



# A beginner's guide to bioinformatics

First lecture

Microbiology

September 11, 2023



# Bioinformatics

Programming  
Maths  
Biology  
Statistics  
Software development  
Health science  
Computational modelling  
Machine Learning  
Multi-omics  
RNA analysis  
DNA sequencing  
Algorithms  
Data bases  
Protein structure



# Context

Bioinformatics has revolutionized the modern life sciences and has become a component of many undergraduate training courses and post-graduate research projects. As such, we are seeing more bioinformatics and programming aspects within undergraduate training and so it is important to understand what bioinformatics is, why it is a necessity in modern research and how young academics can begin their journey as bioinformaticians.

# Introduction to bioinformatics

Bioinformatics or computational biology (which are the same thing) has quickly jumped from a specialist term describing an elite group of biologists that swapped their pipettes for Python and Pearl to a more generic description which encapsulates any form of computational analysis on biological data.

These skills were utilized by higher education institutes during the pandemic as undergraduate and post-graduate students were often asked to undertake alternative capstone projects completely based online.

Bioinformatic skills are also heavily required within the research environment as the volume of data being generated by labs is increasing, so the demand is growing for those with the skills to process and analyse large quantities of data.

# How does bioinformatics work?

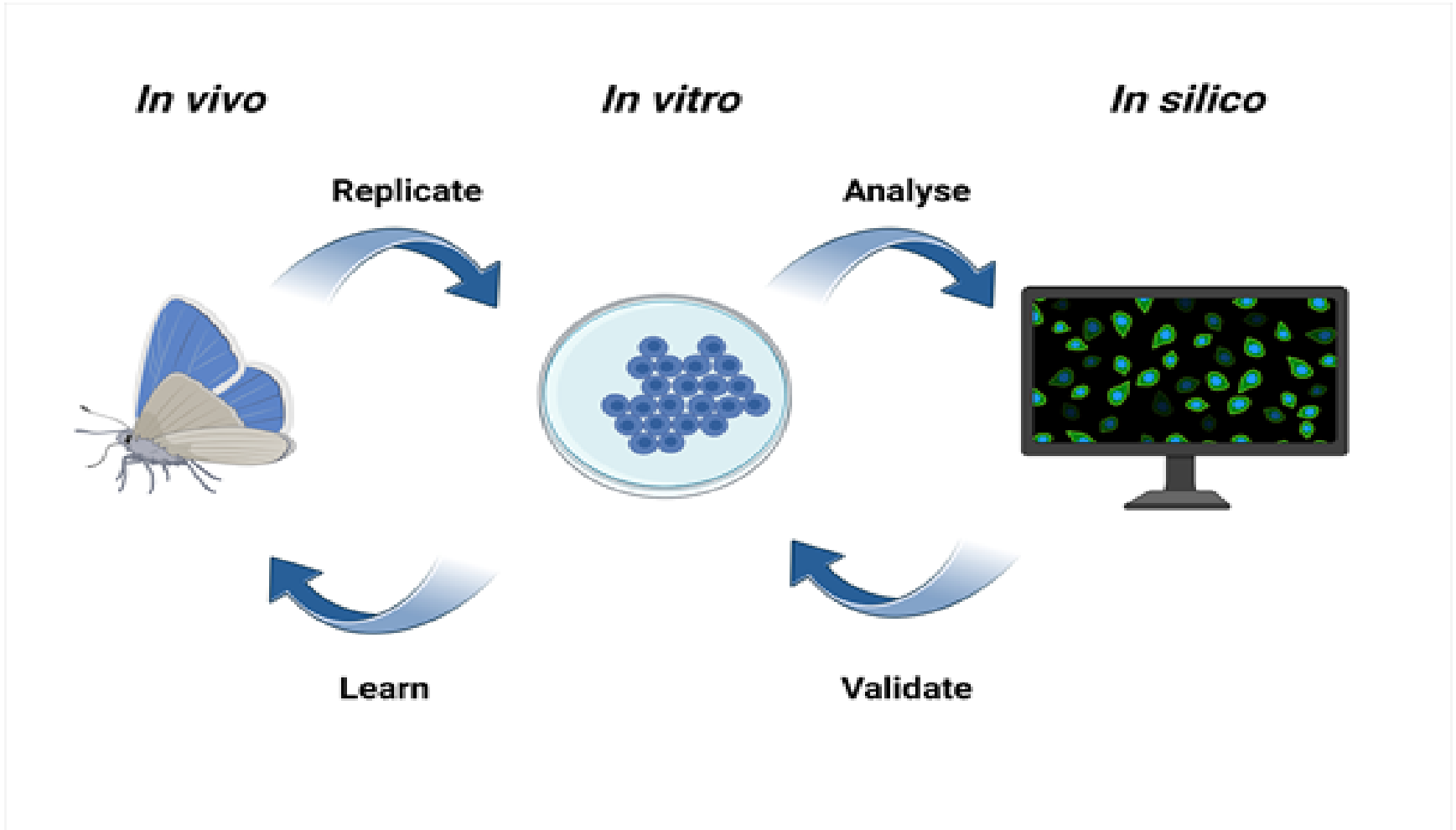
Many questions from life science come from the natural world/in vivo - this can include any range of topics such as investigating the molecular causes of diseases, contrasting virulence of different pathogens or researching the origins of human evolution.

Researchers could then replicate and test what was seen in the natural world in the laboratory using in vitro techniques such as a western blot to measure the effect a gene mutation has on the protein expression level of a known oncogenic gene, or infecting cells with a pathogen and contrasting the viability between infected cells and non-infected cells.

A bioinformatician would then use what has been measured in vitro to inform computational analysis (in silico).

The analysis itself can range from measuring samples in parallel to reduce manual effort, running specialist software for quality control or simply producing graphs in a popular coding language like Python or R. The goal of the analysis could be to generate testable hypothesis which can be validated back in the laboratory, and this in turn would lead to learning something new about the natural world.

# How does bioinformatics work?



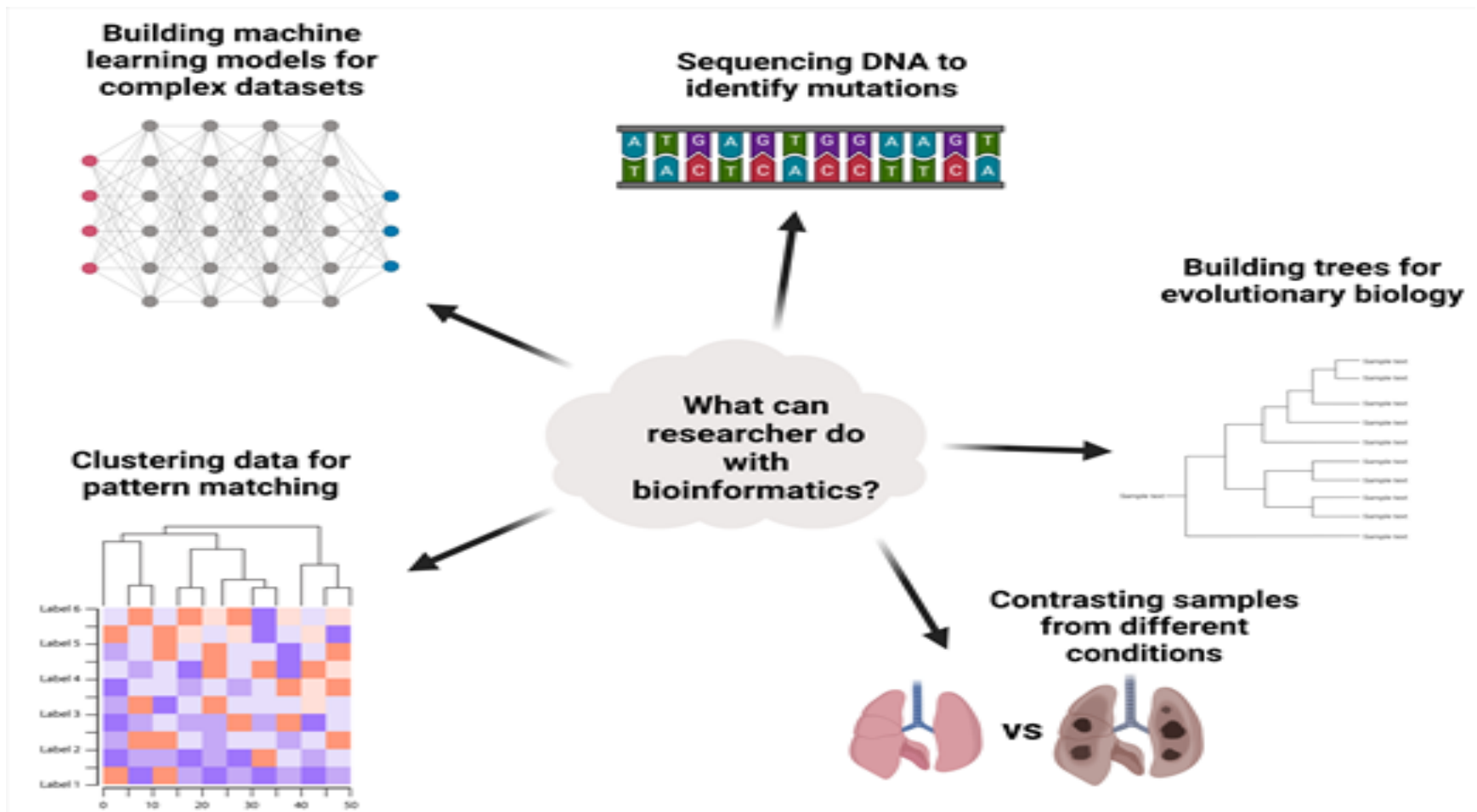
# Why is bioinformatics necessary in biological sciences?

Bioinformatics supports every aspect in modern biological research. In some ways bioinformatics is not a useful term for the vast array of different tasks that can be performed for uses in biology.

Building machine learning models to make predictions or classifications based on biological data is a novel method of using large complex datasets to find patterns which cannot be detected without the use of powerful algorithms.

Finally, graphical representation of large data such as clustering graphs (e.g., heatmaps) to identify similarities between biological samples is useful and can lead to efficient methods of identifying patterns. In each of these examples the specific context is less important than the overall theme of using data science approaches to unearth novel insights from biological information.

# bioinformatics necessary in biological sciences





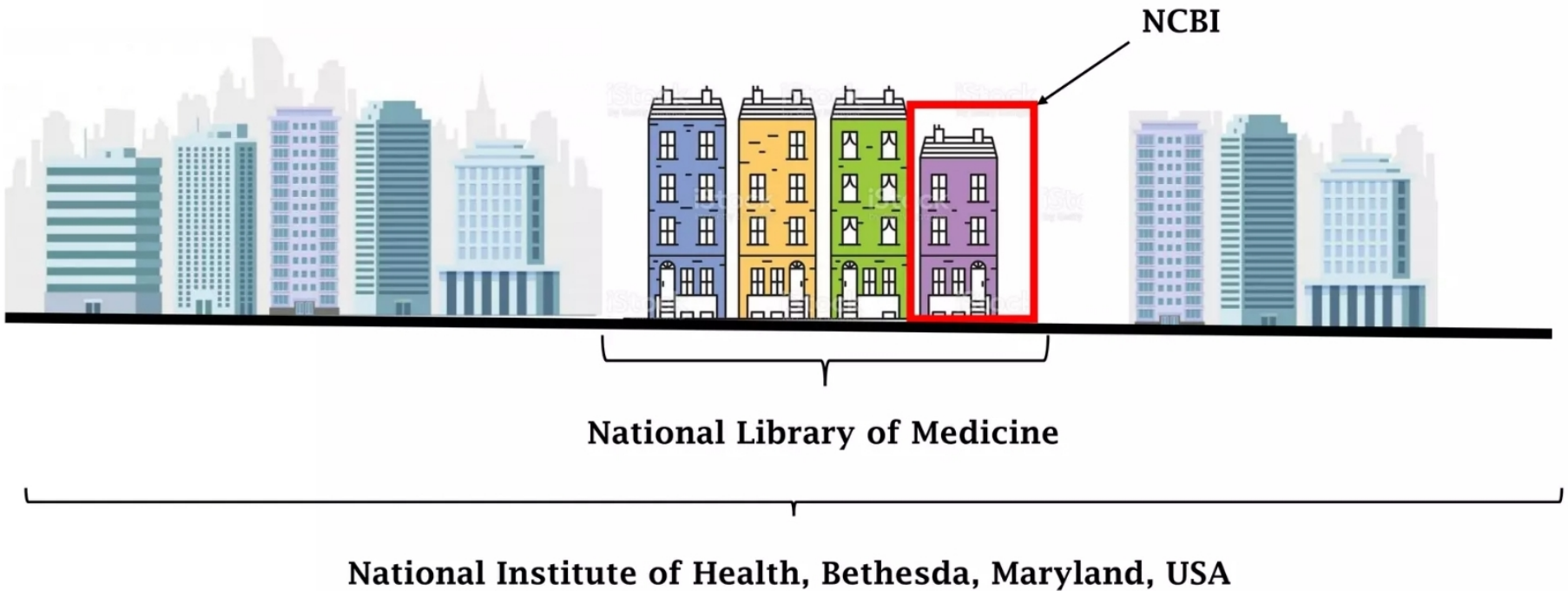



# **National Center for Biotechnology Information**

# **NCBI**

# Introduction

NCBI is established in 1988 as a part of **National Library of Medicine** at the **National Institutes of Health**, Bethesda, Maryland, USA.



- 
- The **NCBI** houses a series of **databases** relevant to biotechnology and biomedicine and consists of important **resources** for bioinformatics tools and services.
  - Major **databases** include **GenBank** for **DNA sequences** and **PubMed**, a bibliographic database for **biomedical literature**.
  - Major **Resources** includes Books, software tools (BLAST), retrieval system (Entrez), submission tools (Bankit, Sequin), specialized tools (ORF Finder, e-PCR, Spidey/Splign).

# Aim of NCBI

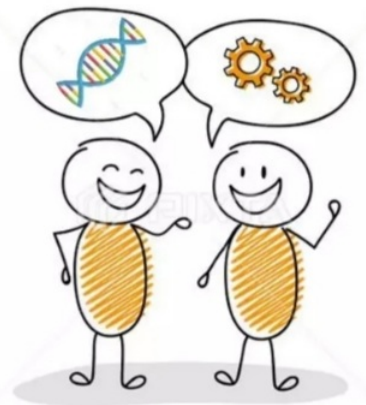
**1. To Develop  
Public Databases**



**2. To Develop Software Tools**



**3. To Spread  
Biomedical  
Information**



# NCBI Database & Resources

Databases	Brief description
Nucleotide	Sequence database, e.g., GenBank
Genome	Complete genomes
Taxonomy	Classification of organisms in NCBI sequence database
Structure	MMDB (Molecular Modelling Database): experimental 3D structure
Domains	CDD (Conserved Domain Database): Conserved protein domains
3D Domains	Compact 3D protein domains in MMDB
OMIM	Online Mendelian Inheritance in Man
SNP	Single Nucleotide Polymorphism
UniSTS	Sequence Tagged Site markers
GEO	Data repository of gene expression data
PopSet	Population study datasets
UniGene	Gene-based expressed sequence clusters
HomoloGene	Eukaryotic homology groups
Cancer Chromosomes	Chromosomal aberrations in cancer database
GENSAT	Gene expression pattern in mouse CNS
Protein	Protein database compiled from various sources
PubMed	Biomedical literature
PubMed Central (PMC)	Free and full text journal articles
Books	Online text books

