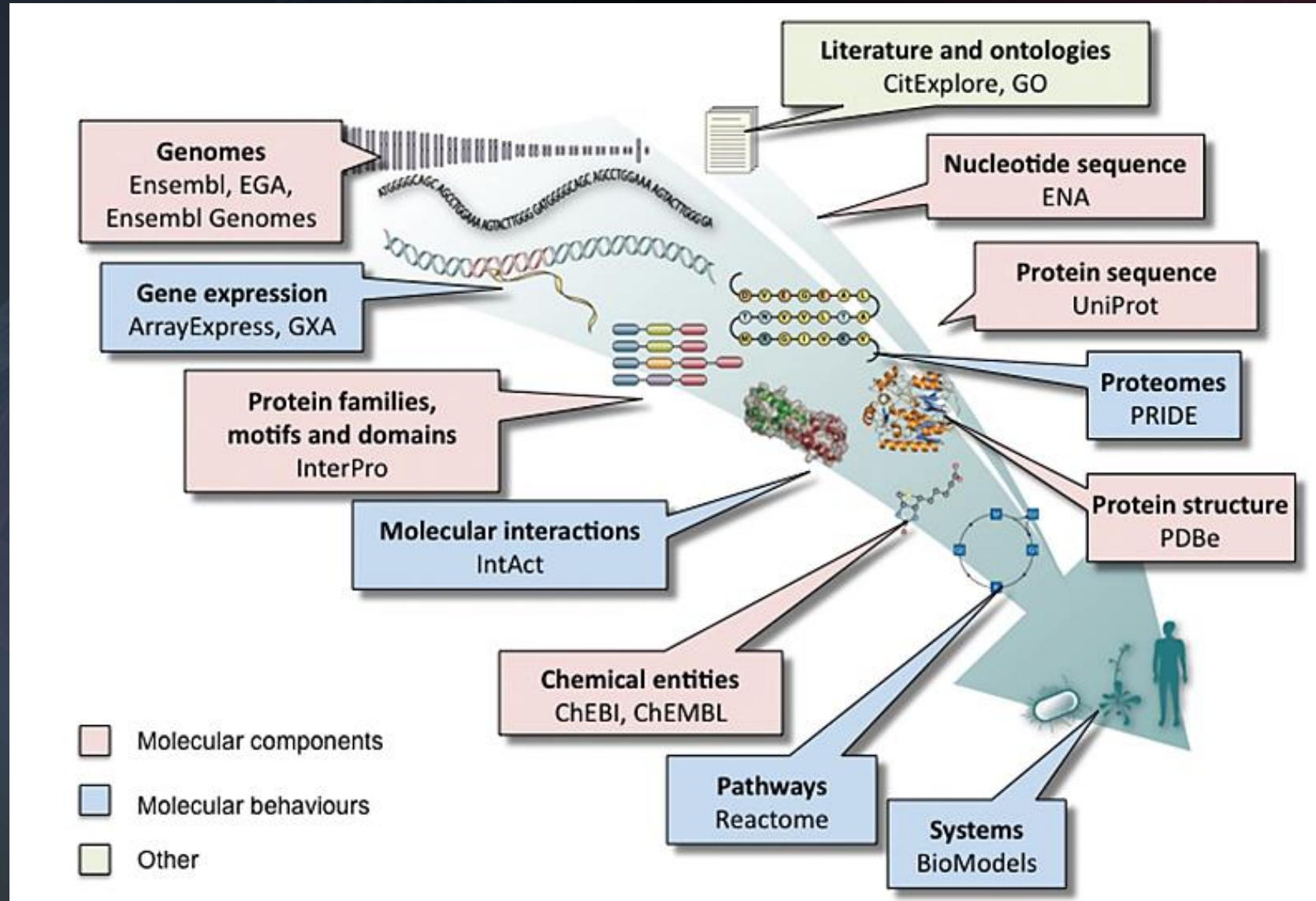
Rosalind Franklin

History

- In 1953, James Watson and Francis Crick determined the 3D double helix structure of DNA.
- Watson and Crick were the first to uncover DNA's molecular structure, which revealed how it could encode and transmit genetic information.

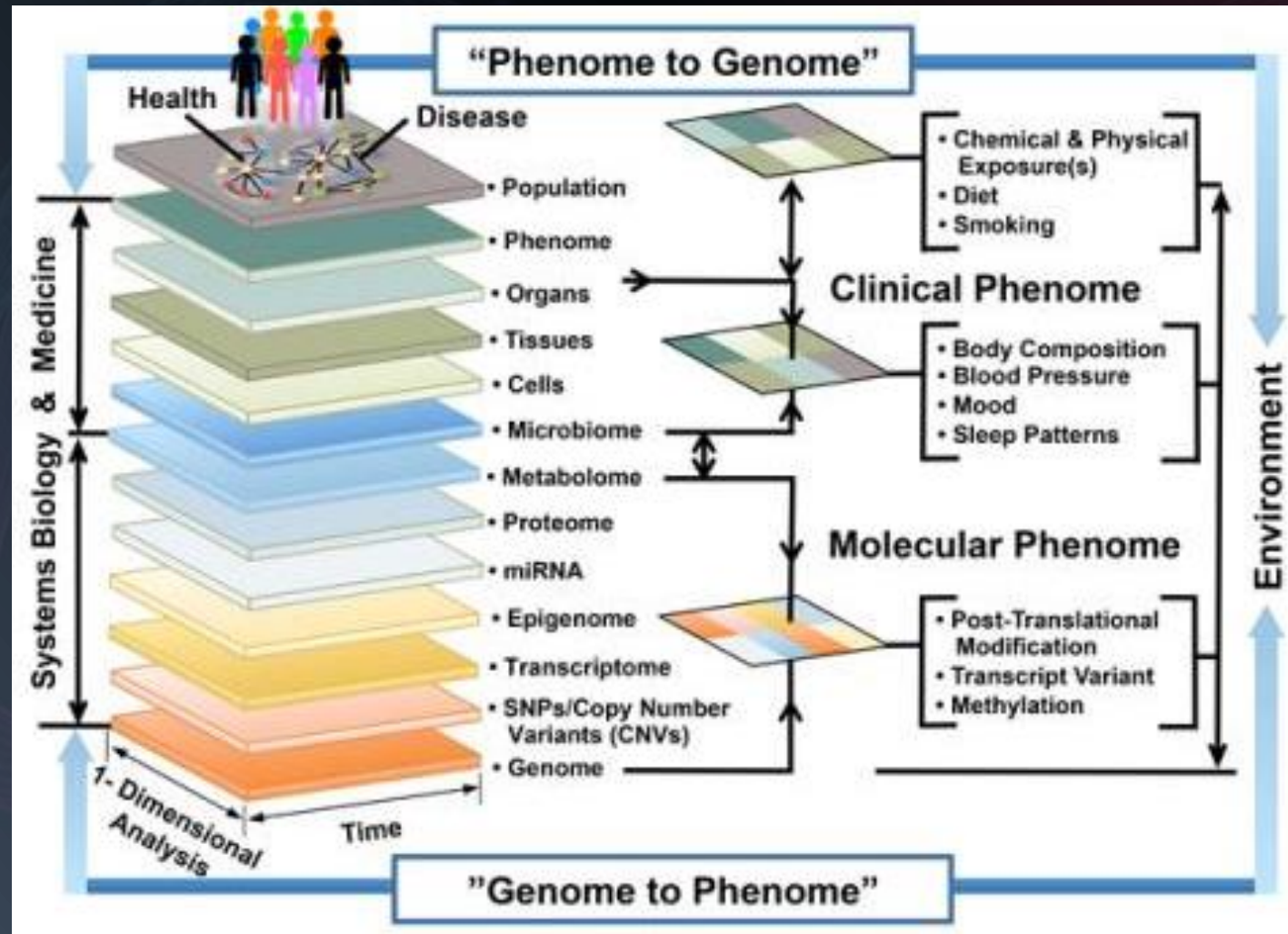


Types of Bioinformatics Data



<http://dx.doi.org/10.1093/nar/gkp986>

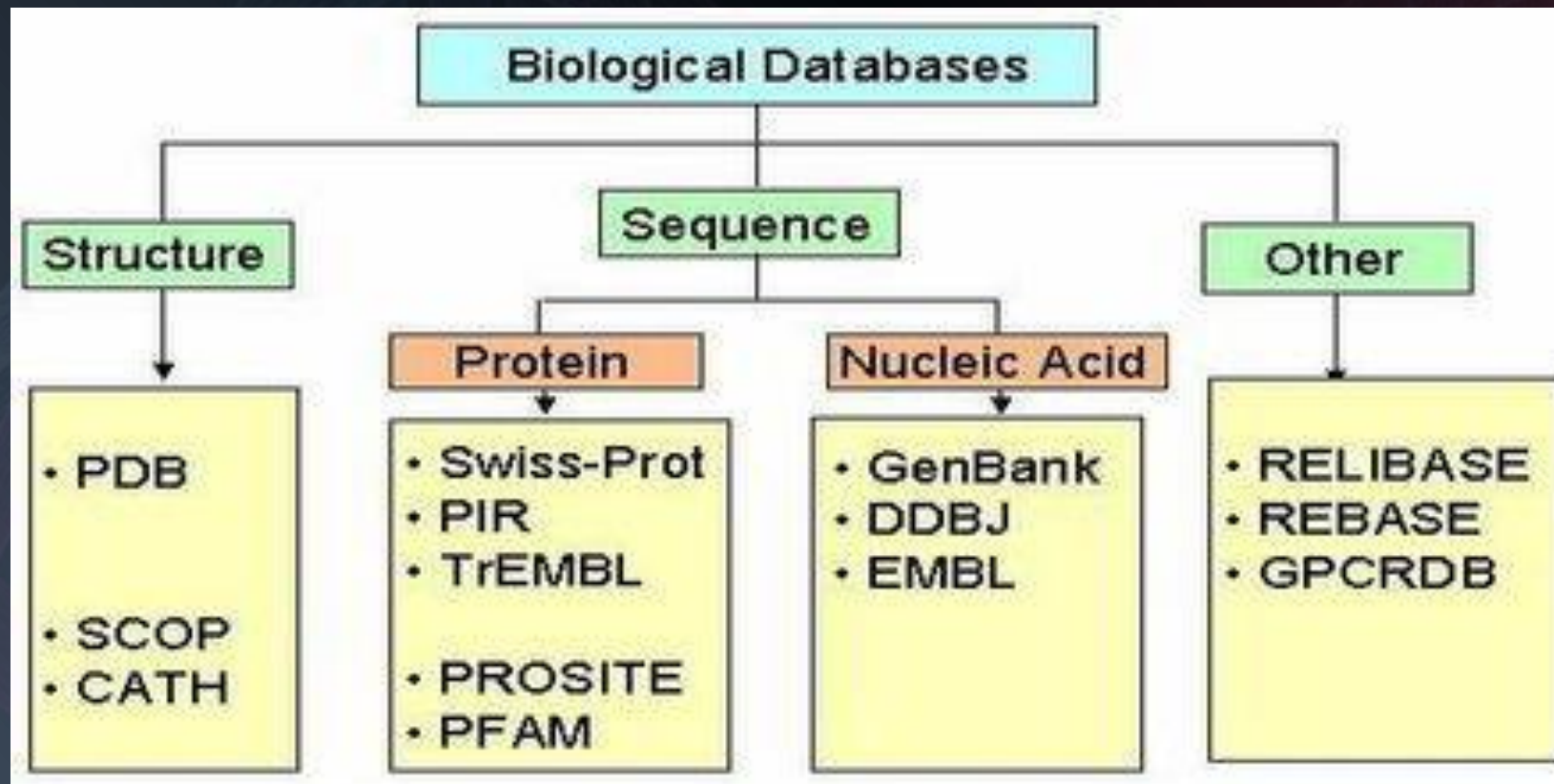
Factors Involved in Data Analysis



Some Bioinformatics Branches

➤ Computational Biology	➤ Using machine learning algorithms to predict protein-protein interactions.
Genetics	Identifying genetic mutations that cause diseases such as cystic fibrosis.
Genomics	Sequencing the entire genome of a species, such as the human genome project.
Proteomics	Identifying and quantifying all the proteins in a biological sample, such as a blood sample.
Metagenomics	Studying the microbial communities in the human gut to understand their role in health and disease.
Transcriptomics	Analyzing gene expression patterns in cancer cells to identify potential drug targets.
Phylogenetics	Reconstructing the evolutionary history of species using DNA sequence data.
Metabolomics	Identifying small molecules in biological samples to understand metabolic pathways and disease mechanisms.

Biological Database



<https://assignmentpoint.com/biological-database/>

Bioinformatics and Computer Engineering

Artificial Intelligence (AI)

➤ AI can be used to develop predictive models for drug discovery, protein structure prediction, and gene expression analysis.

Machine Learning (ML)

ML algorithms can be used to classify and predict gene expression patterns, identify disease biomarkers, and analyze genomic data.

Data Mining

Data mining techniques can be used to identify patterns and relationships in large-scale genomic and proteomic data sets, and to discover new drug targets.

Algorithm Design and Analysis

Algorithms can be designed and analyzed to solve complex bioinformatics problems such as sequence alignment, gene expression analysis, and protein structure prediction.

Bioinformatics and Computer Engineering

Computer Vision

➤ Computer vision techniques can be used to analyze and interpret images of biological samples, such as microscopy images of cells and tissues.

Natural Language Processing (NLP)

NLP techniques can be used to extract information from scientific literature, such as identifying gene-disease associations and drug targets.

Cloud Computing

Cloud computing can be used to provide scalable and cost-effective computing resources for large-scale bioinformatics analysis.

Database Management

Databases can be used to store and manage large-scale genomic and proteomic data sets, and to provide efficient access to this data for analysis.



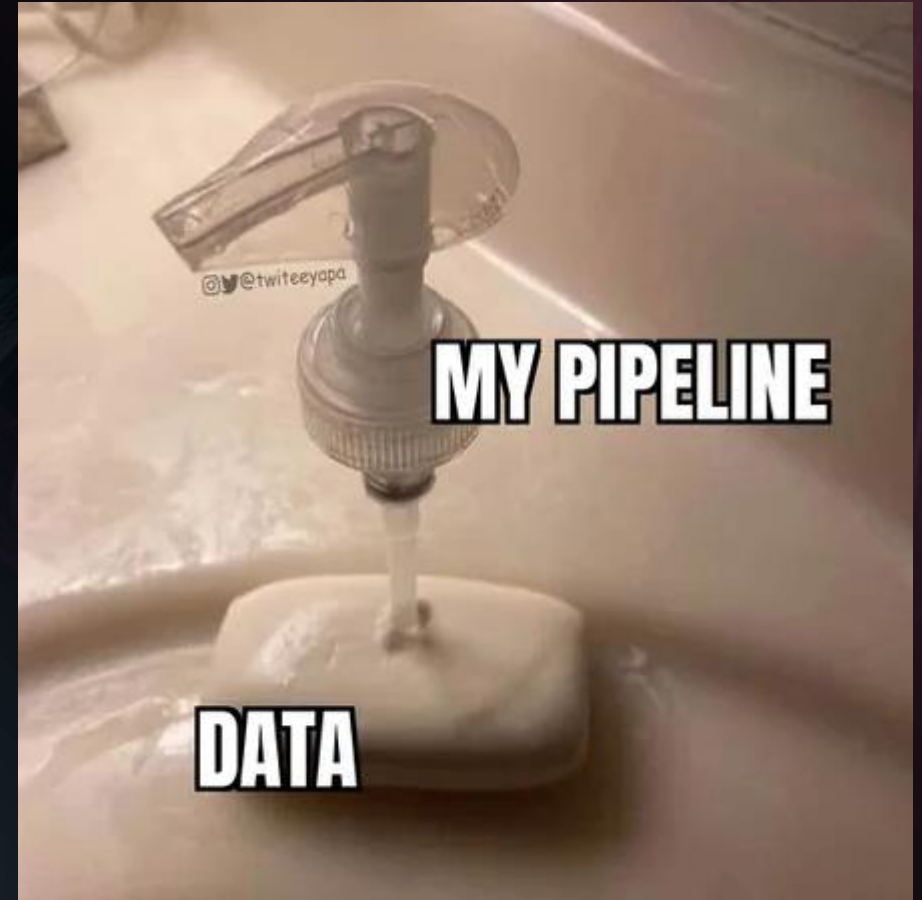
Course Overview

Course Syllabus

- Structure of the Cells
- DNA/RNA/Protein Structure
- Replication/Transcription/Translation
- Cell Fate Determination, Microarray, NGS
- Genome Assembly
- Sequence Comparison (Pairwise/Multiple Seq. Alignment)
- Phylogenetic Tree
- Hypothesis Testing, t-test, permutation test, p-value
- P-value adjustment, Fisher, GO/Pathway Analysis
- Motif Finding
- Markov Models

A Few Recommendations

- Make time for it now that you have taken the course in college.
- Always understand the problem first and make it precise.
- Master the biology data formats you use.
- When solving problems, consider consulting with biologists.
- Have a broad view and recognize the interrelatedness of different areas in the world.
- Search for your interests as different fields may be related!



○
Thank you for
your attention

Any Question?