

# Introduction to bioinformatics

## Chapter 1 Introduction

Jiaxing Chen

# Course Information

Instructor: Jiaxing Chen

Email: jiaxingchen@uic.edu.cn

Office T7-102

Teaching assistant: Rubin Zhou

Email: rubinzhou@uic.edu.cn

Office: T3-502-R26

# Evaluation

Type of Assessment Methods	Weighting	CILOs to be Addressed	Description of Assessment Tasks
Quizzes	30%	1,2	Quizzes will test students' understanding of basic concepts and algorithms on bioinformatics.
Labs	20%	2,3	Lab practices will measure students' understanding of basic concepts and algorithms on bioinformatics, and make sure students are prepared for solving real-life bioinformatic problems.
Assignments	20%	1,3	Written and programming assignments will measure the students' understanding of the theory and their ability to solve practical problems.
Group project and essay writing	30%	2-4	Group project will test students' ability in selecting proper methods to explain, analyse and solve real-world problems.

# Course Contents

---

- Introduction
    - What is bioinformatics
    - Examples on application of bioinformatics on daily research
    - Major research directions of bioinformatics
  - Genome, gene, RNA, protein and classic algorithms
    - Understanding genomes/genes/human genetic studies
    - RNA Structure Prediction and Non-Coding RNAs
    - introduction to protein structures
  - Bioinformatic Data
    - Data Format, eg. Fasta, gtf, vcf, RPKM/FPKM
    - Database and tool, eg. National Center for Biotechnology Information (NCBI), Ensembl (EMBL-EBI/Sanger Institute), Protein Data Bank and other important data resources
-

# Course Contents

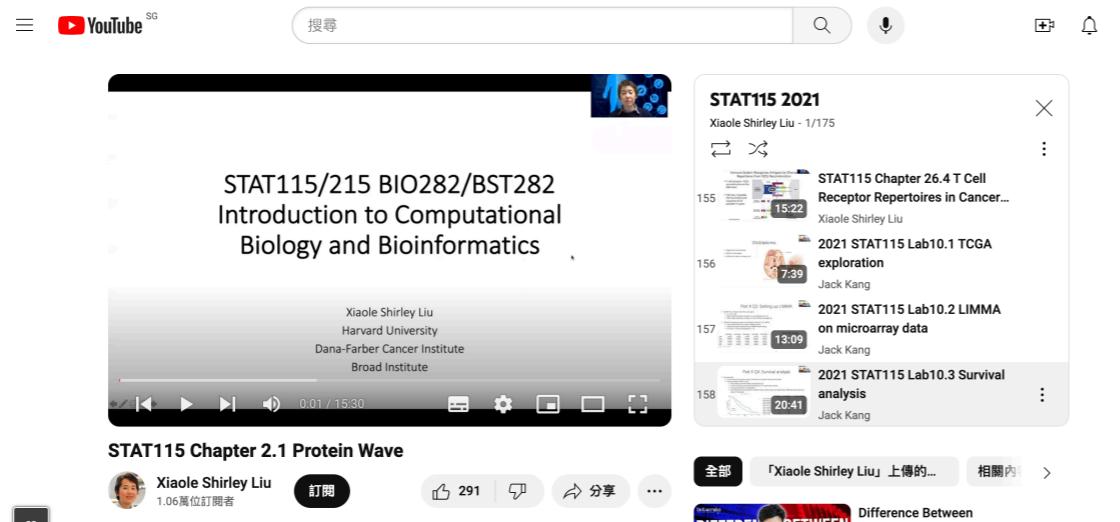
---

- Statistical analysis and machine learning in bioinformatics
    - statistical model, eg. Data preprocessing, Differential expression, Enrichment
    - Machine learning in bioinformatics, Clustering and classification for gene expression, Medical image segmentation, Protein structure prediction, Imputation
  - State-of-art technologies and case study
    - Single-cell sequencing and Human Cell Atlas
    - Spatial transcriptomics
    - Genome editing
    - Biological networks
-

# Further reading

Harvard University: STAT115 2021 Introduction to Computational Biology and Bioinformatics  
(more bioinformatics topics)

<https://www.youtube.com/watch?v=LvtPXZHNoCc&list=PLeB-Dlq-v6taAXK6ZCGfqlmrNWJzFt3p3>



CMU: Bioinformatics  
(more focus on related algorithms)  
<https://www.cs.cmu.edu/~ckingsf/bioinfo-lectures/>

## Bioinformatics Lectures

(b) indicates slides that contain primarily background information. (a) indicates "advanced" material. All slides (and errors) by Carl Kingsford unless noted. Often the material for a lecture was derived from some source material that is cited in each PDF file.

Introduction to Computers and Biology

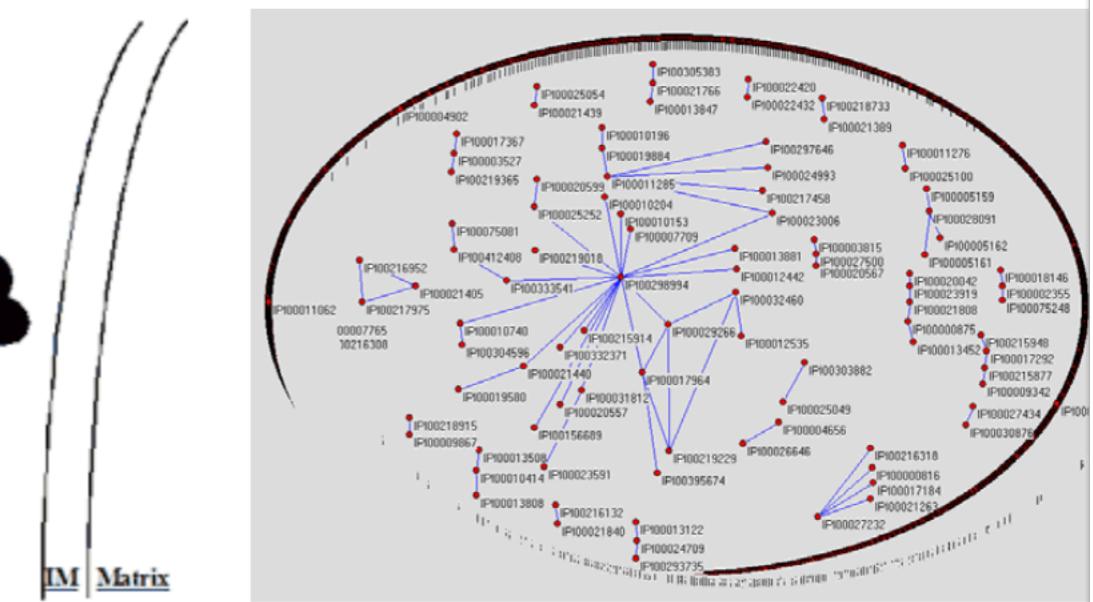
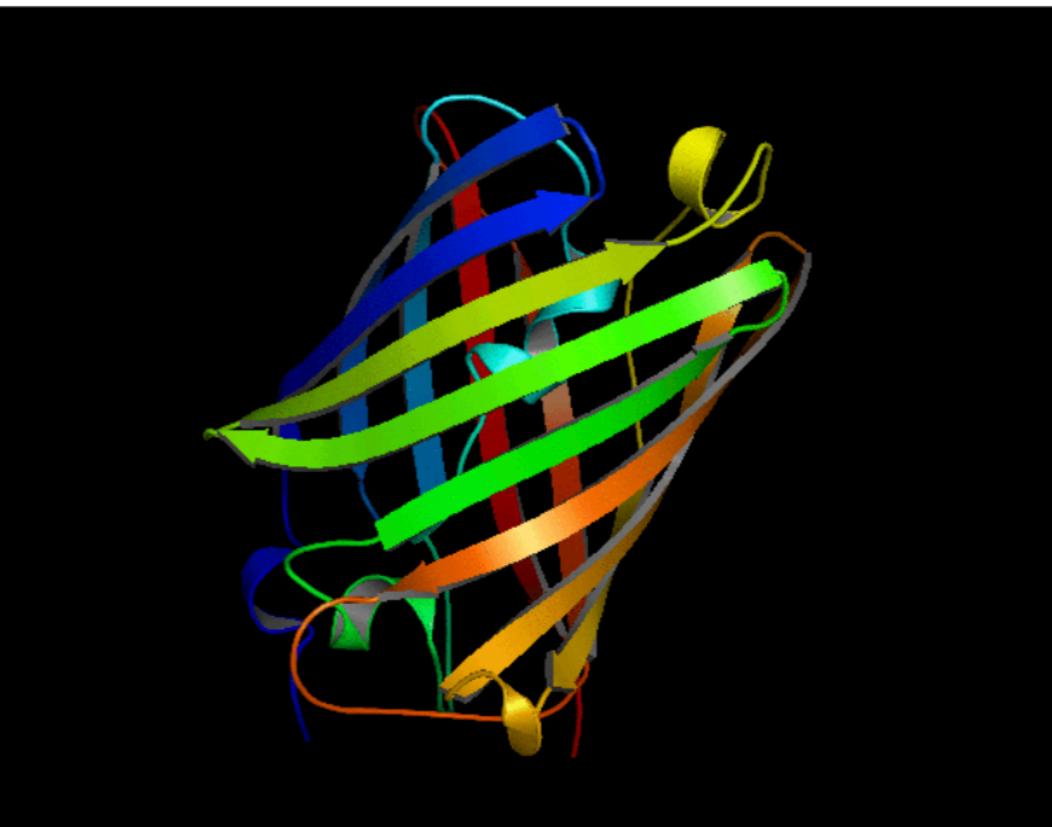
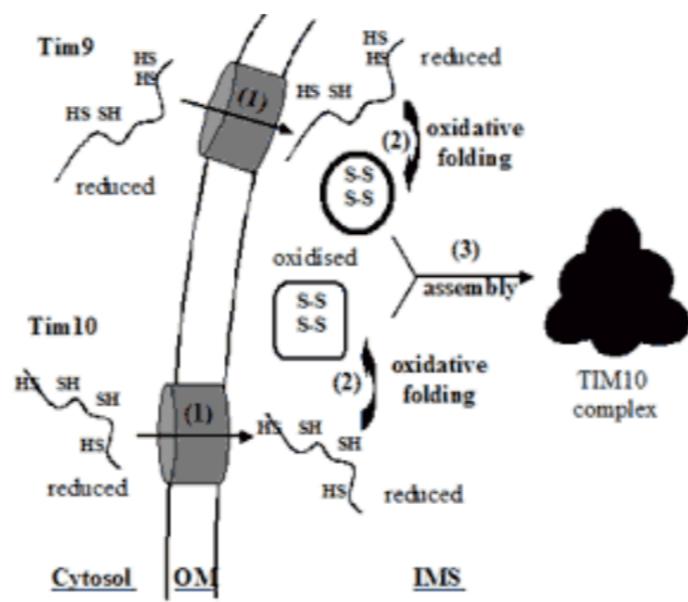
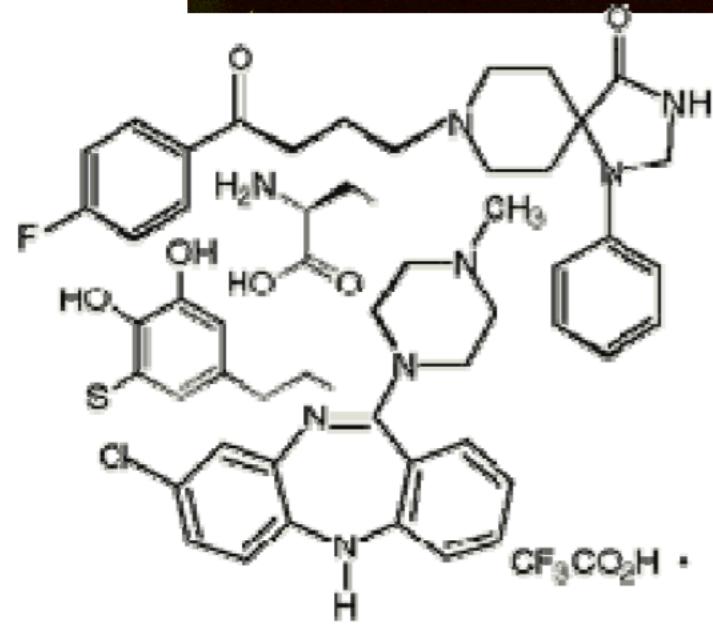
1. [Introduction and Computational Successes](#)
2. [Quick Biology Introduction](#) (b)

Peking University: Bioinformatics: Introduction and Methods  
(more information on sequence related problems)  
<https://www.coursera.org/course/pkubioinfo>



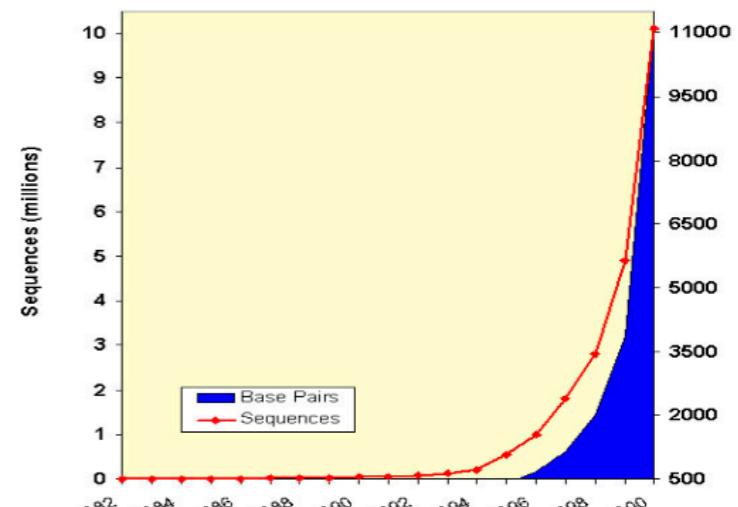
<https://www.coursera.org/course/pkubioinfo>

## Various biological data

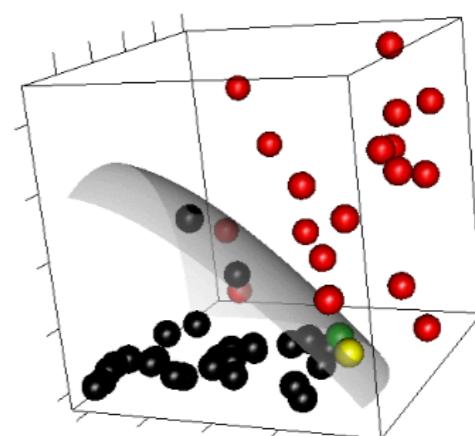


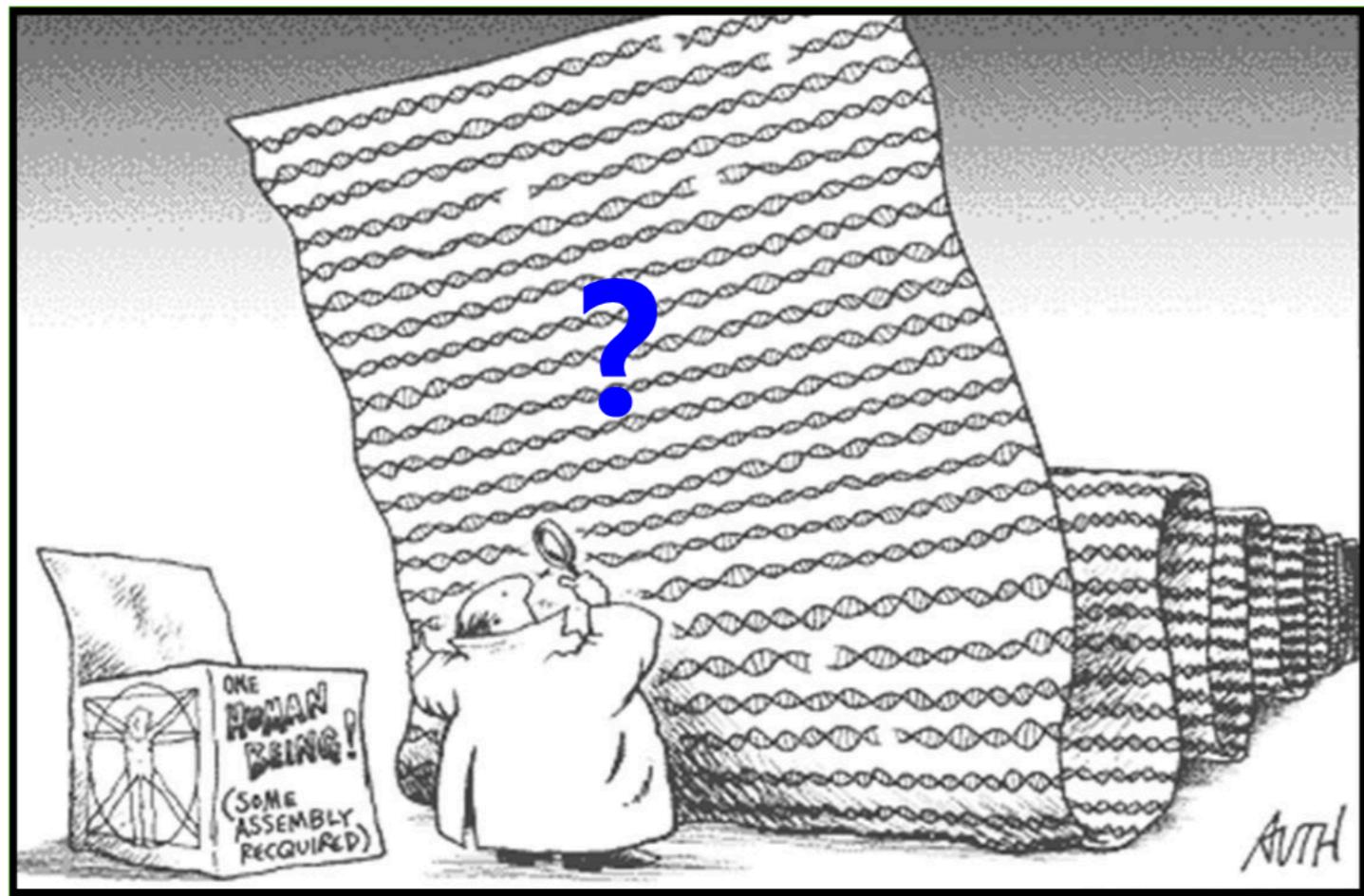
# Opportunities and challenges hand-in-hand: the driving forces of bioinformatics

- High-throughput data
  - Huge amount
  - Explosive growth
  - Low signal-to-noise ratio
  - Multiple types
- Requirements for the methods
  - Data needs to be stored in efficient **ontology-based database** systems
  - The huge amount of data requires **efficient** algorithms
  - Exponential growth requires **scalable** methods
  - The low signal-to-noise ratio requires **accurate** methods
  - Multiple types of data require data **integrative** methods



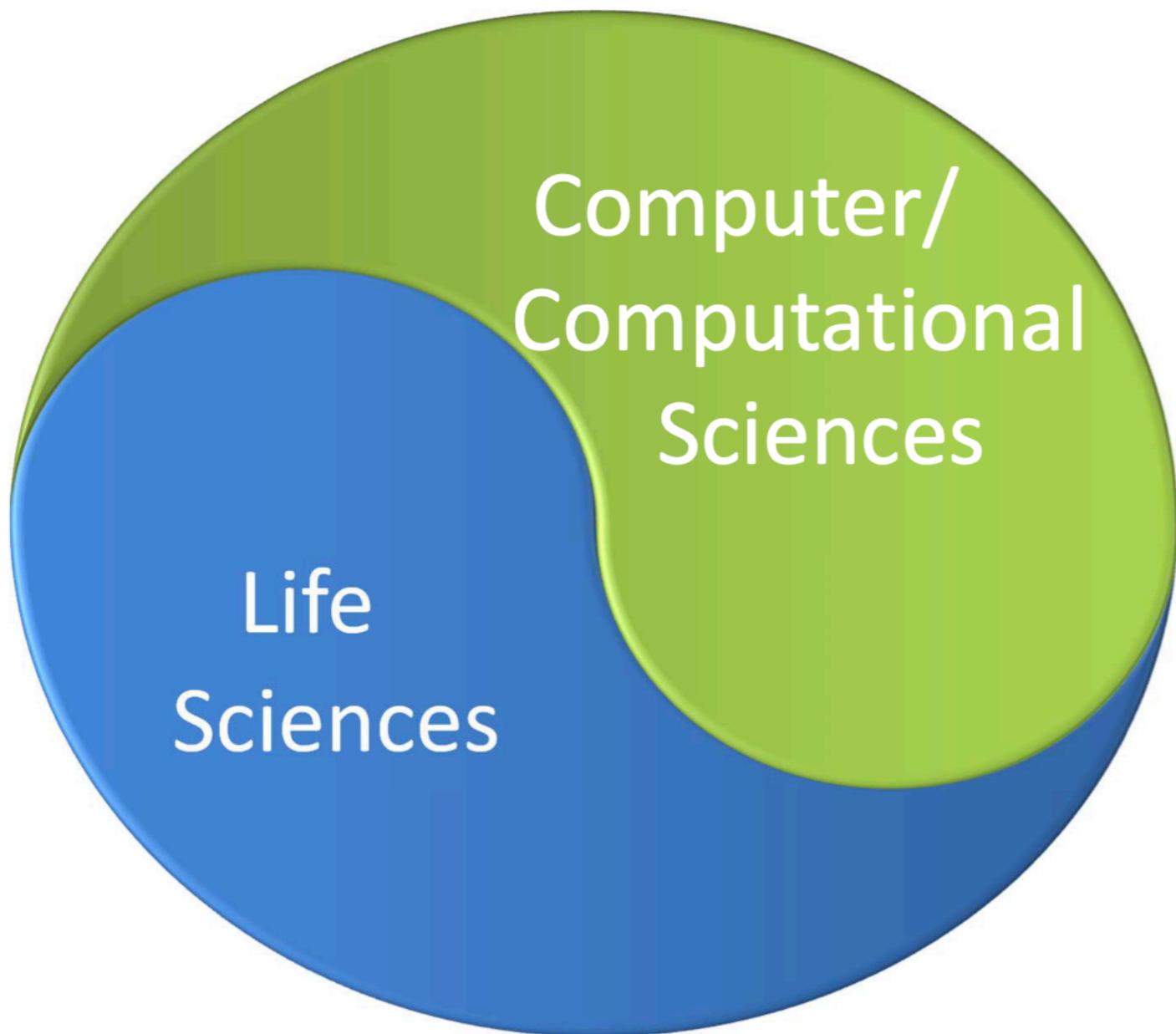
<ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt>





3 billion of 'letters'  
per person

**How can I find the 'fish' I want?**



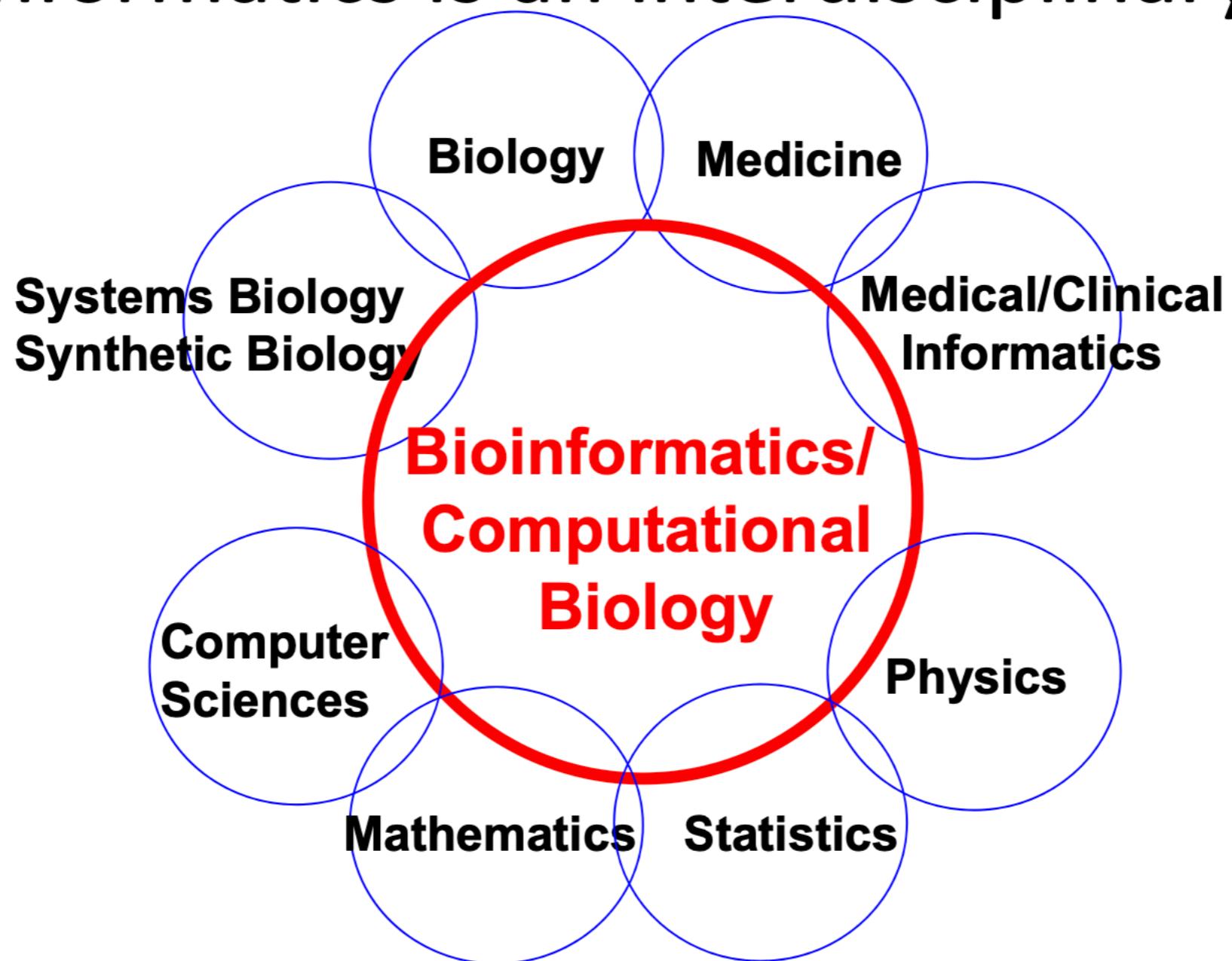
# Bioinformatics

# **What is bioinformatics?**

**Bioinformatics: an interdisciplinary field that develops and applies computer and computational technologies to study biomedical questions**

- As a technology, bioinformatics is a powerful technology to manage, query, and analyze big data in life sciences.
- As a methodology, bioinformatics is a top-down, holistic, data-driven, genome-wide, and systems approach that generates new hypotheses, finds new patterns, and discovers new functional elements.

# Bioinformatics is an interdisciplinary field

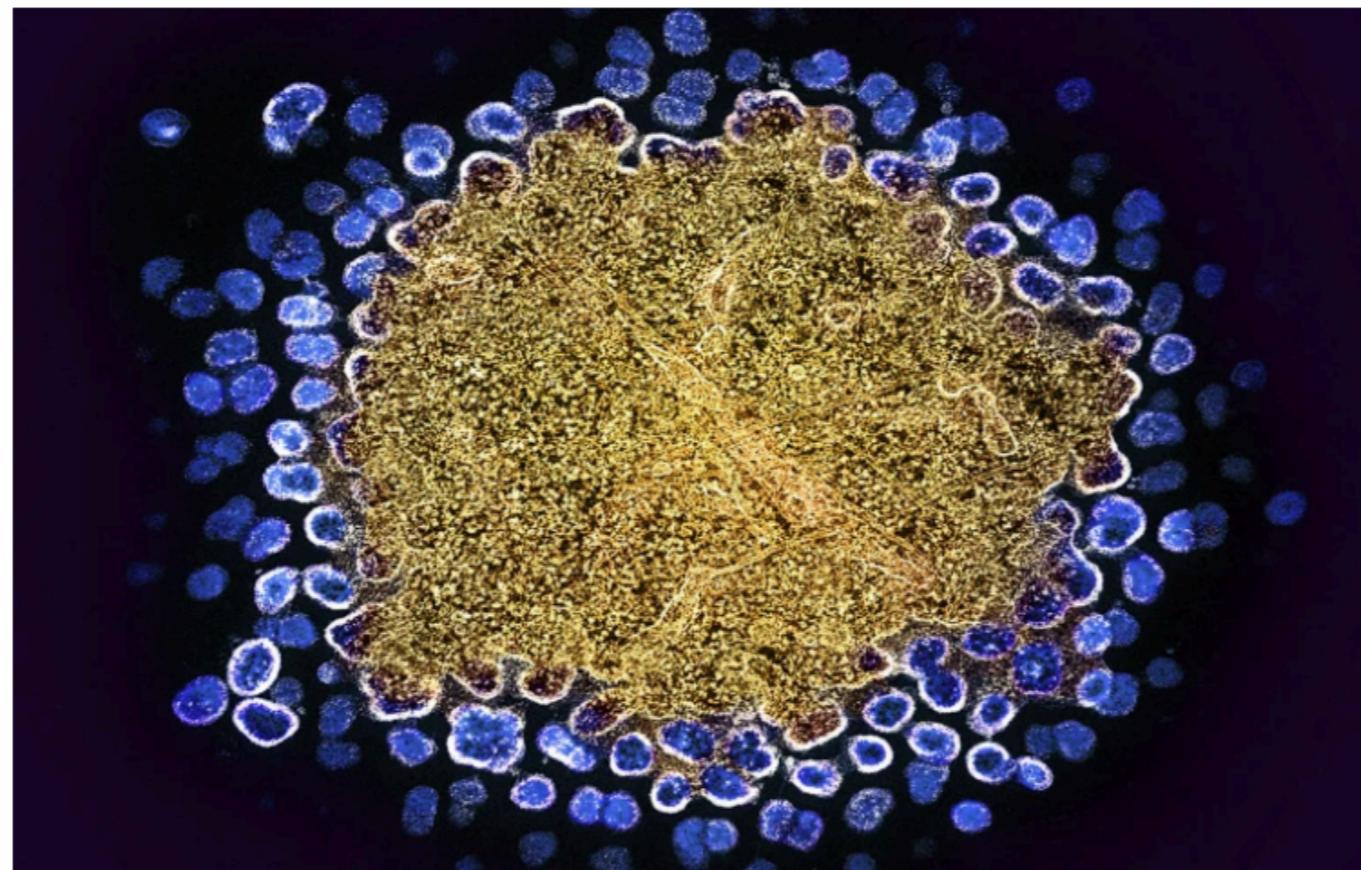


What we can doing in bioinformatics?

NEWS | 26 July 2024

# Seventh patient ‘cured’ of HIV: why scientists are excited

A man in Germany is HIV-free after receiving stem cells that are not resistant to the virus.

By [Smriti Mallapaty](#)

Mutations in the gene that encodes a receptor called CCR5 can stop HIV (blue) entering immune cells. Credit: NIAID/National Institutes of Health/SPL

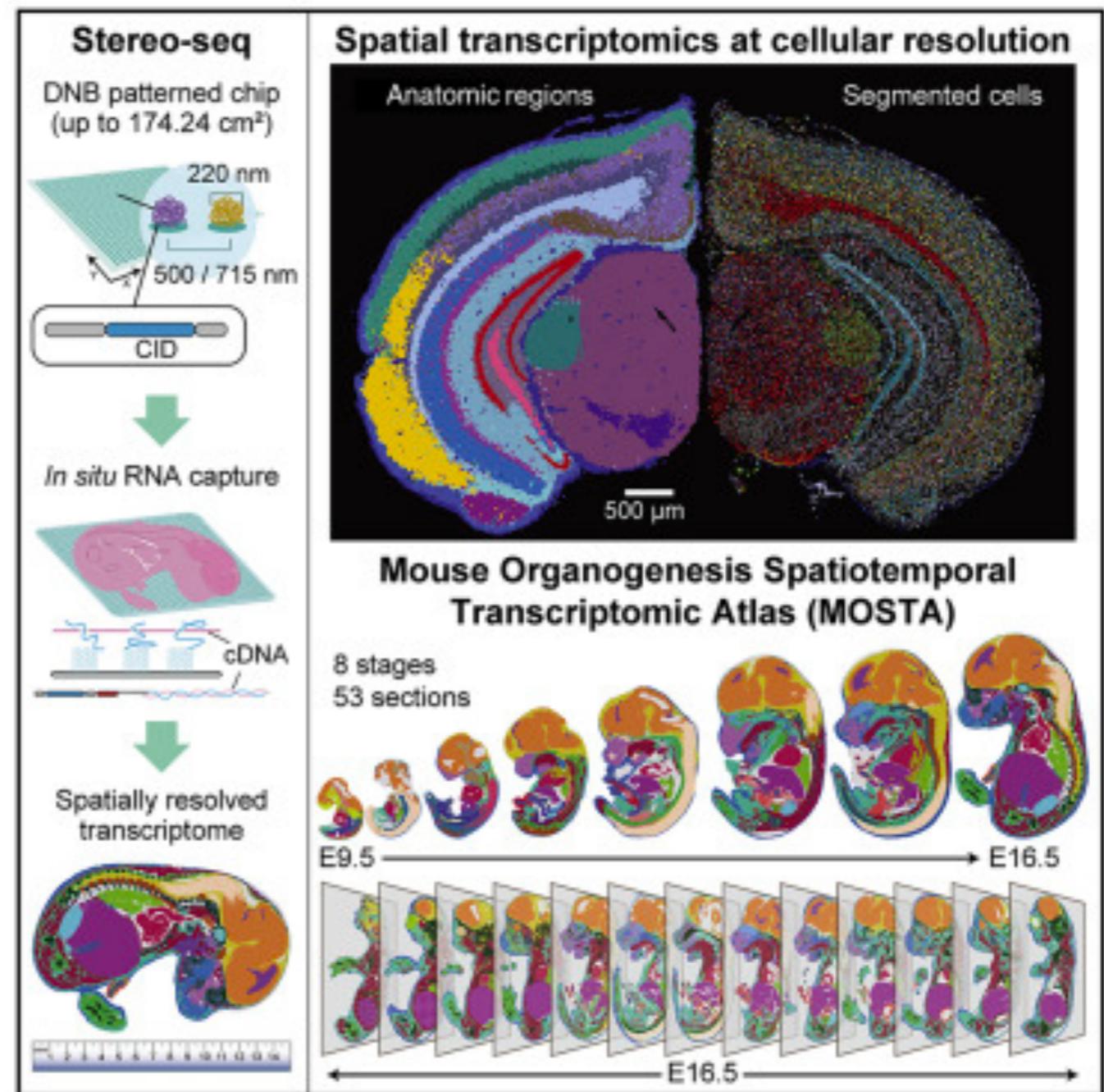
[nature](#) > [articles](#) > [article](#)Article | [Published: 23 August 2023](#)

# Increased hyaluronan by naked mole-rat Has2 improves healthspan in mice

[Zhihui Zhang](#), [Xiao Tian](#), [J. Yuyang Lu](#), [Kathryn Boit](#), [Julia Ablaeva](#), [Frances Tolibzoda Zakusilo](#), [Stephan Emmrich](#), [Denis Firsanov](#), [Elena Rydkina](#), [Seyed Ali Biashad](#), [Quan Lu](#), [Alexander Tyshkovskiy](#), [Vadim N. Gladyshev](#), [Steve Horvath](#), [Andrei Seluanov](#)✉ & [Vera Gorbunova](#)✉

# nature methods

Method of the Year 2020:  
Spatially resolved transcriptomics



News Feature | Published: 01 June 2023

# Inside the nascent industry of AI-designed drugs

[Carrie Arnold](#)

[Nature Medicine](#) **29**, 1292–1295 (2023) | [Cite this article](#)

**43k** Accesses | **27** Citations | **504** Altmetric | [Metrics](#)

**Artificial intelligence tools are beginning to upend the drug discovery pipeline, with several new compounds entering clinical trials.**

Drug discovery is expensive, inefficient, and fraught with failure. An estimated 86% of drug candidates developed between 2000 and 2015 [did not meet their stated endpoints](#).



# Precision Medicine?

Identifying Potential  
Drug Targets

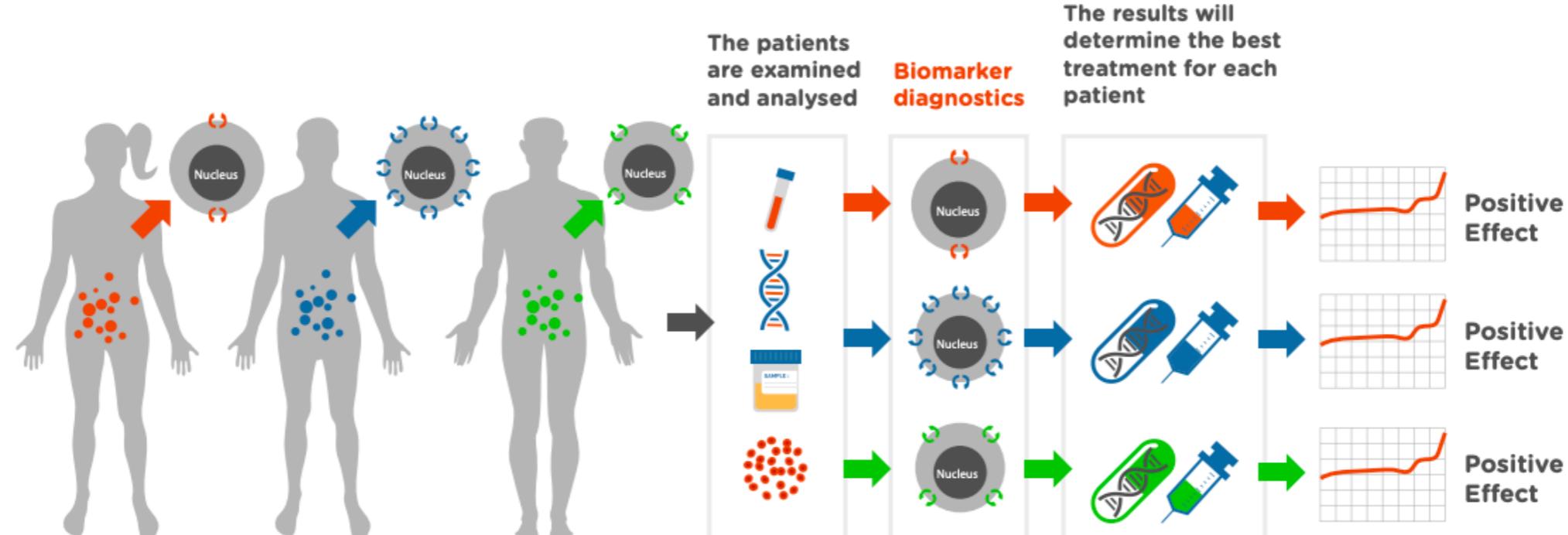
Real-Time Monitoring  
of Patient Health

AI-Powered  
Early Detection  
of Diseases

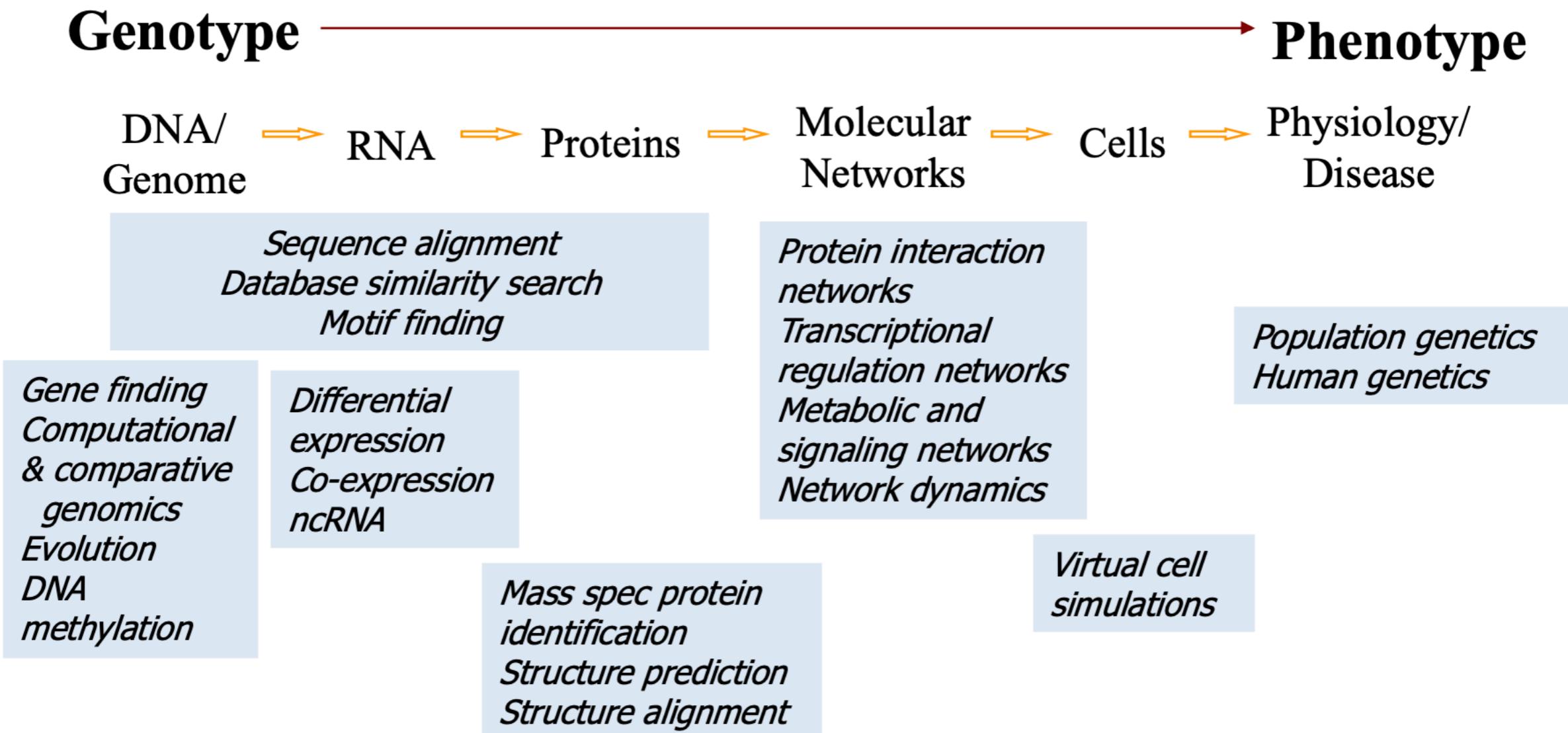
Identifying  
Causal Genes

Phenotypic & Genetic  
Heterogeneity

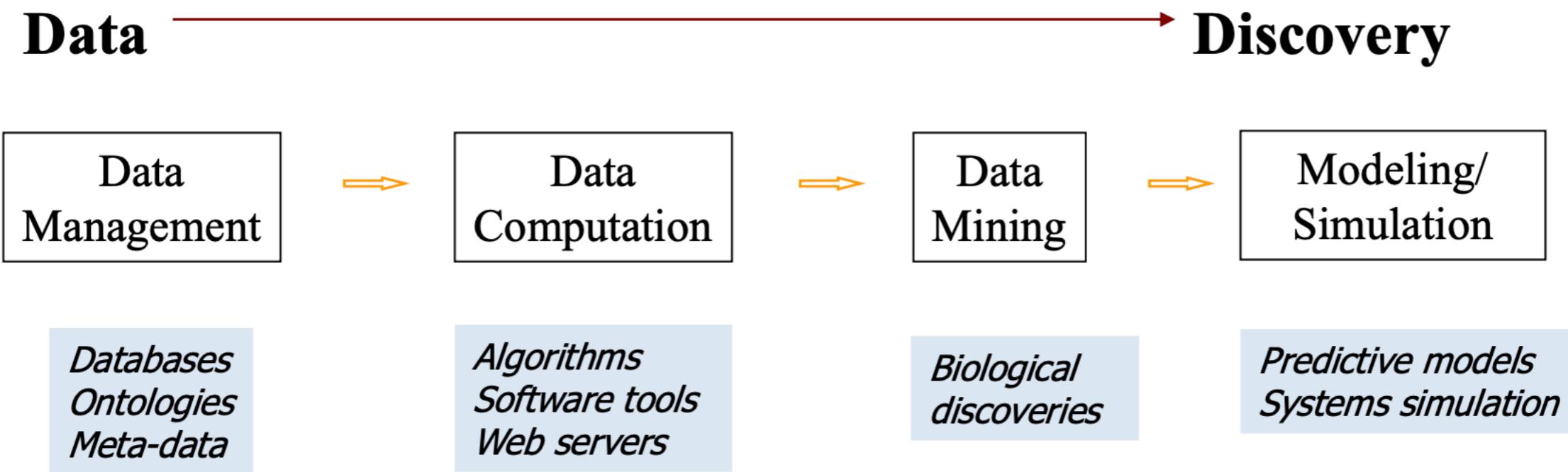
Changing Role of  
Physicians



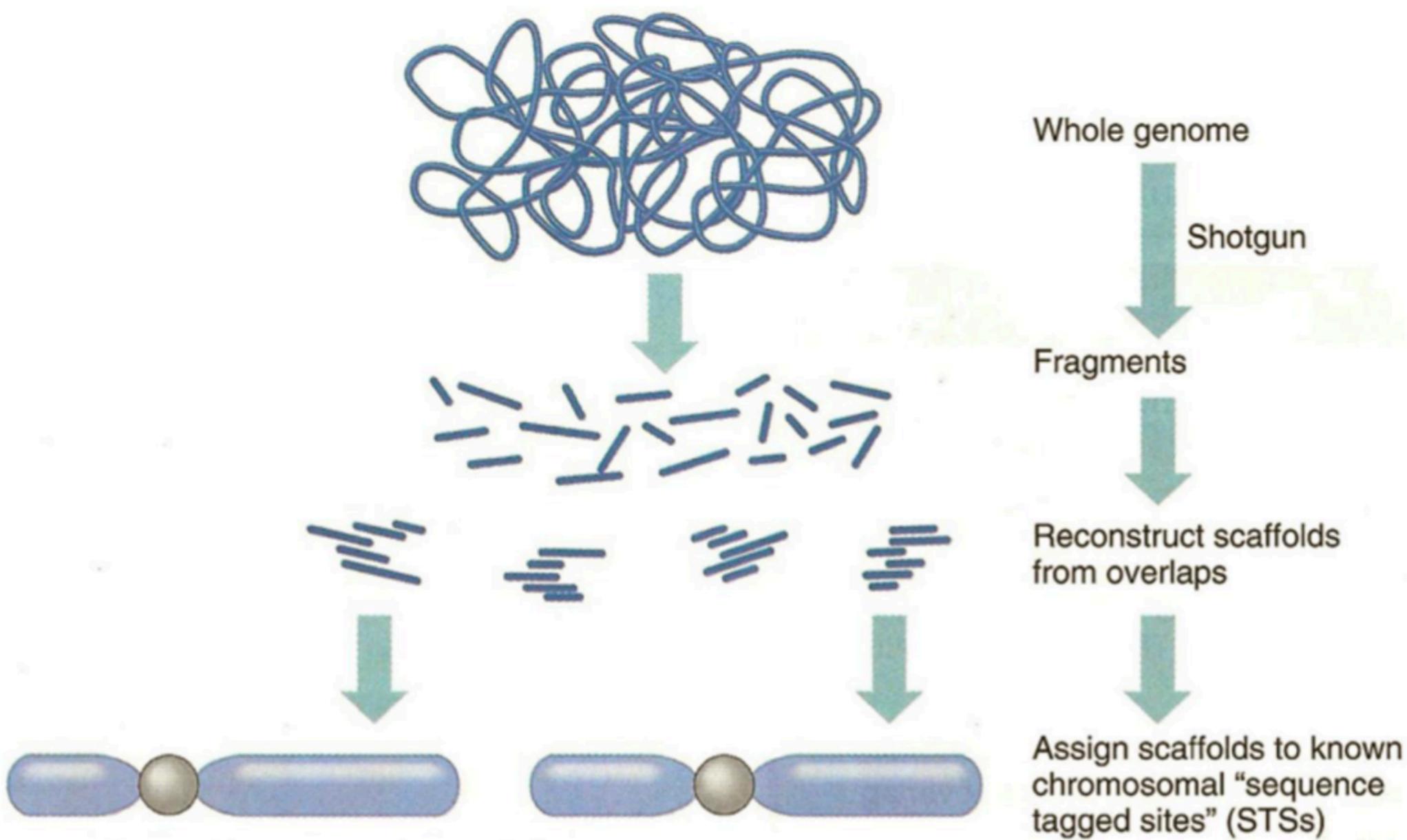
# The Bio- in Bioinformatics



# The -informatics in Bioinformatics

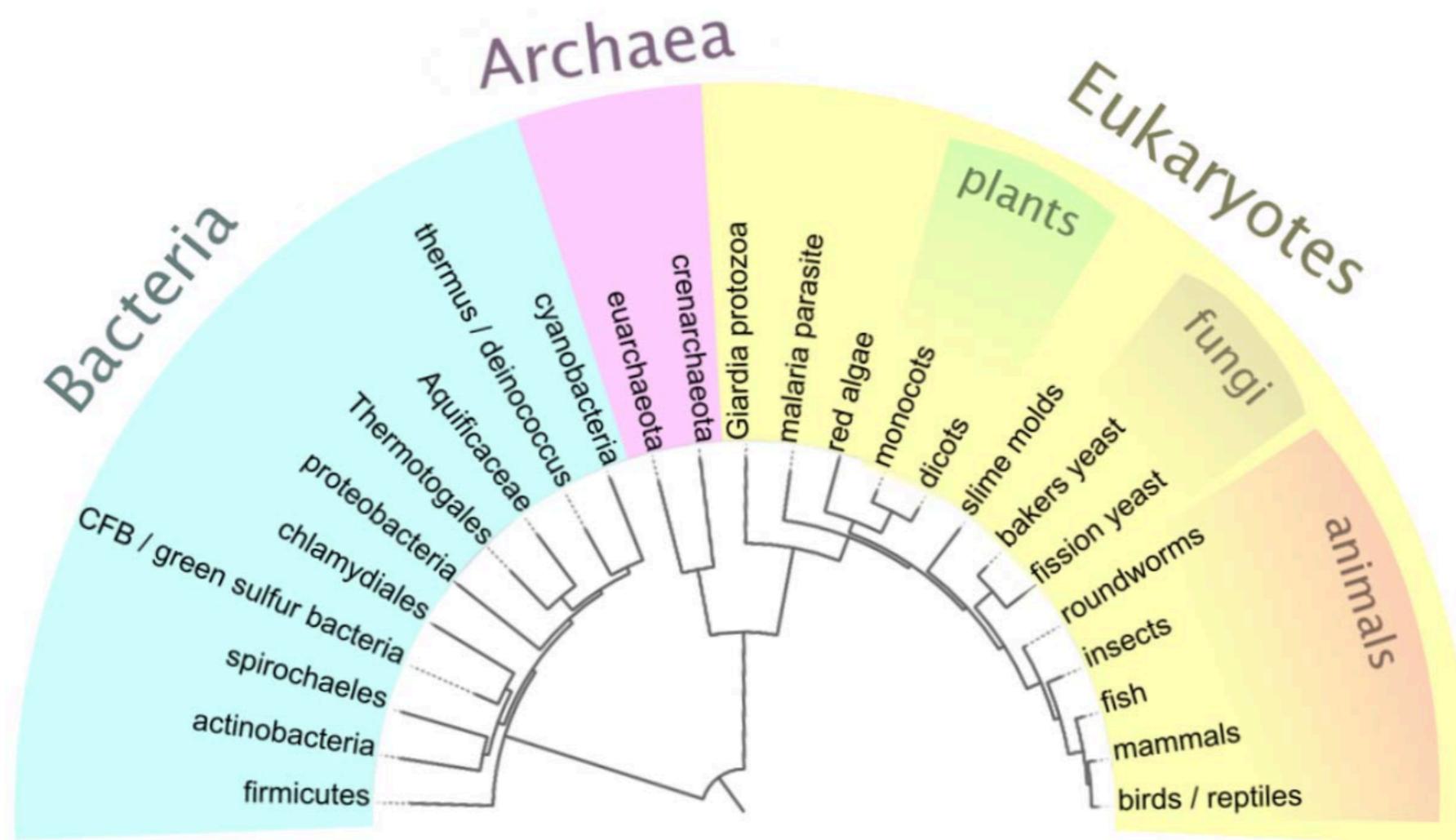


## Research Directions of Bioinformatics — General process of genome assembly



# The universal code: Other species' genomes

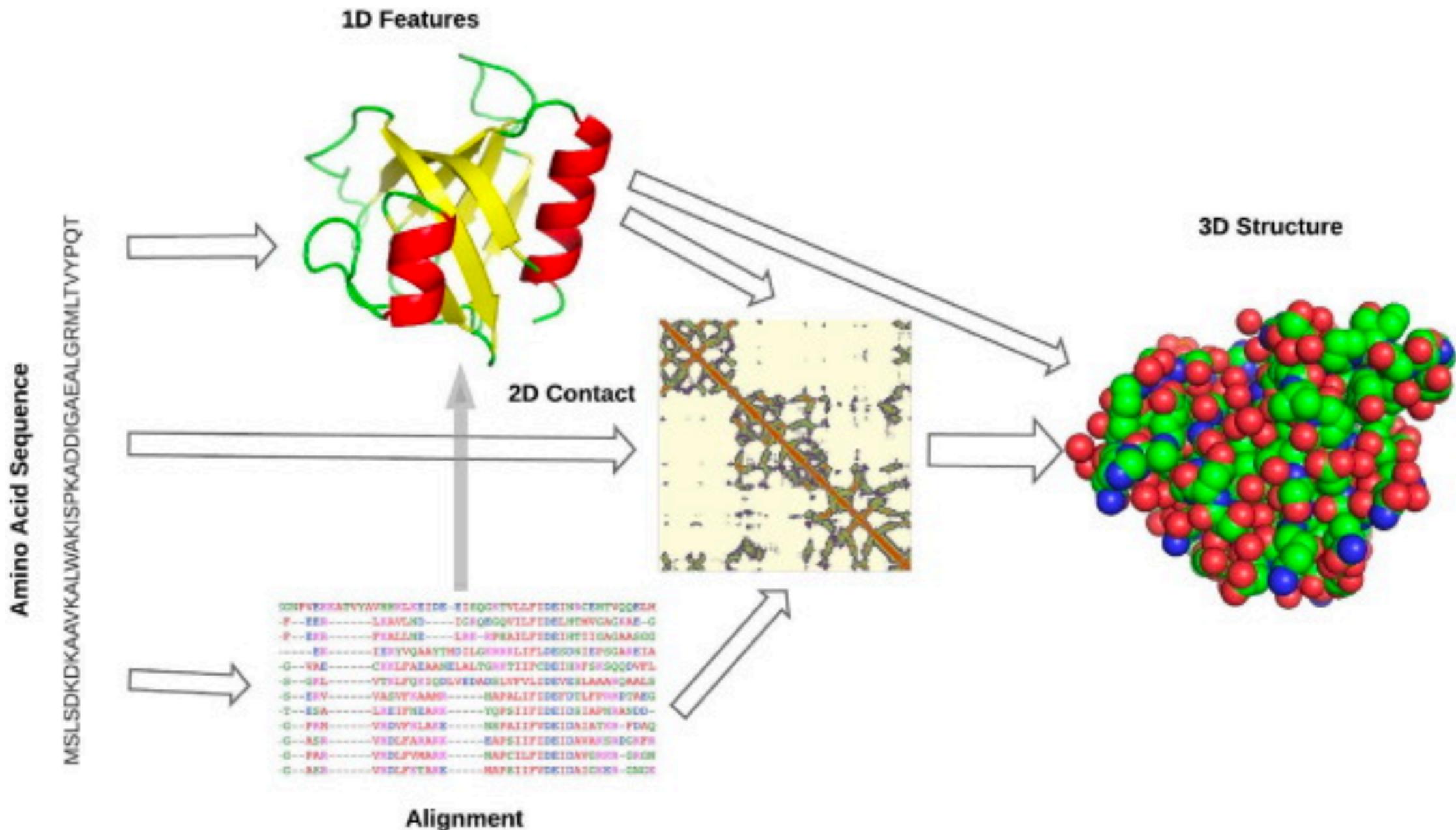
## The Tree of Life



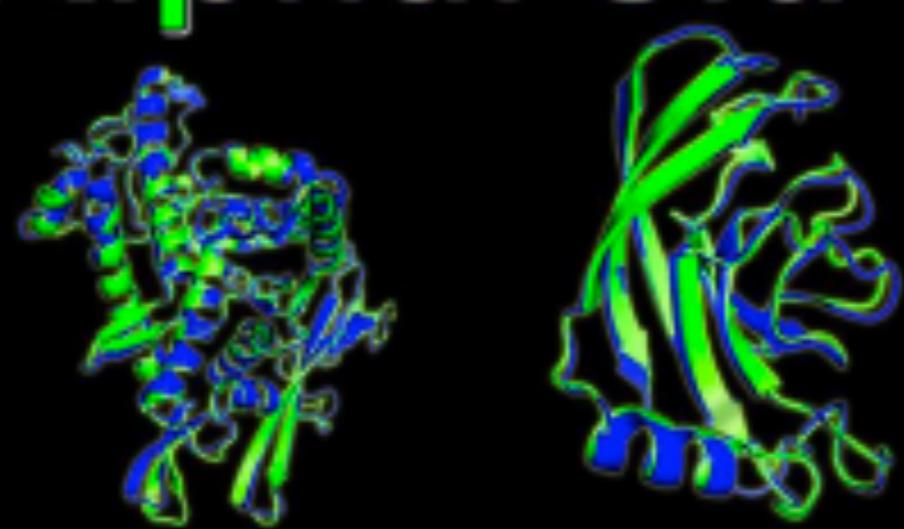
[http://commons.wikimedia.org/wiki/File:Simplified\\_tree.png](http://commons.wikimedia.org/wiki/File:Simplified_tree.png)

# **Research Directions of Bioinformatics —**

## **Protein Structure Prediction**



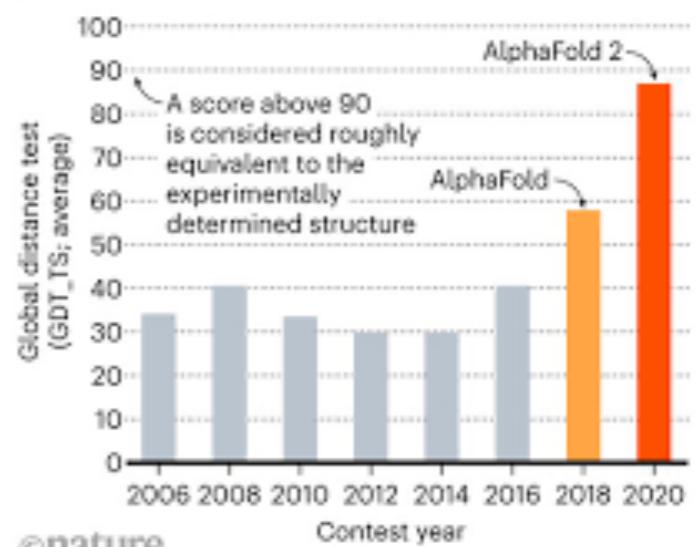
# Google DeepMind's **AlphaFold 2**



## AI Breakthrough in Biology

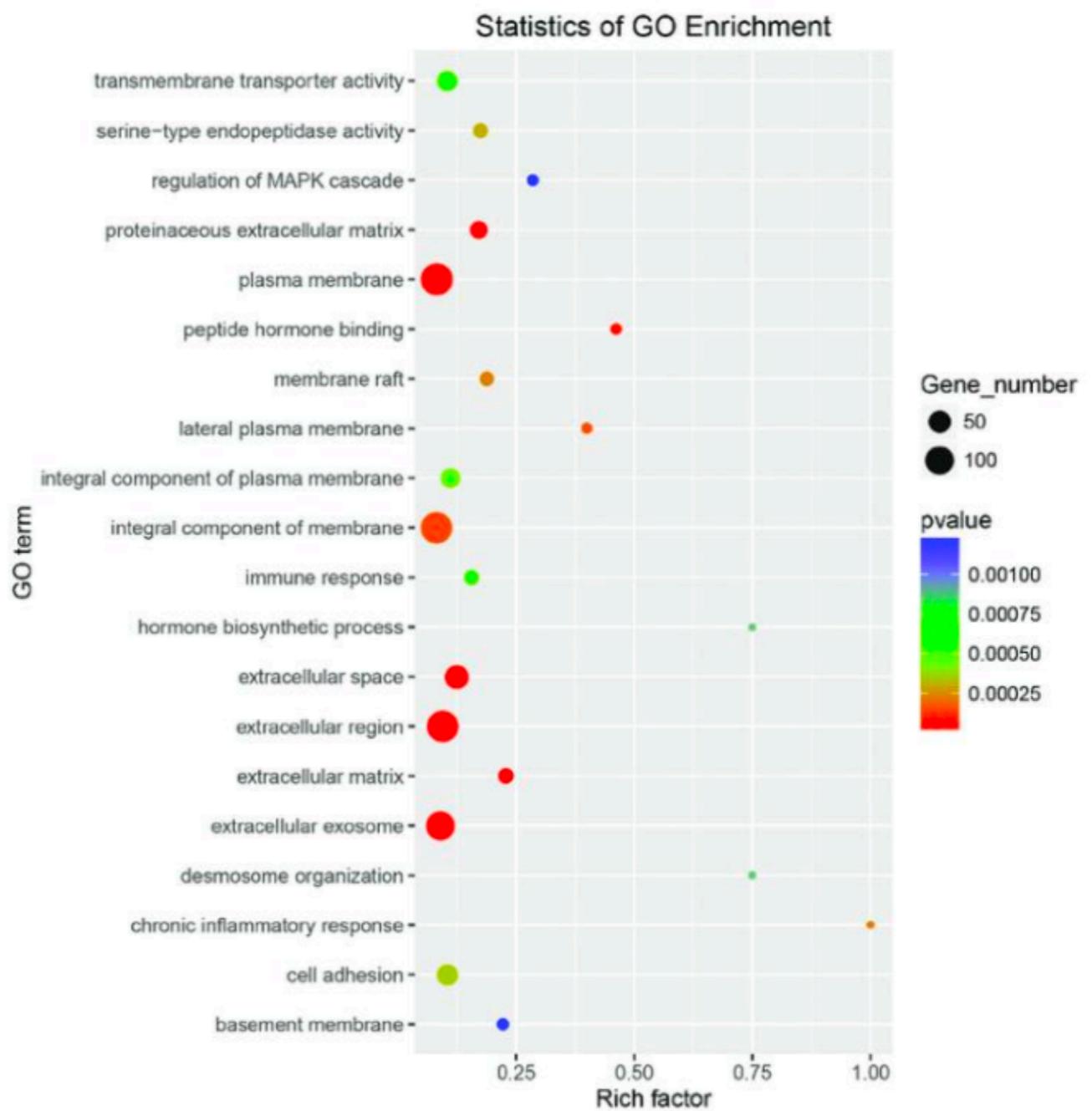
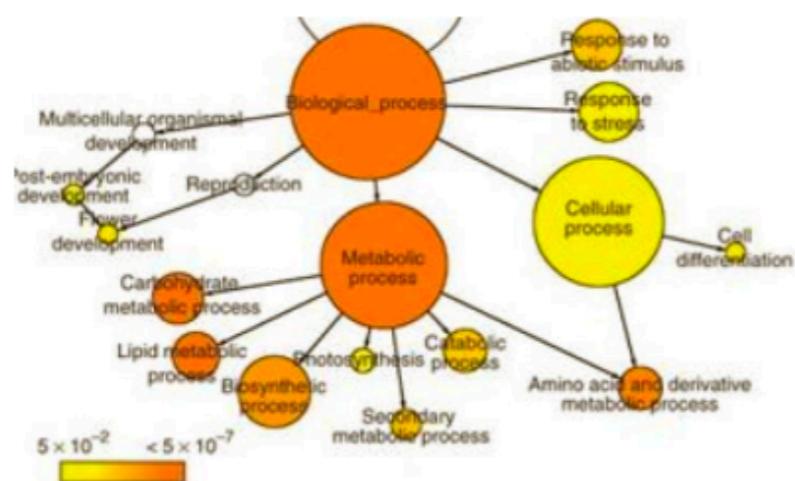
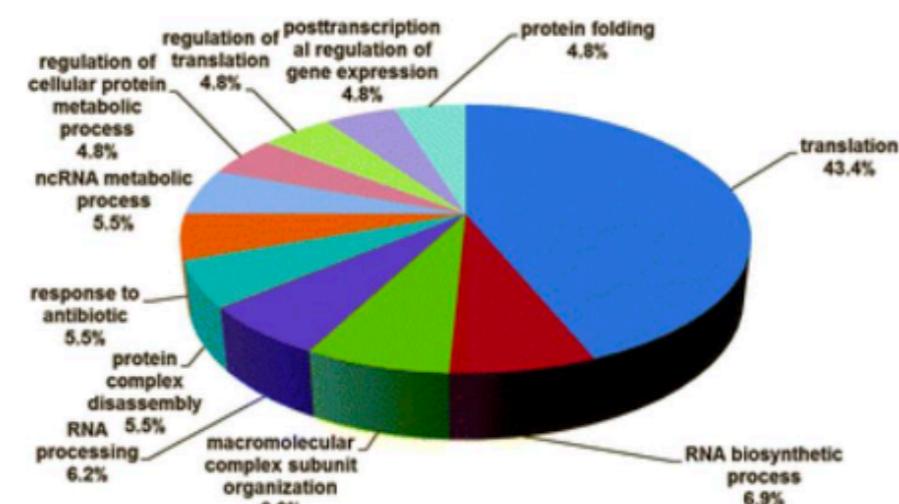
### STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



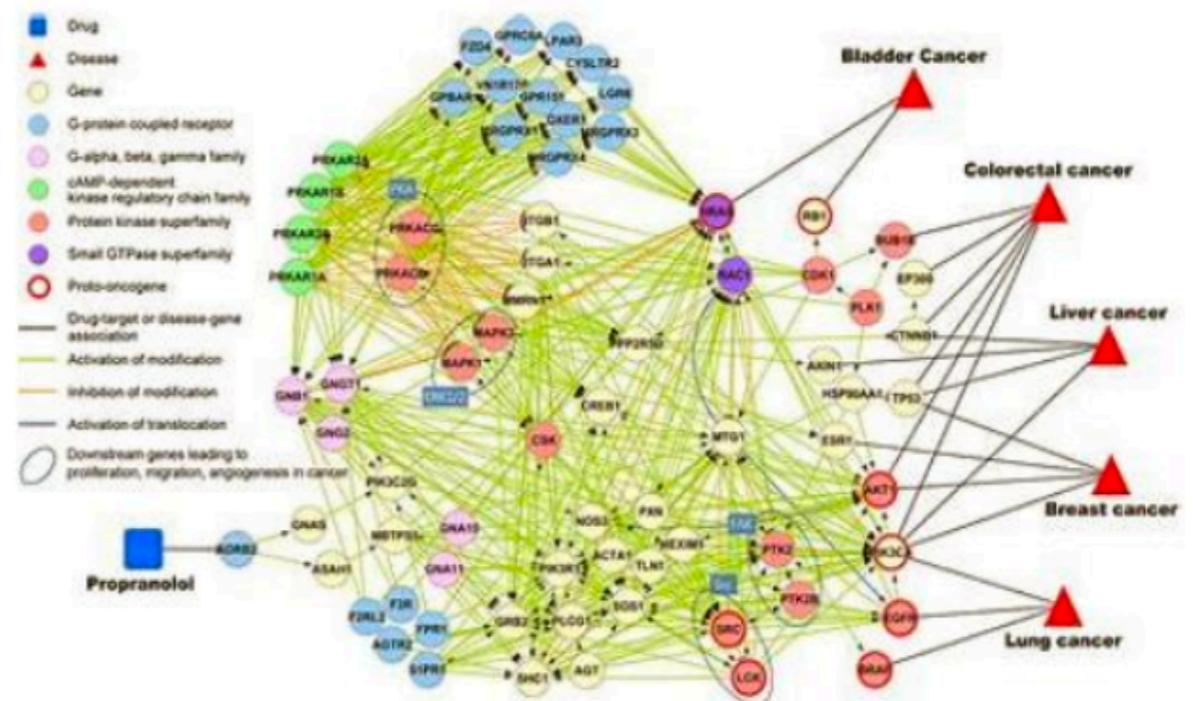
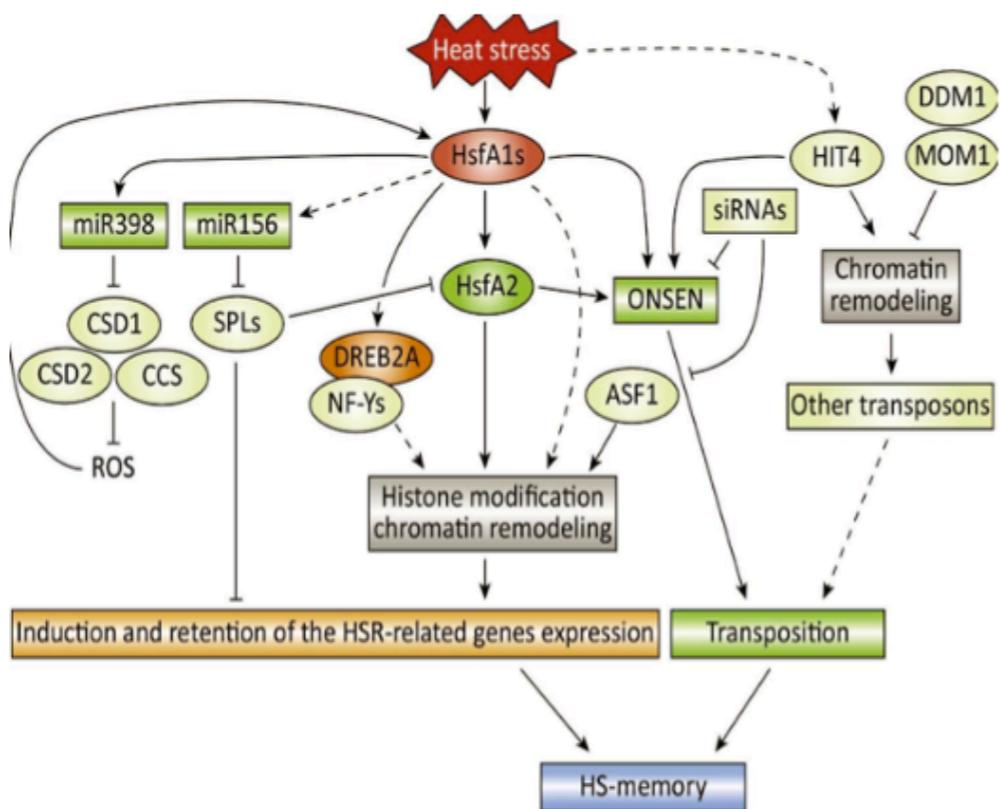
©nature

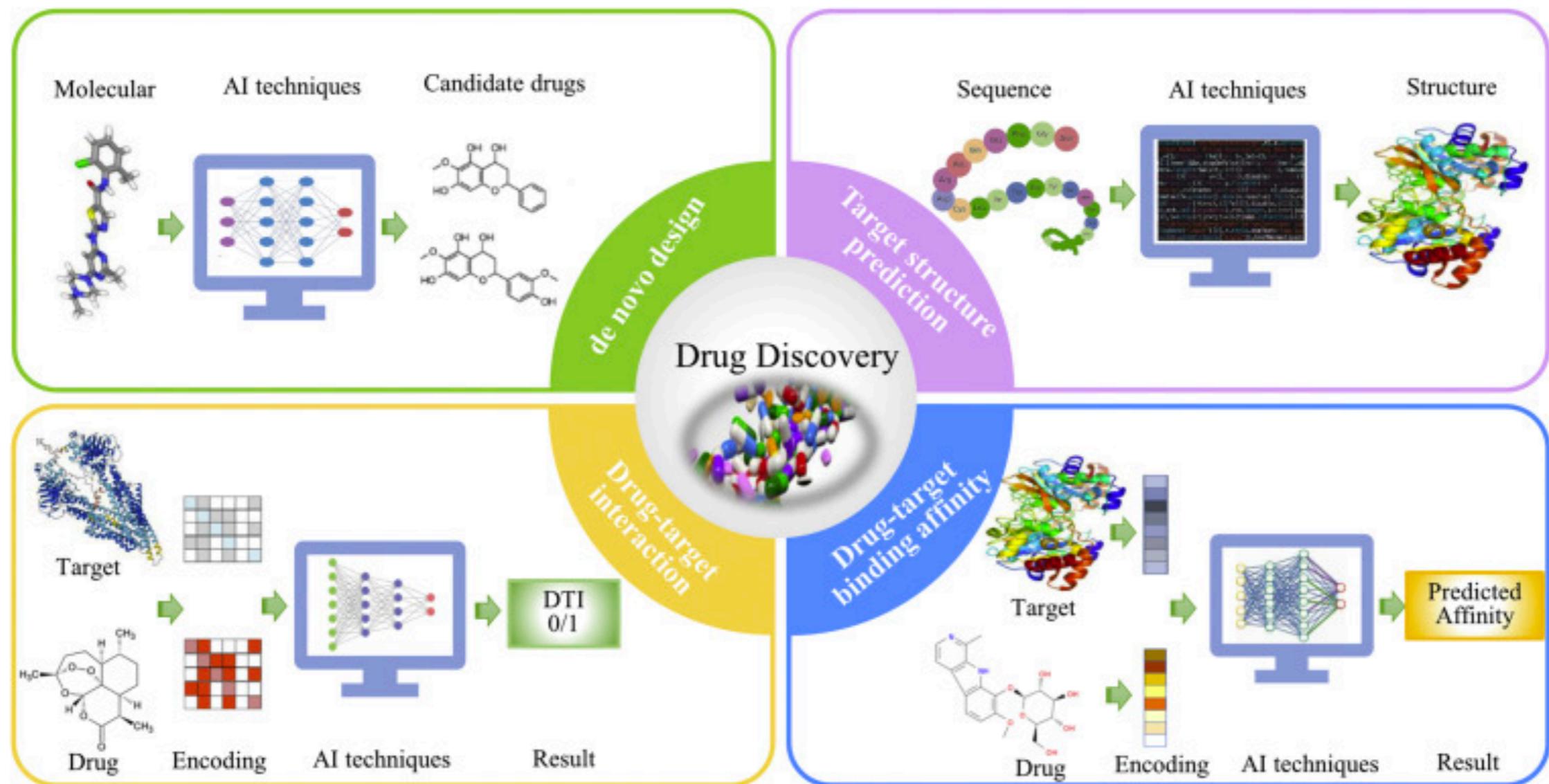
# Research Directions of Bioinformatics — Gene Function Prediction



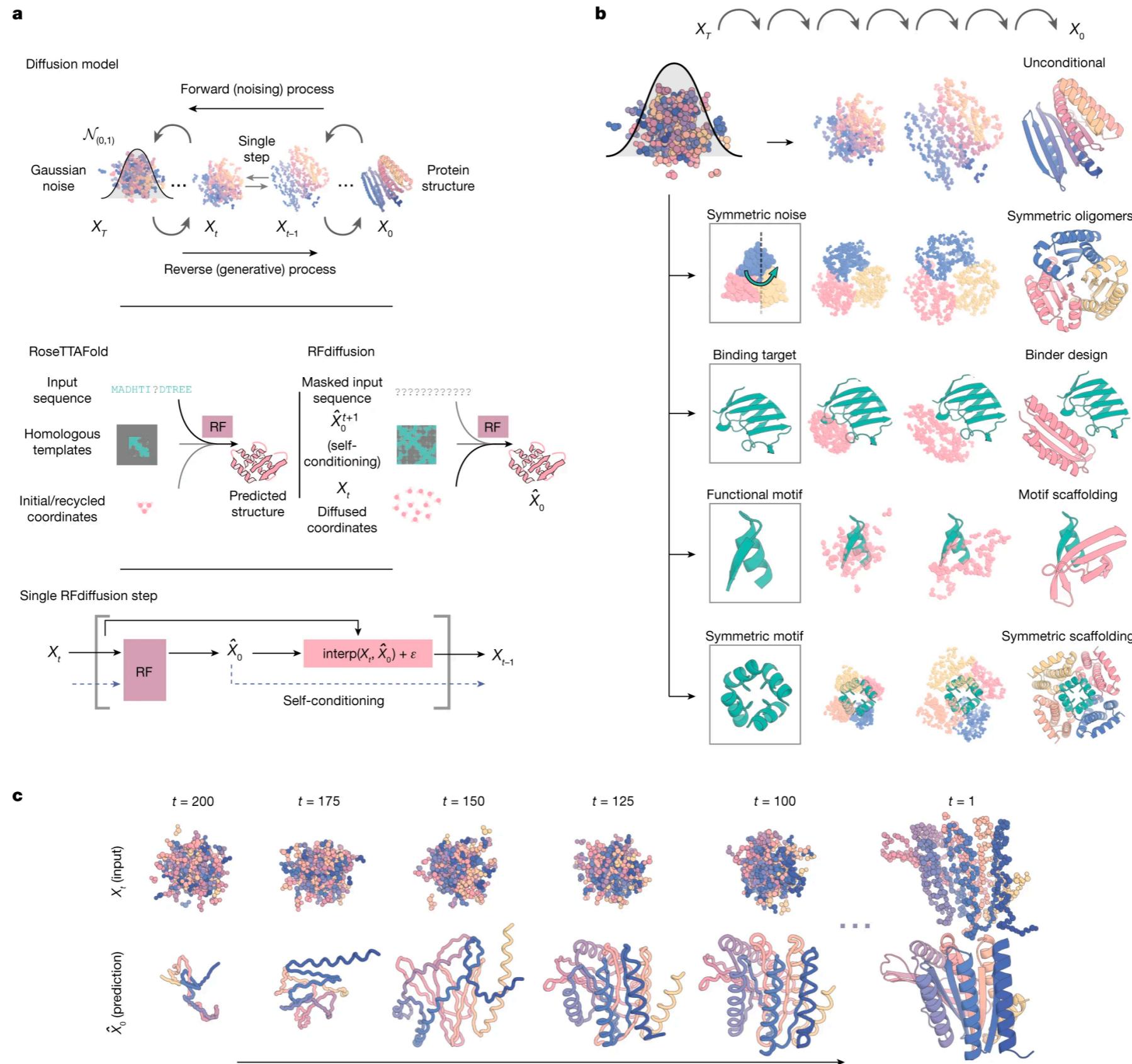
# **Research Directions of Bioinformatics —**

## **Regulatory Network Analysis**





# De novo design of protein structure and function with RFdiffusion





THANK YOU