أكاديمية كاوست
KAUST ACADEMY

Introduction to Bioinformatics

Day 1: Introduction to Bioinformatics and Biological Resources

12th January 2026

KAUST

Instructors

- Dr. Muhammad Waqas Ali (Oxford, OBU, UK)

Course Schedule and Content Overview

Day	Content overview
Day 1	Introduction to Bioinformatics and Biological Resources
Day 2	Sequence Alignment and Phylogenetics
Day 3	Genomics, Metagenomics and Transcriptomics
Day 4	Proteomics and Structural Bioinformatics
Day 5	Exam

Course Timing and Duration
Training Sessions: Total of 6 hours,
Morning (9:00 AM -12:00 PM)
Afternoon (2:00 PM - 5:00 PM)
Breaks: 12.:00 PM to 2:00 PM (Prayer and Lunch)
Operational Days: 5 consecutive days, (12th - 16th of Jan. 2026)

Content
Composition: 30% theory + 70% hands on
Targeted Students: Life Sciences + Computer Sciences

Learning Objectives

Bioinformatics Basics:

- Define bioinformatics and its key applications.
- Understand molecular biology fundamentals relevant to bioinformatics.

Database Utilization:

- Navigate and query biological databases like GenBank, PDB, and KEGG using web tools and programming.

Sequence Alignment

- Perform pairwise and multiple sequence alignments.

Phylogenetics:

- Construct and interpret phylogenetic trees.

Genomics and Metagenomics:

- Analyse genomic and metagenomic data

Transcriptomics:

- Conduct RNA-seq analysis and interpret gene expression data.

Proteomics and Structural Bioinformatics:

- Access and Use Proteomics resources, Predict protein structures

Learning Outcomes

Knowledge Appllication:
- Explain bioinformatics concepts and molecular biology processes.

Database Skills:
- Efficiently query and utilize biological databases.

Alignment Proficiency:
- Conduct and interpret sequence alignments.

Phylogenetic Analysis:
- Build and analyse phylogenetic trees.

Genomic and Metagenomic Insights:
- Analyse and interpret genomic and metagenomic data.

Transcriptomic Skills:
- Perform and interpret RNA-seq data analyses.

Proteomics and Structural Analysis:
- Predict, visualize, and analyse protein structures and interactions.

أكاديمية كاوست
KAUST ACADEMY

Day 1: Introduction to Bioinformatics and Biological Resources

Morning Session

KAUST

Outline Day 1

Morning (9-12 pm):
Introduction to Bioinformatics:

- Overview of bioinformatics:
  - Basics of molecular biology relevant to bioinformatics
  - Genomics, Metagenomics, Transcriptomics, Proteomics

Practical Session:
Exploring the central dogma (using R).

Afternoon (2-5 pm):

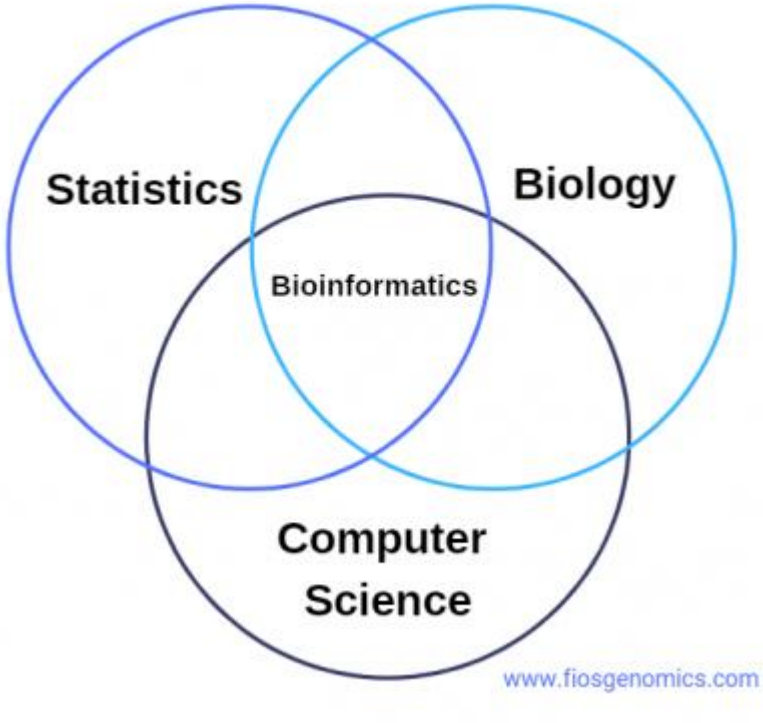
- Biological Resources:
  - Types of biological databases/resources: Sequence, structure, functional.
  - Overview of genomic, transcriptomic, and proteomic databases/resources.

Practical Session:

- Primary Option: Query databases like GenBank, PDB, and KEGG (web-based).
- Alternative: Access and query these databases using Ensembl.

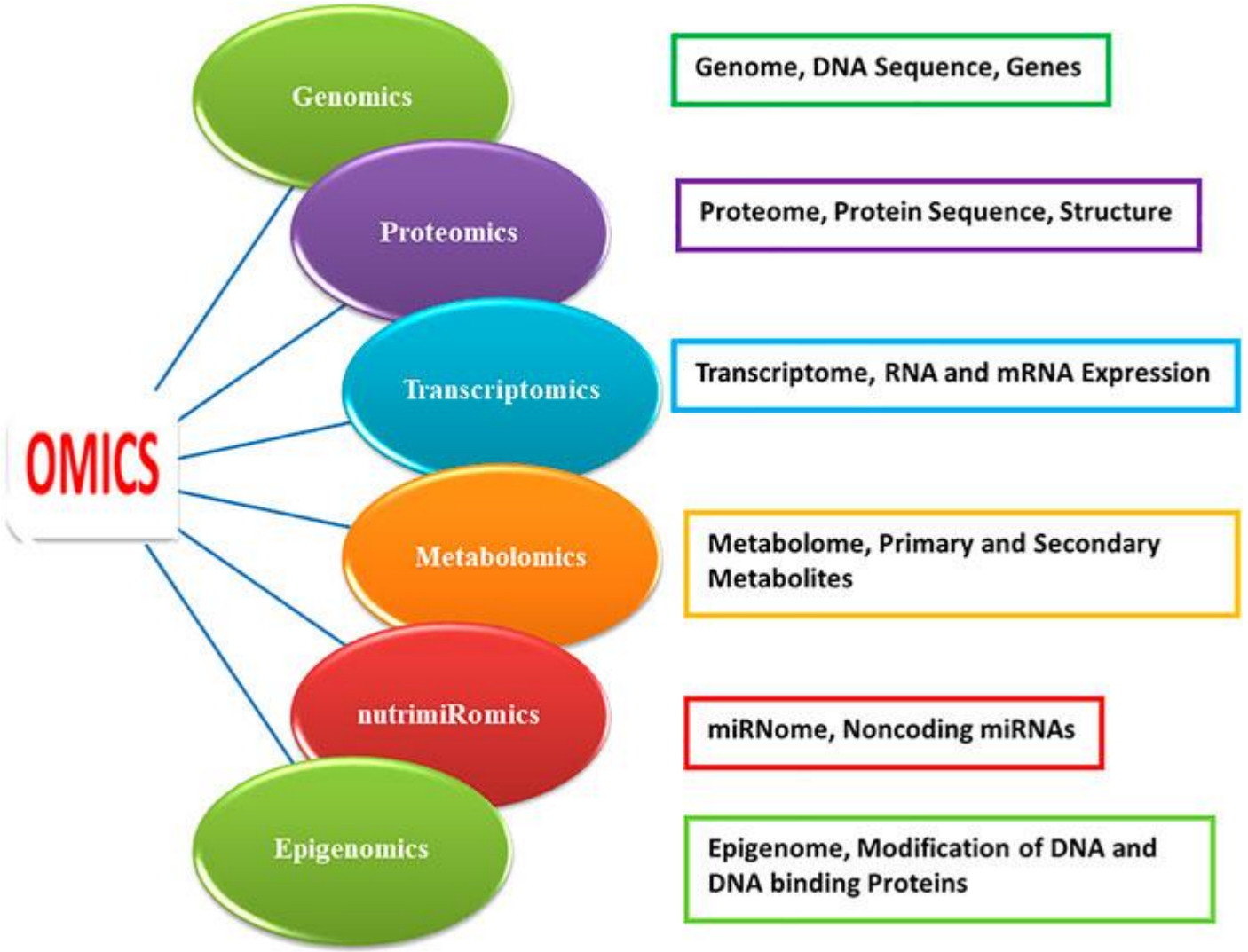
What is Bioinformatics?

Interdisciplinary field that combines life sciences, computer sciences, and statistics to analyse and interpret biological data.



Source: https://www.fiosgenomics.com/what-is-bioinformatics-overview-and-examples/

Key Omics Disciplines in Bioinformatics



Why do we need Bioinformatics ?

1. Key Complexity Factors

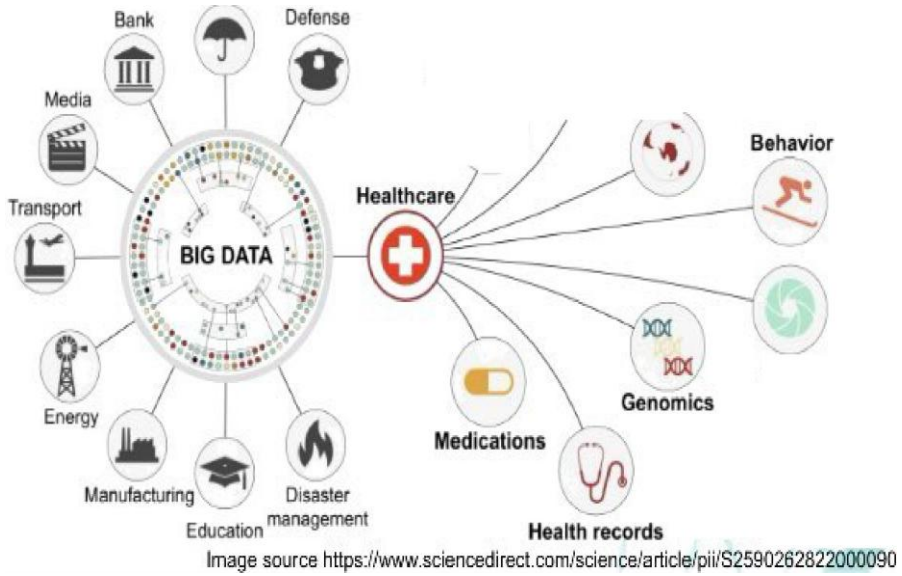
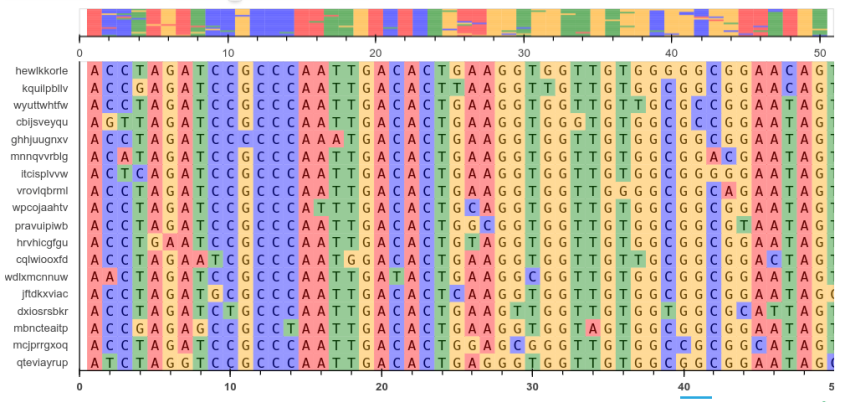


Image source https://www.sciencedirect.com/science/article/pii/S2590262822000090

- **Volume:** Massive datasets from high-throughput technologies
- **Variety:** Different data types (Genomics, proteomics, transcriptomics, and more) and various formats (sequences., structures, networks) require specialized analysis
- **Velocity:** Rapid data generation and updates
- **Variability:** Inherent noise and inconsistencies in biological systems



Why do we need Bioinformatics?

Challenges

Integration: Merging diverse data sources.

interpretation: Extracting meaningful insights.

Scalability: Effidently handling large datasets.

The Role of Bioinformatics

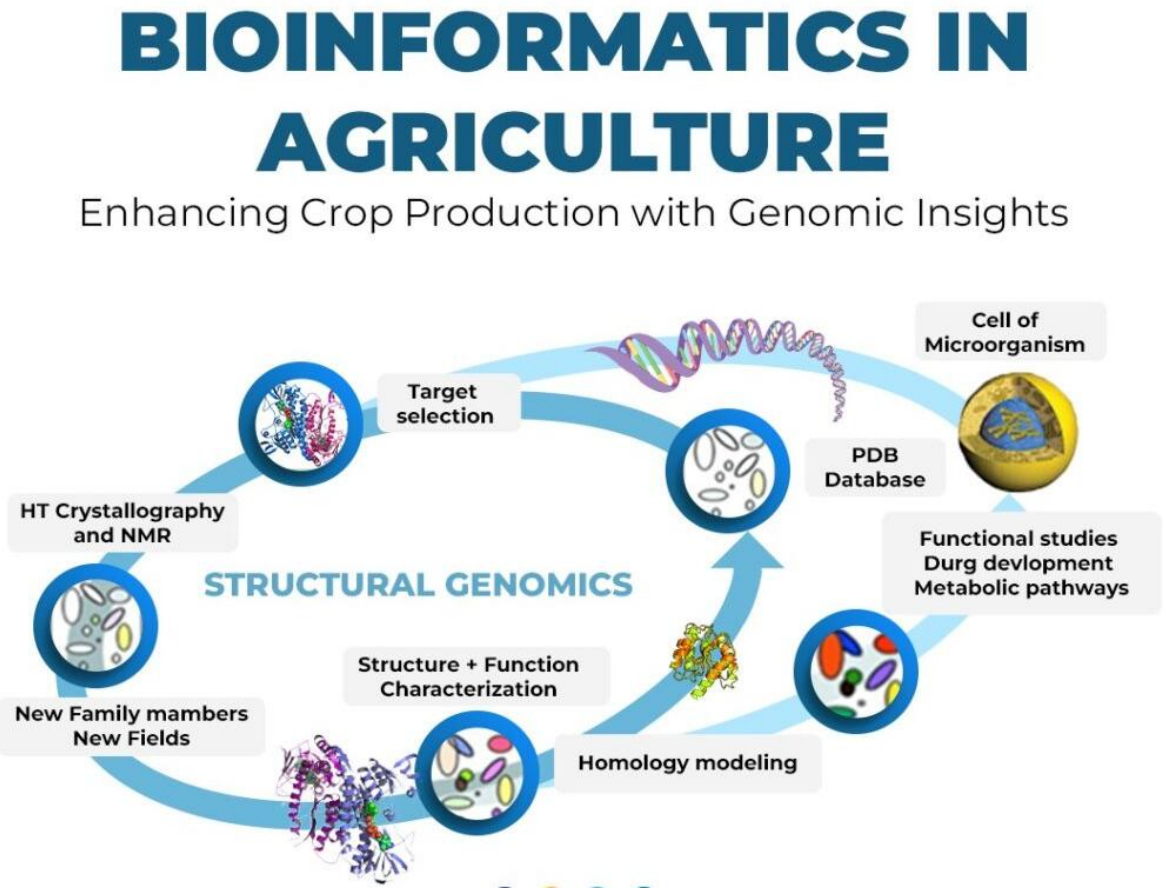
Solution: Computational tools and interdisciplinary approaches to manage and interpret complex biological data.

Applications of Bioinformatics

Biotechnology & Biofuels:
Enhancing enzyme production, Genetically Modified Organisms., and developing efficient biofuels.
Drug Discovery & Preventive Medicine:
Identifying drug targets, designing drug1s, and predicting disease risks.
Gene Therapy & Stem Cell Therapy:
Optimizing gene edit1ing and improving stem cell treatments.

Applications of Bioinformatics

Environmental & Climate Applications:
Bioremediation, studying climate impacts, and engineering resistant crops.
Agriculture & Nutrition:
Improving crop yield, resistance, and nutritional quality.
Microbial & Evolutionary Studies:
Analysing microbial genomes and tracing species evolution.
Forensic & Veterinary Sciences:
DNA analysis for identity verification, criminal investigations, and improving animal health.
Antibiotic Resistance & Insect Resistance:
Studying resistance mechanisms and engineering crops against pests.



أكاديمية كاوست
KAUST ACADEMY

Bioinformatics in Saudi Arabia

KAUST

Saudi Human Resource Genome

Saudi gene hunters comb country's DNA to prevent rare diseases

Research could help prevent disorders that result from marriages between relatives

8 DEC 2016 • BY JOCELYN KAISER

KMAP Platform

KAUST Metagenomic Analysis Platform (KMAP), enabling access to massive analytics of re-annotated metagenomic data

Intikhab Alam, Allan Anthony Kamau, David Kamanda Ngugi, Takashi Gojobori, Carlos M. Duarte & Vladimir B. Bajic

Scientific Reports **11**, Article number: 11511 (2021) | Cite this article

A Publisher Correction to this article was published on 01 July 2021

This article has been updated

Abstract

Exponential rise of metagenomics sequencing is delivering massive functional environmental genomics data. However, this also generates a procedural bottleneck for on-going re-analysis as reference databases grow and methods improve, and analyses need be updated for consistency, which require access to increasingly demanding bioinformatic and computational resources. Here, we present the KAUST Metagenomic Analysis Platform (KMAP), a new

Microbial Habitats

1 Samples

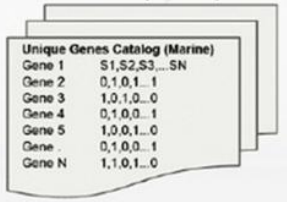
- Metadata [Temp., Salinity]
- Shotgun metagenomic sequencing
- 40,000-60,000 free living or host associated shotgun metagenomic samples available at EBI .

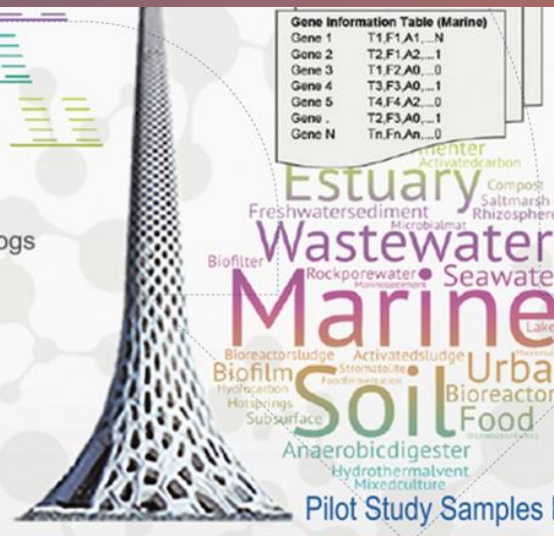
Microbes

Microbes are everywhere, an estimated number is ~1 trillion species.

2 Assemble Metagenomes

- Get full-length genes
- Make unique genes catalogs & map sample reads





Pilot Study Samples Habitats

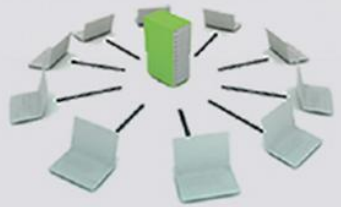
4 Data Sharing

Shaheen and KMAP were used to produce Pilot Study: 40 Gene Catalogs, 275 million genes.

+ KMAP

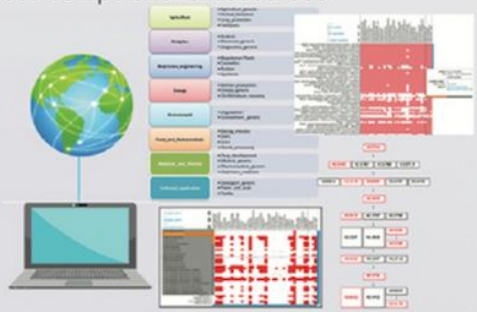
Research Groups

Gene information tables (AAMG TSV) available for research organizations with advanced computational resources and skills.



Individuals

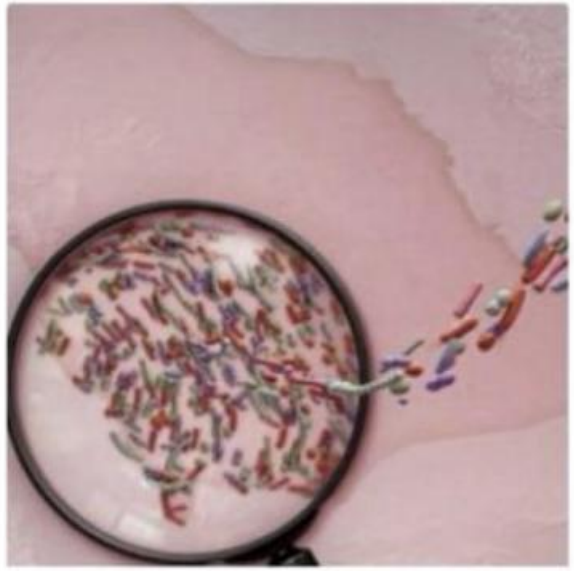
Graphic User Interface access to indexed gene information tables for browsing and comparisons. Available for individuals with less computational resources.



أكاديمية كاوست
KAUST ACADEMY

KAUST Smart-Health Initiative

Health and Wellness: Accelerating Impact in KSA



04 December, 2023

The weird and wonderful world of Saudi Arabia's microbiomes

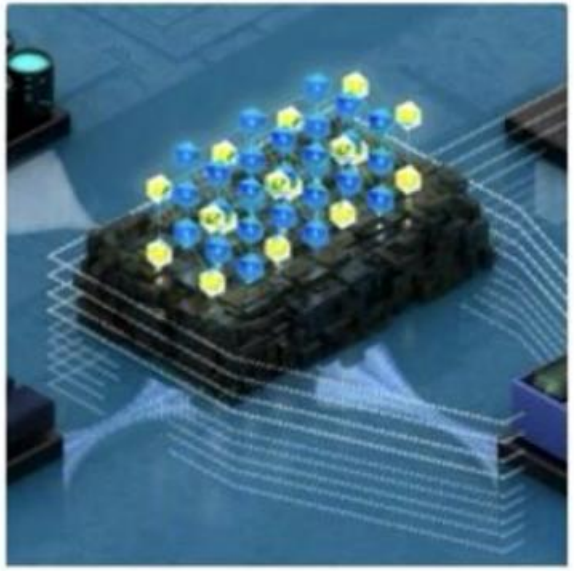
Saudi Arabia is home to microbial communities that can boost coral reef health, sequester carbon in the desert and enable plants to survive in the harshest environments — and that is just the beginning.

03 December, 2023

The first ride across Saudi Arabia on a hand bike

To showcase the remarkable potential of people with physical disabilities, KAUST Professor Matteo Parsani will travel from the east to the west of Saudi Arabia by hand-bicycle.

Read more



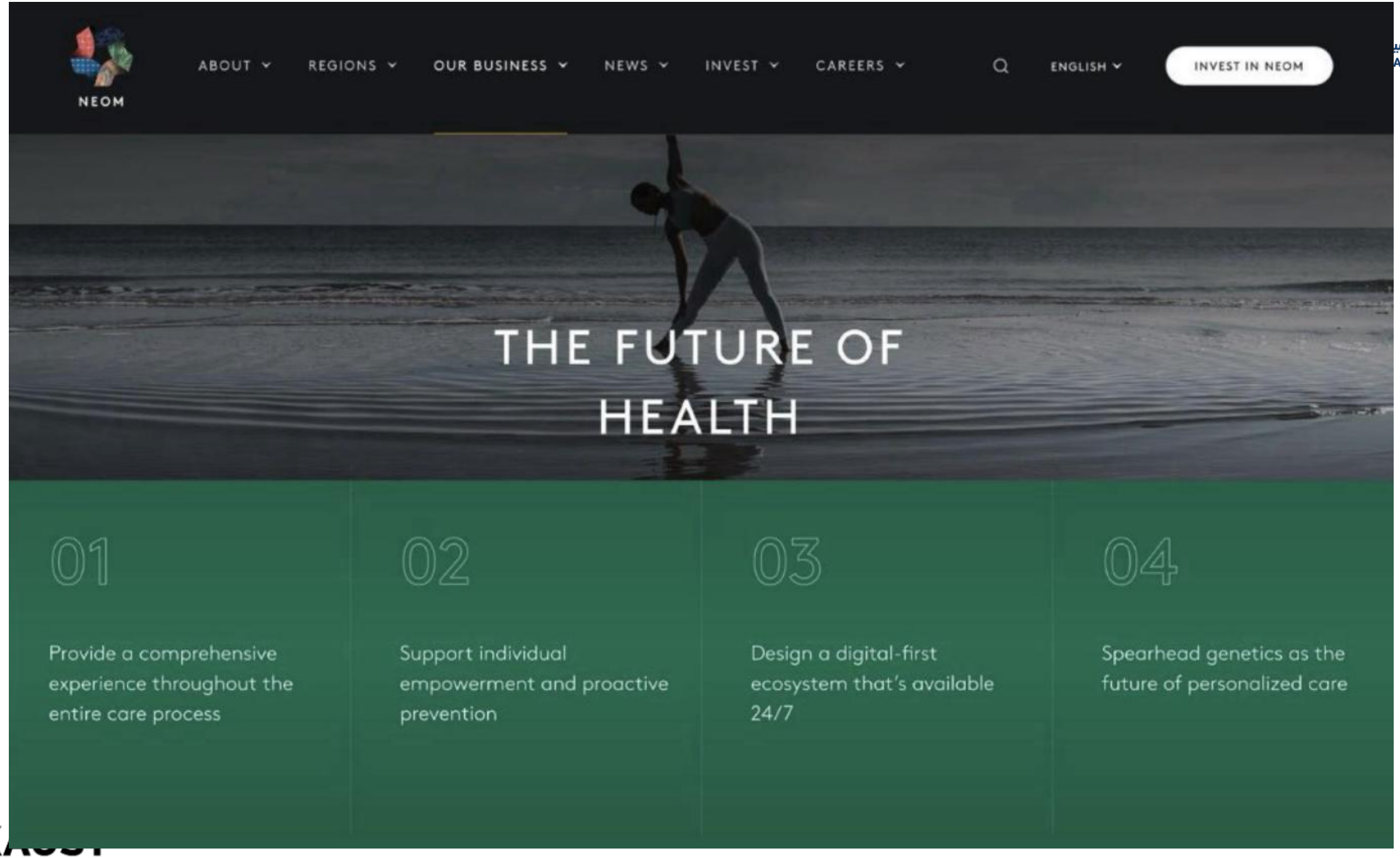
22 November, 2023

Safeguarding the right to be forgotten

An open-source software can help align artificial intelligence applications in healthcare with data privacy regulations.

Read more

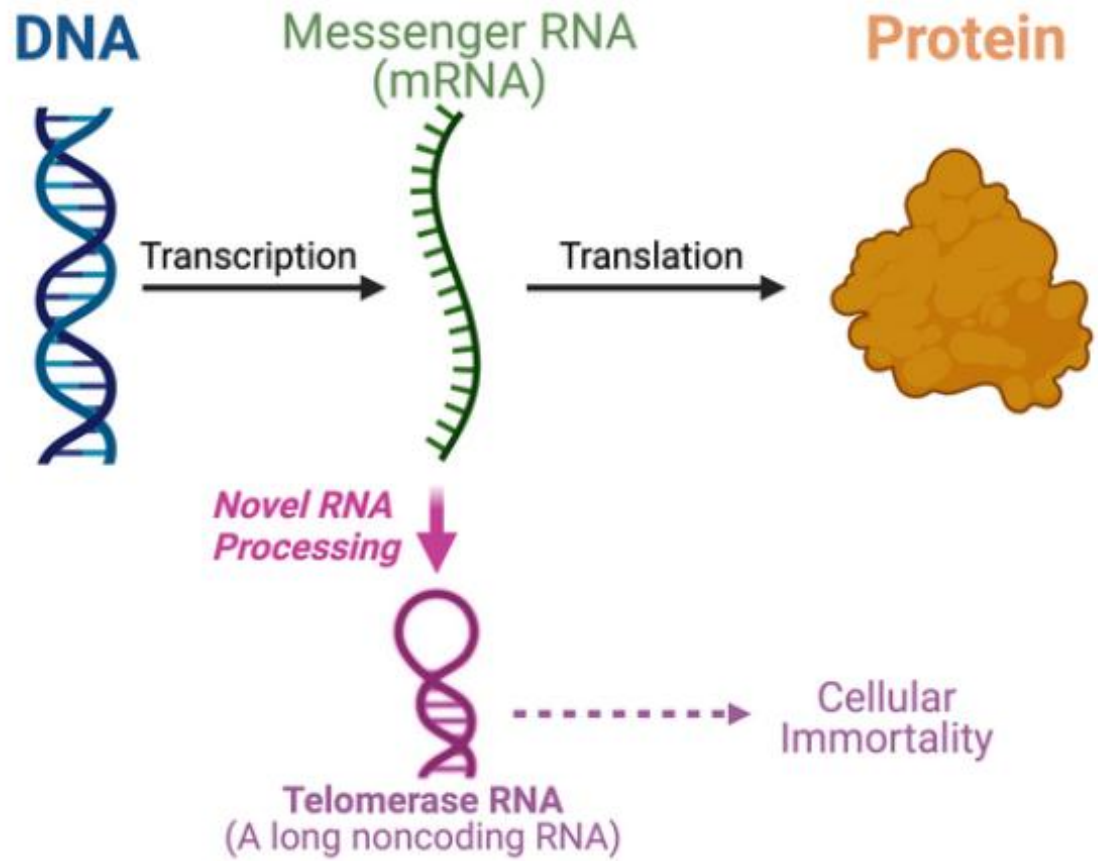
KAUST


NEOM
ABOUT
REGIONS
OUR BUSINESS
NEWS
INVEST
CAREERS
ENGLISH
INVEST IN NEOM
THE FUTURE OF HEALTH
01
Provide a comprehensive experience throughout the entire care process
02
Support individual empowerment and proactive prevention
03
Design a digital-first ecosystem that's available 24/7
04
Spearhead genetics as the future of personalized care

What is Molecular Biology?

The study of structure and function of nucleic acids and proteins of biological molecules to understand the molecular mechanisms underlying various biological processes

Basis of Molecular Biology: The Central Dogma

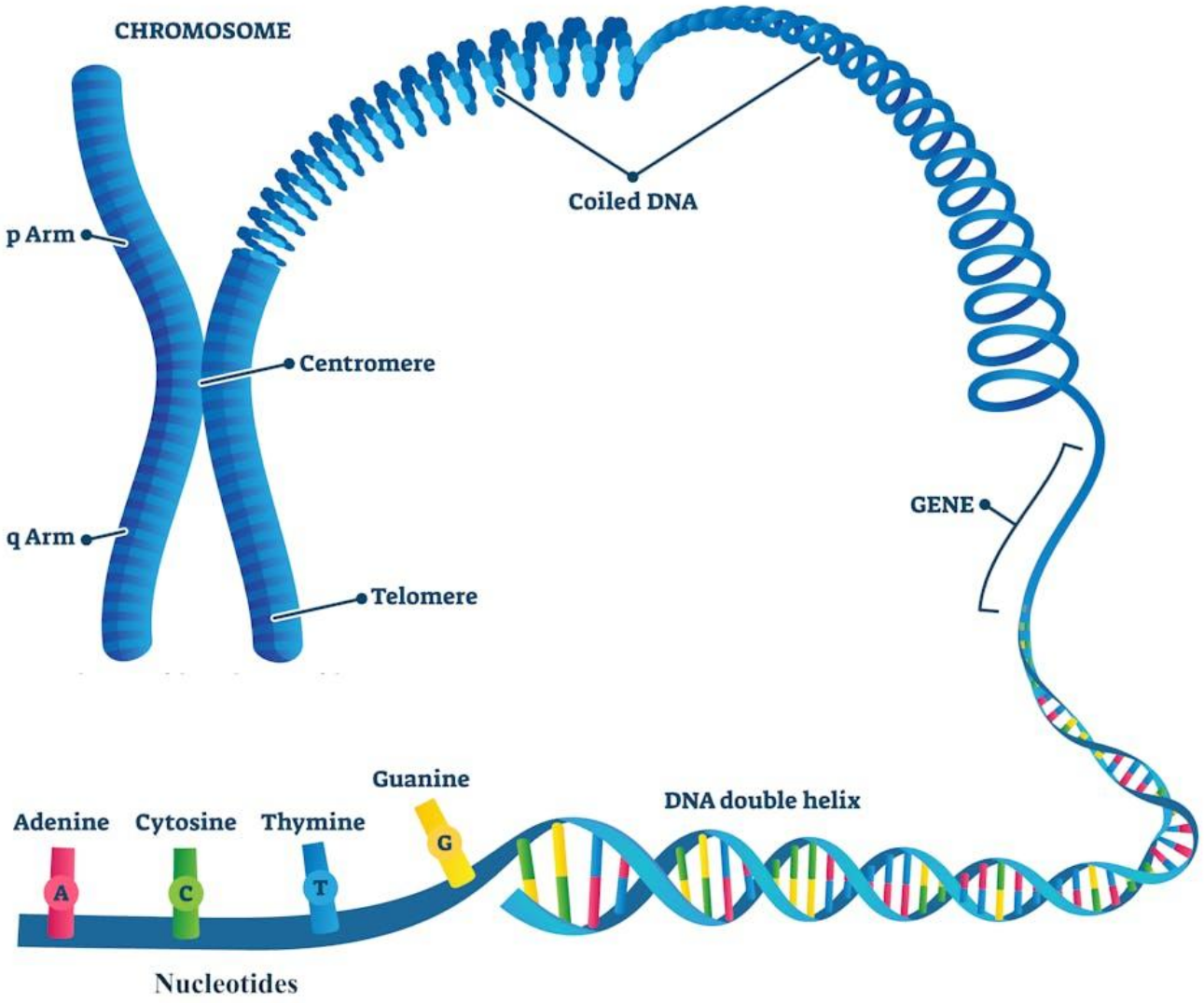


Human Genome

Chromosomes:
Humans have 46 chromosomes (22 pairs of autosomes pairs + 1 pair of sex chromosomes), containing1our genetic information.

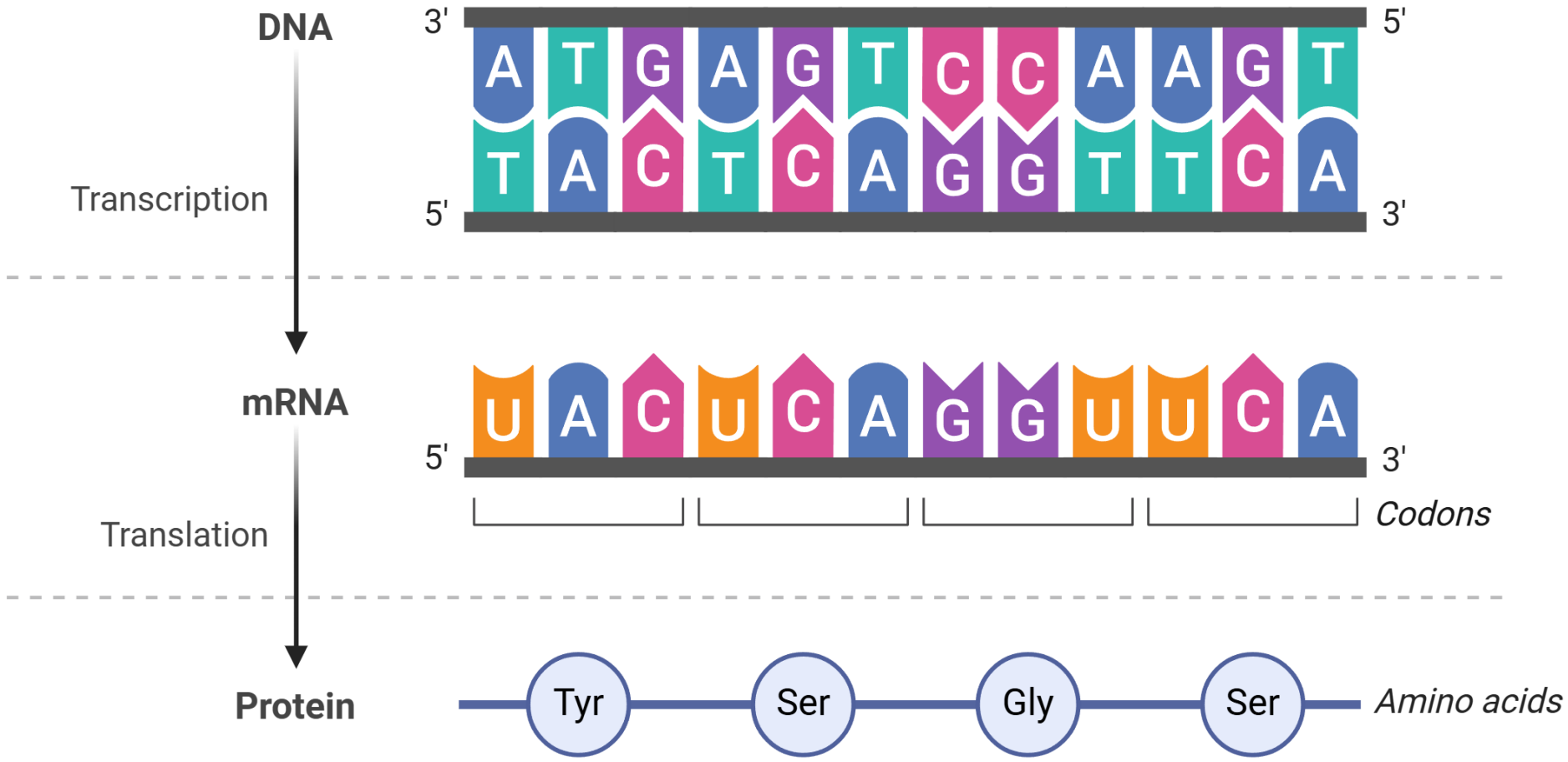
Genes:
Specific segments of DNA that code for proteins, which perform various functions in the body.

Nucleotide:
It is the building block of DNA and RNA, made up of a sugar, phosphate, and a nitrogenous base.

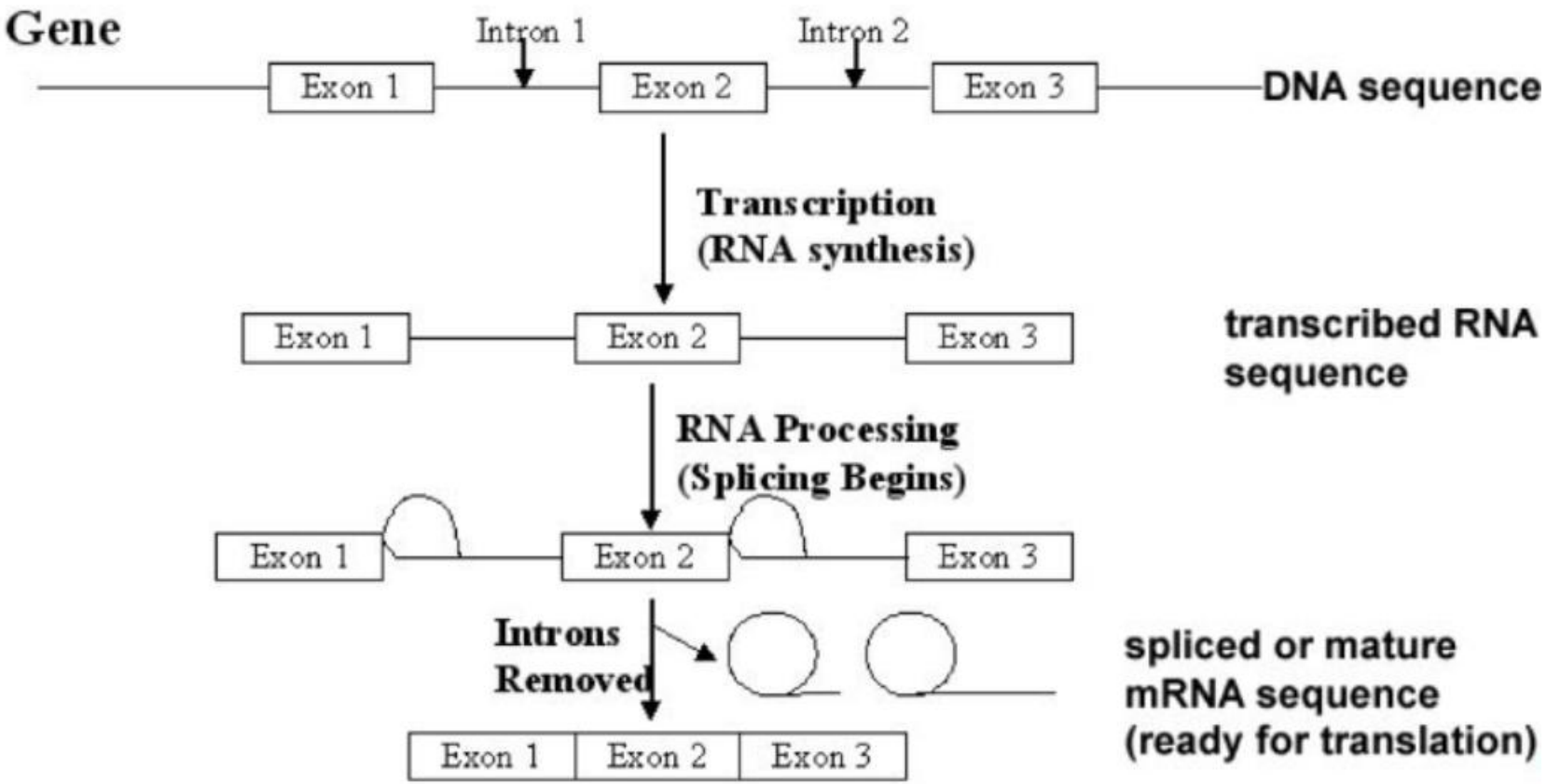


KAUST

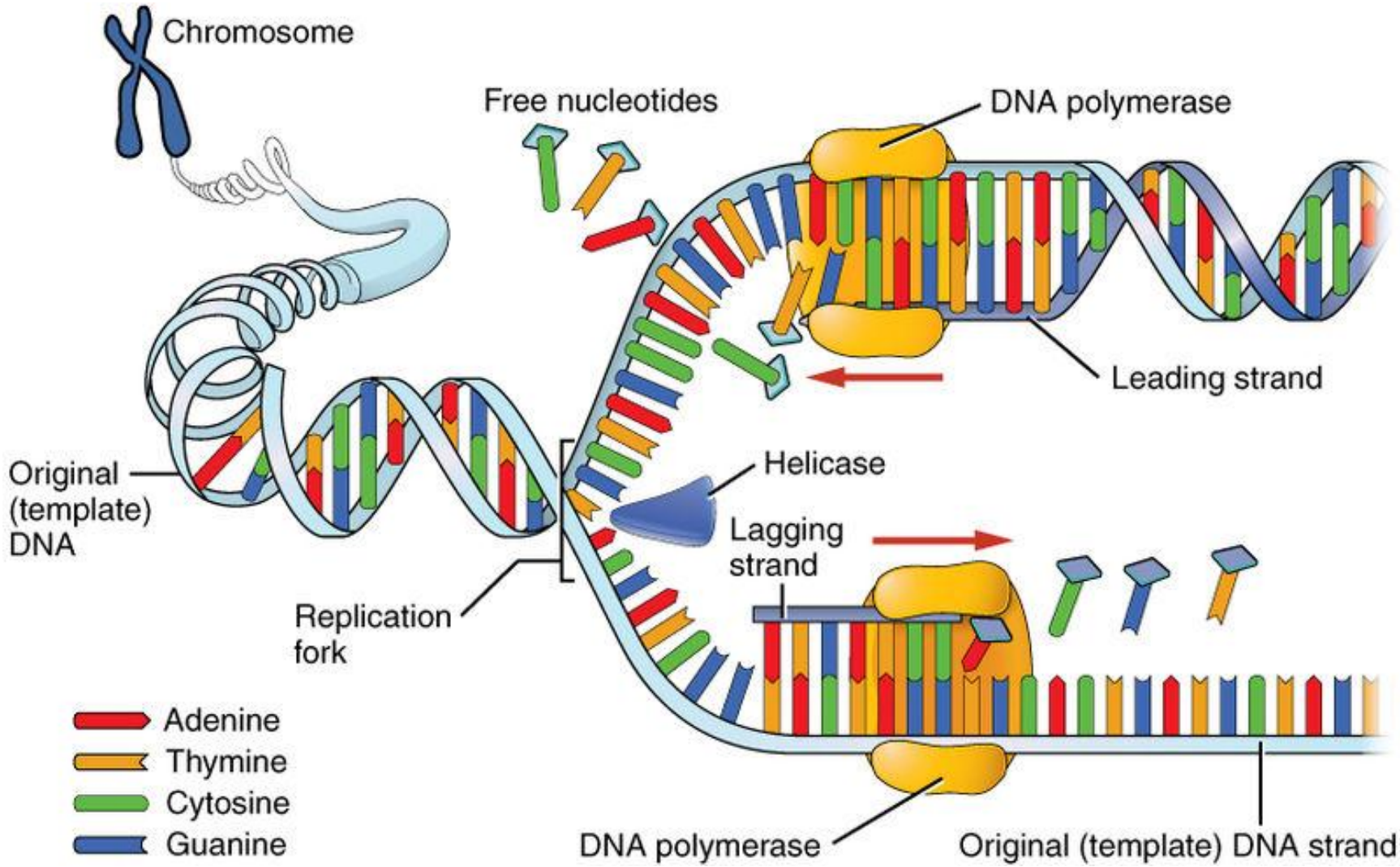
The Central Dogma Main Process



Exons (coding) and Introns (non-coding)

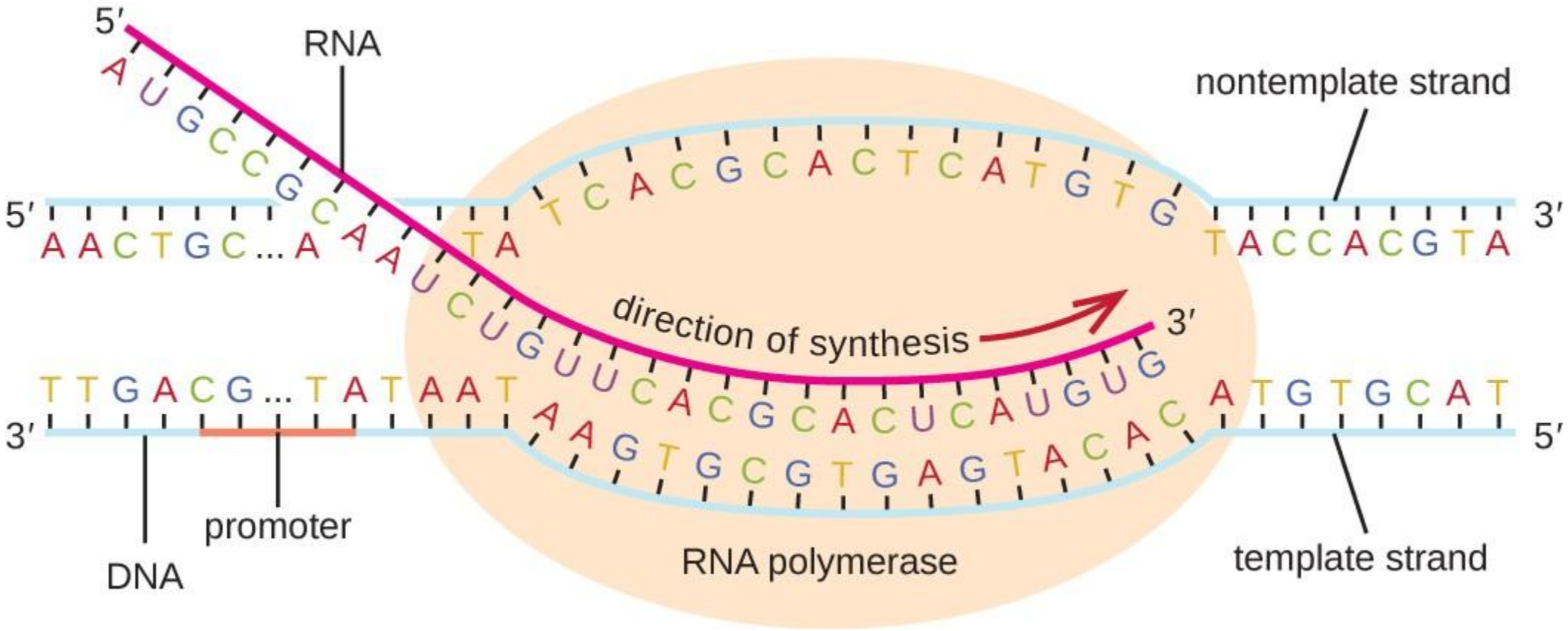


DNA Replication and Repair

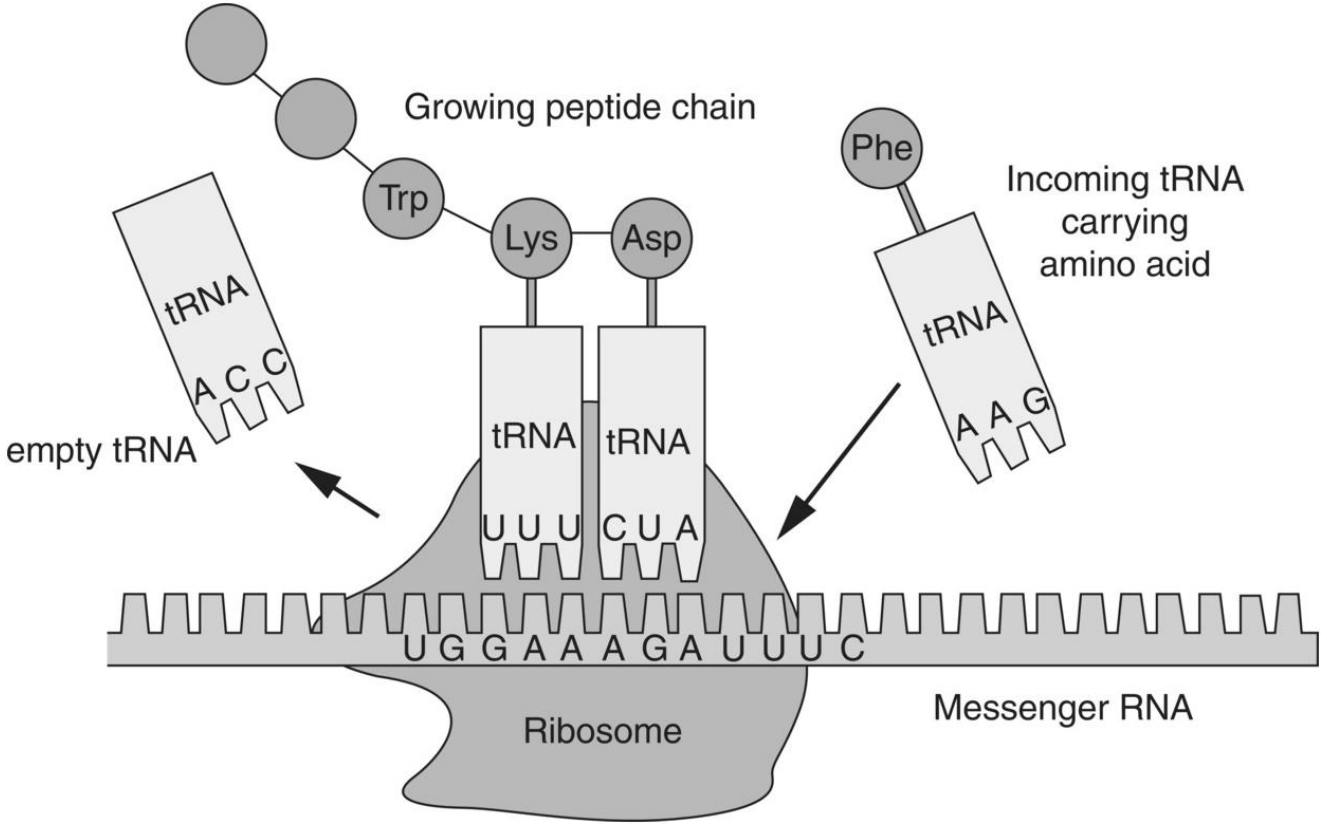


Transcription

Genetic information encoded in DNA is transcribed into RNA molecules, particularly messenger RNA (mRNA),which serves as a template for protein synthesis.

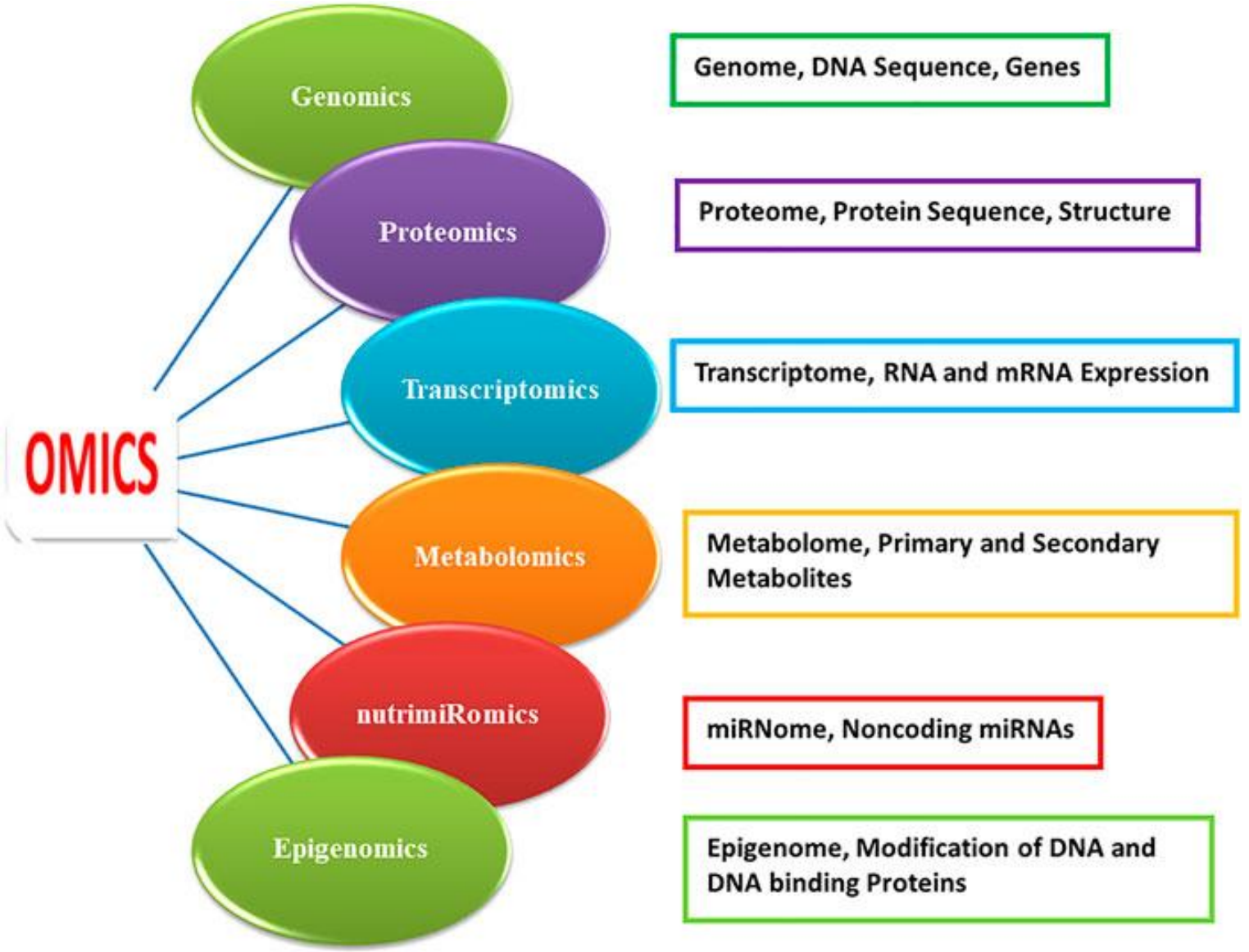


Translation

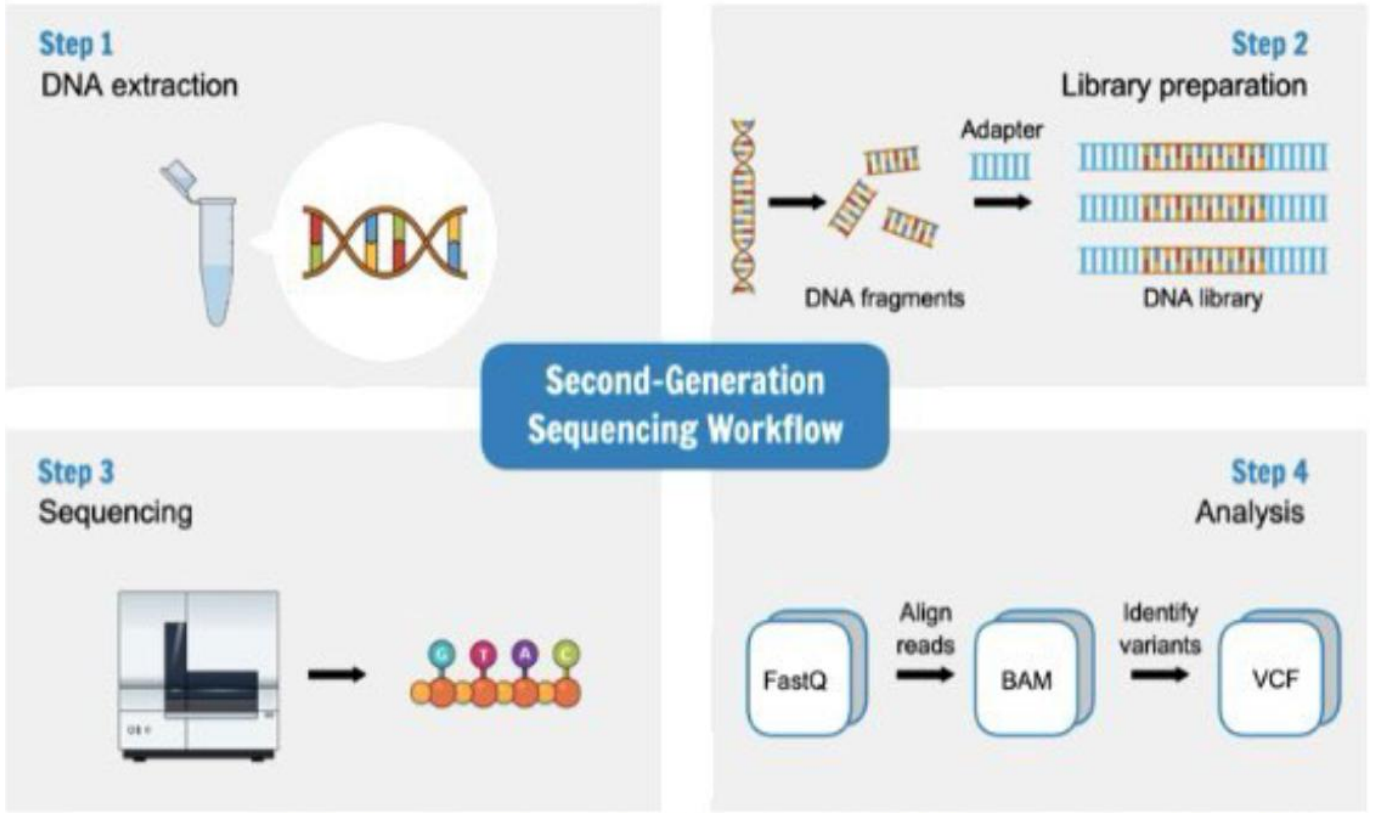


The process by which the information in mRNA is used to synthesize proteins, involving the interaction between mRNA, ribosomes, transfer RNA (tRNA), and amino acids.

Key Omics Disciplines in Bioinformatics



Genomics: Exploring DNA and Genetic Variations



- **DNA Sequencing:**
  - Methods: Sanger sequencing (experimental), Next Generation Sequencing (NGS) (Computational)
  - Applications: Whole-genome sequencing, targeted sequencing

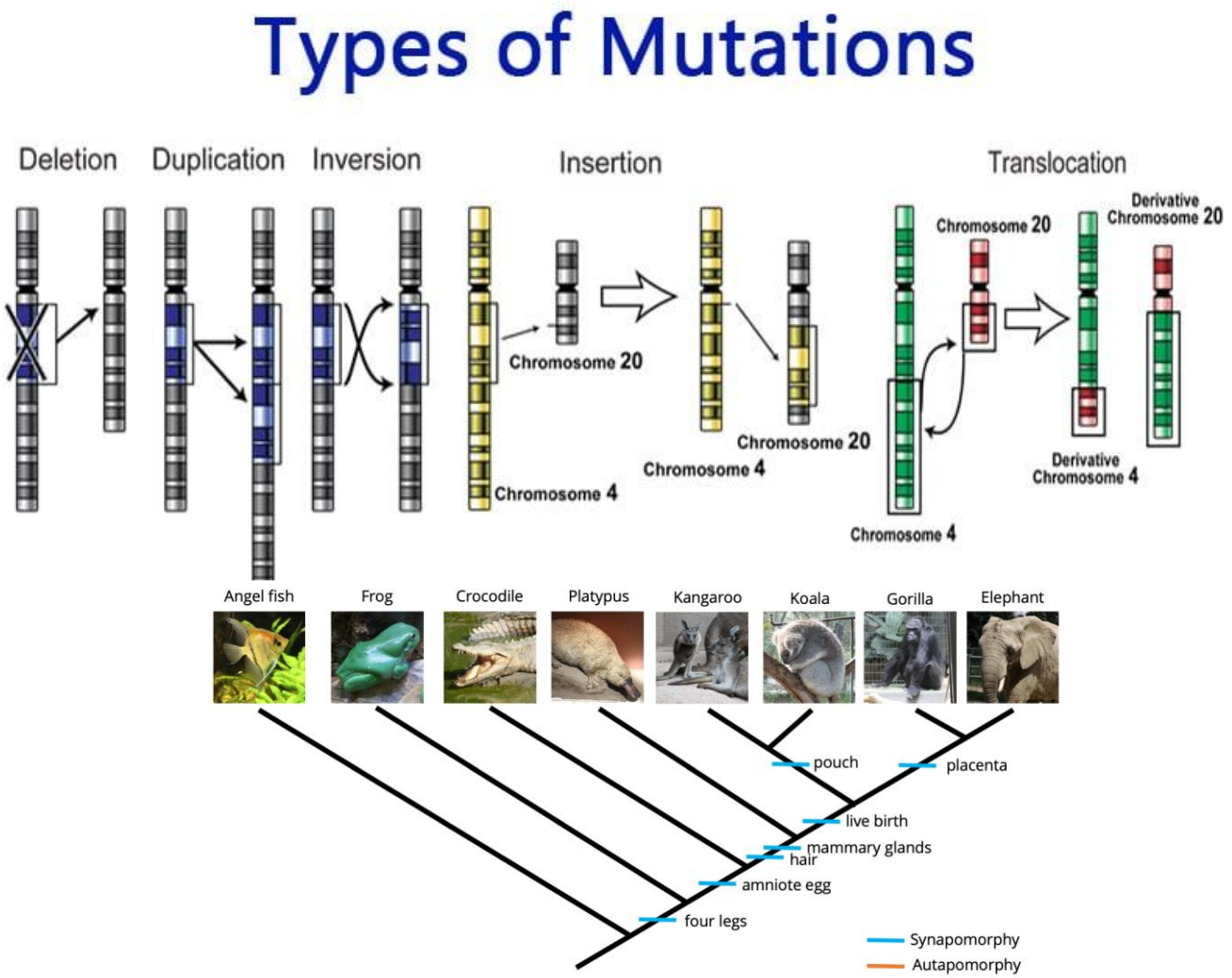
Genomics: Exploring DNA and Genetic Variations

Genetic Variation and Evolution:

Mutations: Types (e.g., point mutations, insertions, deletions)

Linkage Disequilibrium: measures how often genetic variants are inherited together more than expected by chance due to their proximity on the same chromosome.

Phylogenetic Analysis: Tree construction, evolutionary relationships



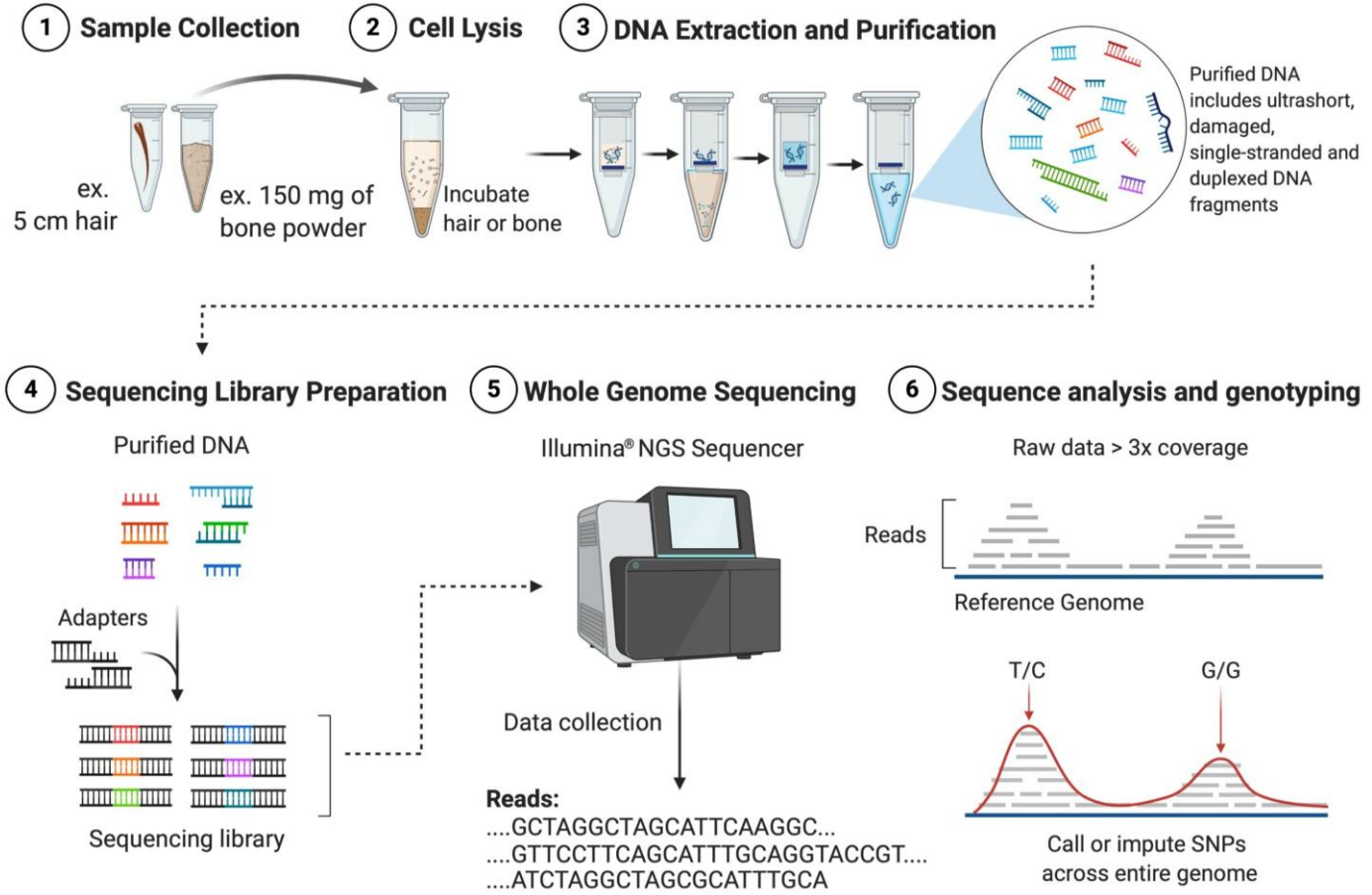
MetaGenomics: Exploring Microbial Communities

Analysis of Genetic-Material from Environmental Samples:

Techniques: DNA extraction, sequencing of environmental samples

Applications:

Studying microbial diversity, functional capabilities

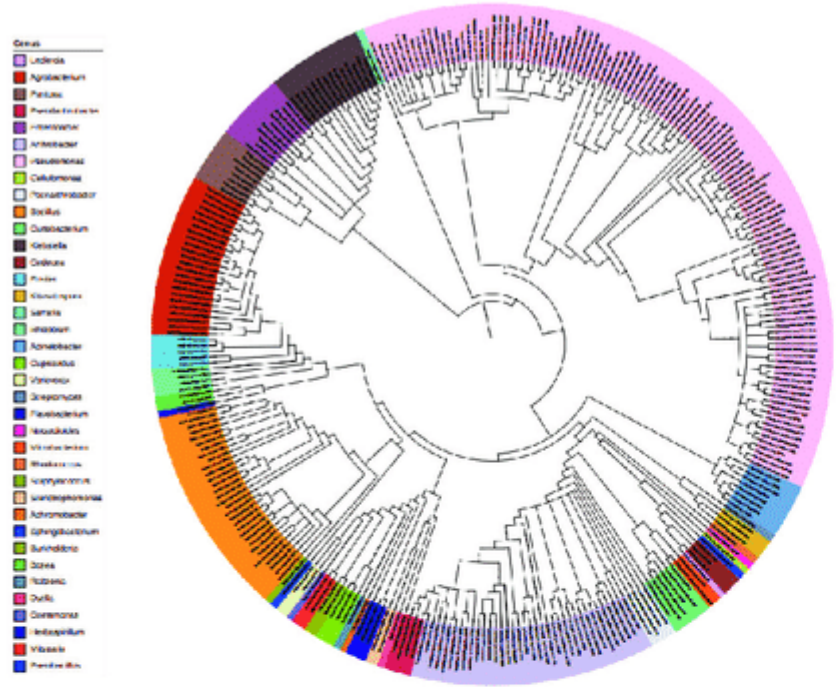


MetaGenomics: Exploring Microbial Communities

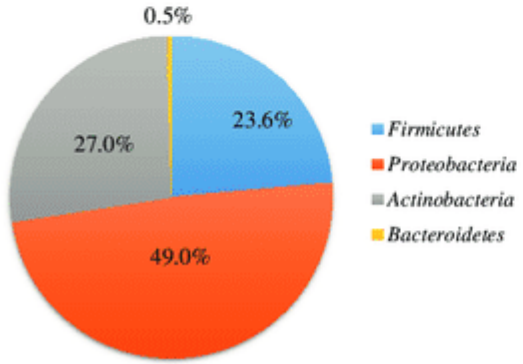
Microbial Diversity and Community Structure

- Diversity: Richness, evenness of microbial species
- Structure: Community composition, interactions

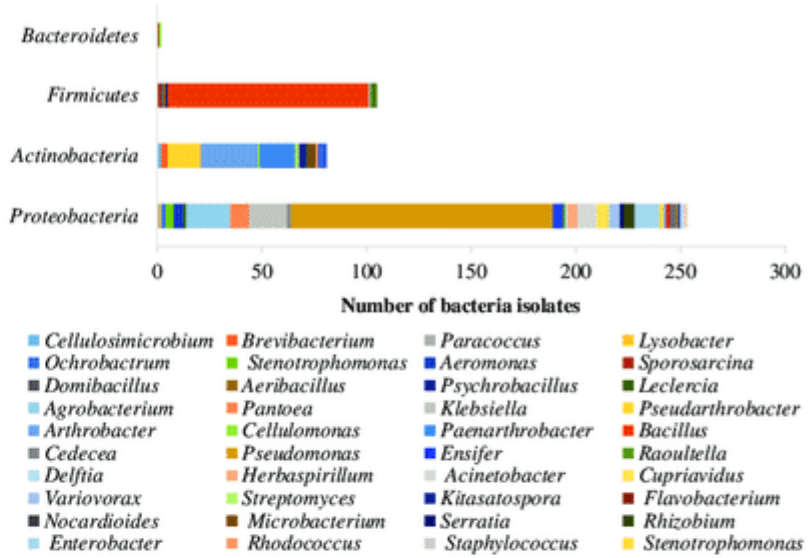
A



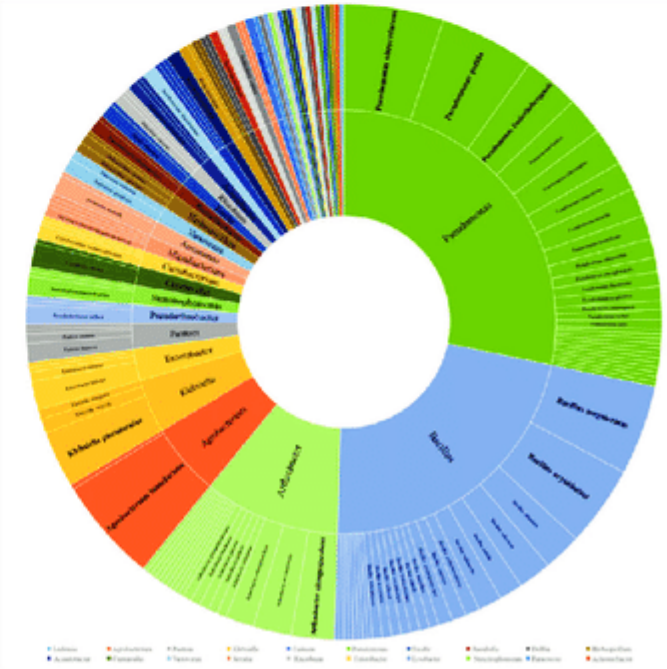
B



C



D

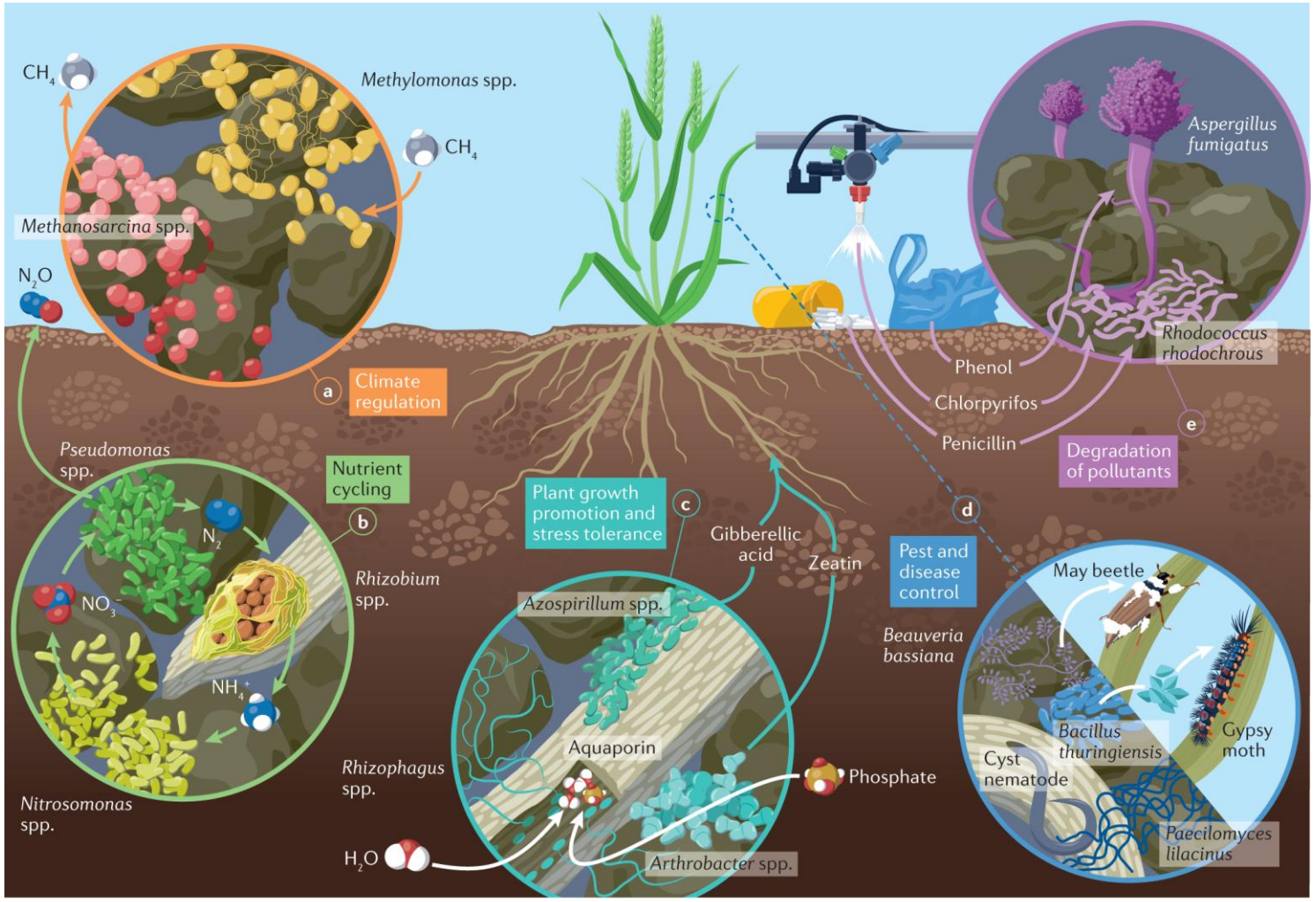


MetaGenomics: Exploring Microbial Communities

Functional Profiling:

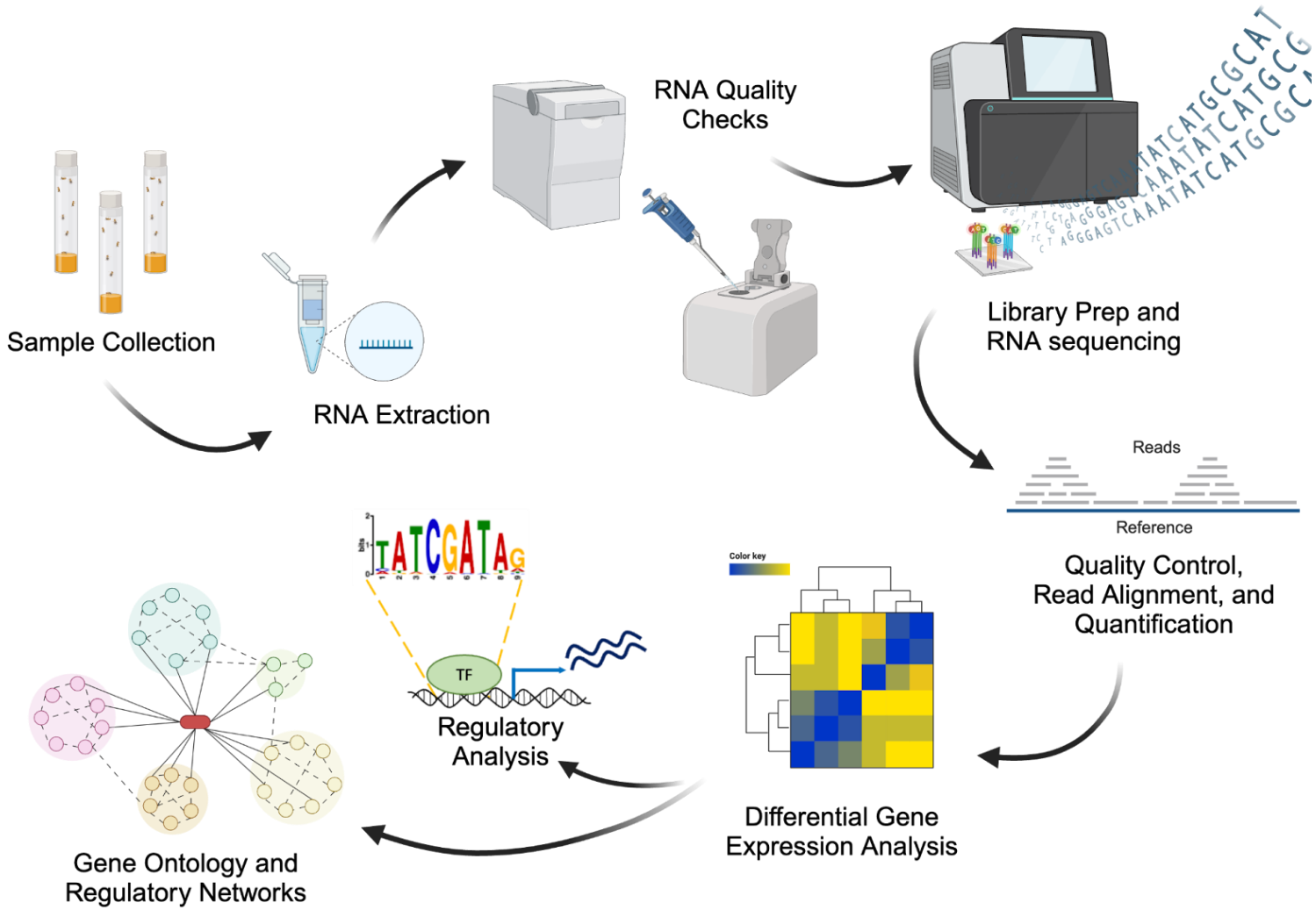
Identifying Functions of microbial communities, potential roles in ecosystems.

Example: A metagenomics study of soil microbiomes revealed that the combination of Nitrosomonas (ammonia-oxidizing bacteria) and Nitrobacter (nitrite-oxidizing bacteria) works together to complete the nitrogen cycle. This collaboration converts ammonia into nitrate, enriching the soil for plant growth.



Transcriptomics: Understanding Gene Expression

RNA Sequencing (RNA-seq):
Technique: Sequencing of RNA to measure gene expression
Sample Applications :
i. Gene expression analysis
ii. Transcriptome profiling
(Mapping the full range of RNA molecules expressed in a cell or organism)



Transcriptomics: Understanding Gene Expression

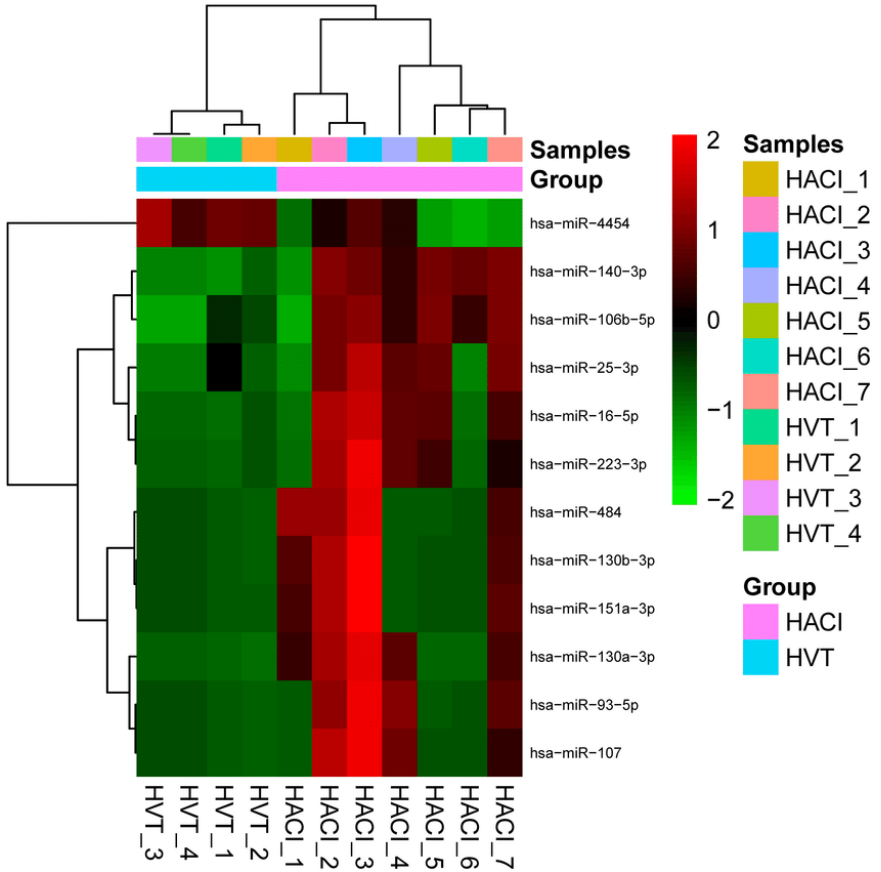
Gene Expression and Regulation:

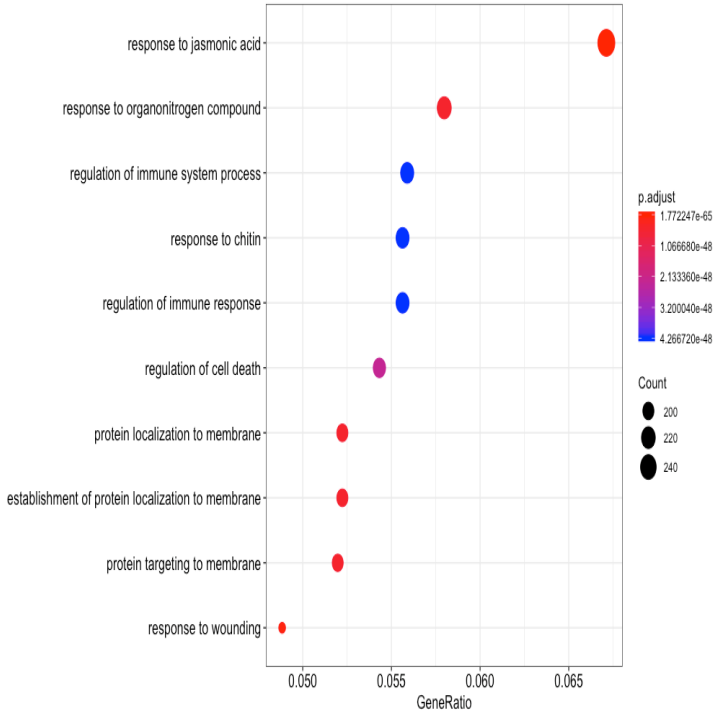
Microarrays:

Technique for measuring expression levels of thousands of genes

Gene Function:

Insights into regulatory mechanisms, functional annotation



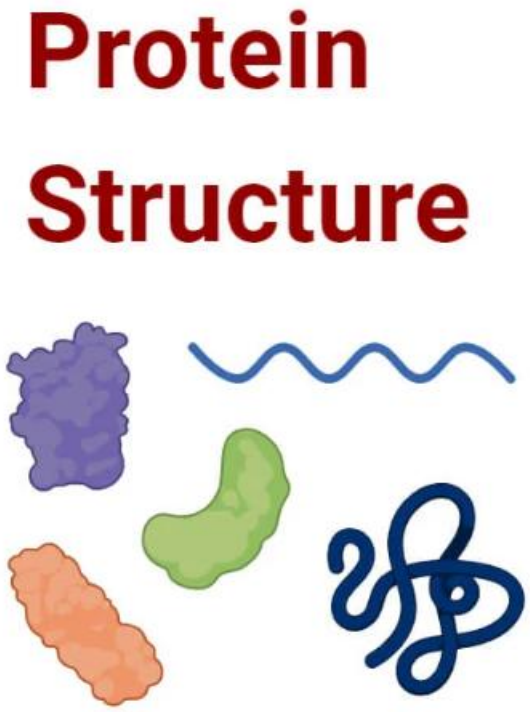


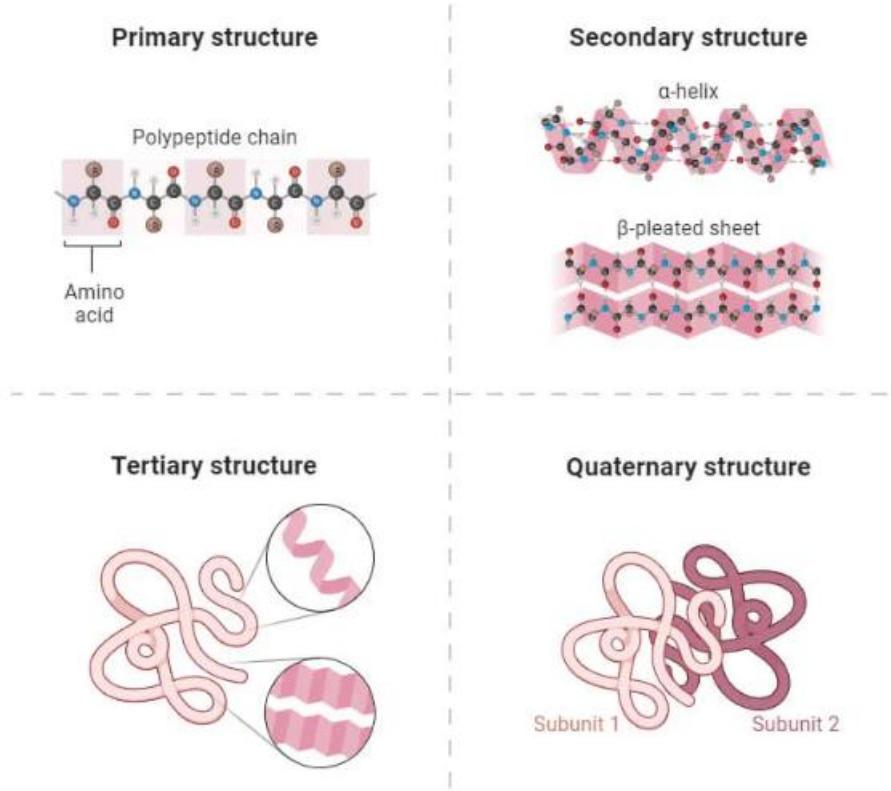
Proteomics: Analysing Proteins and Their Functions

Protein Sequences:
Techniques: Mass spectrometry, protein sequencing
Applications: Identifying proteins, characterizing protein sequences

Protein Folding and Function:
Folding: importance of protein structure (primary, secondary, tertiary)

Function: Enzymatic activity, interaction with other molecules

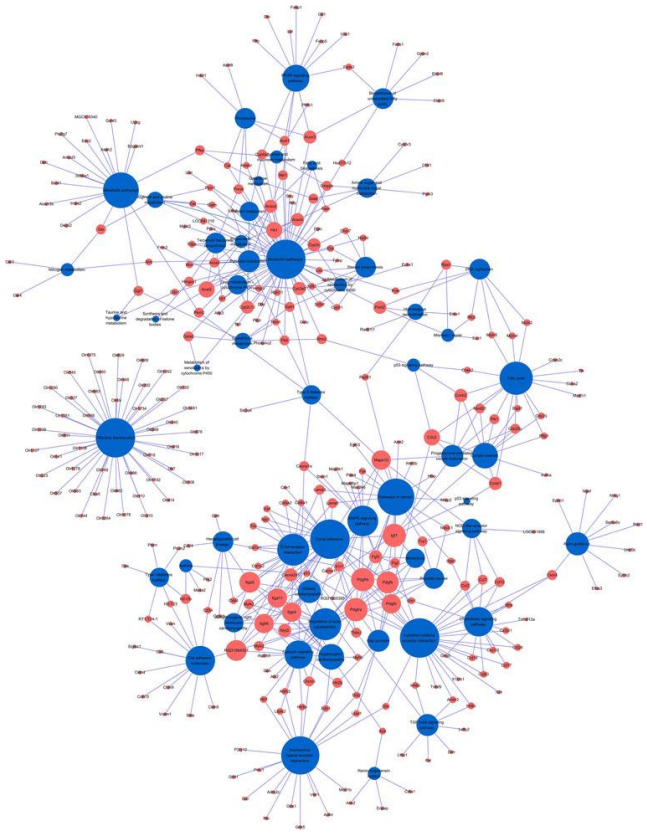


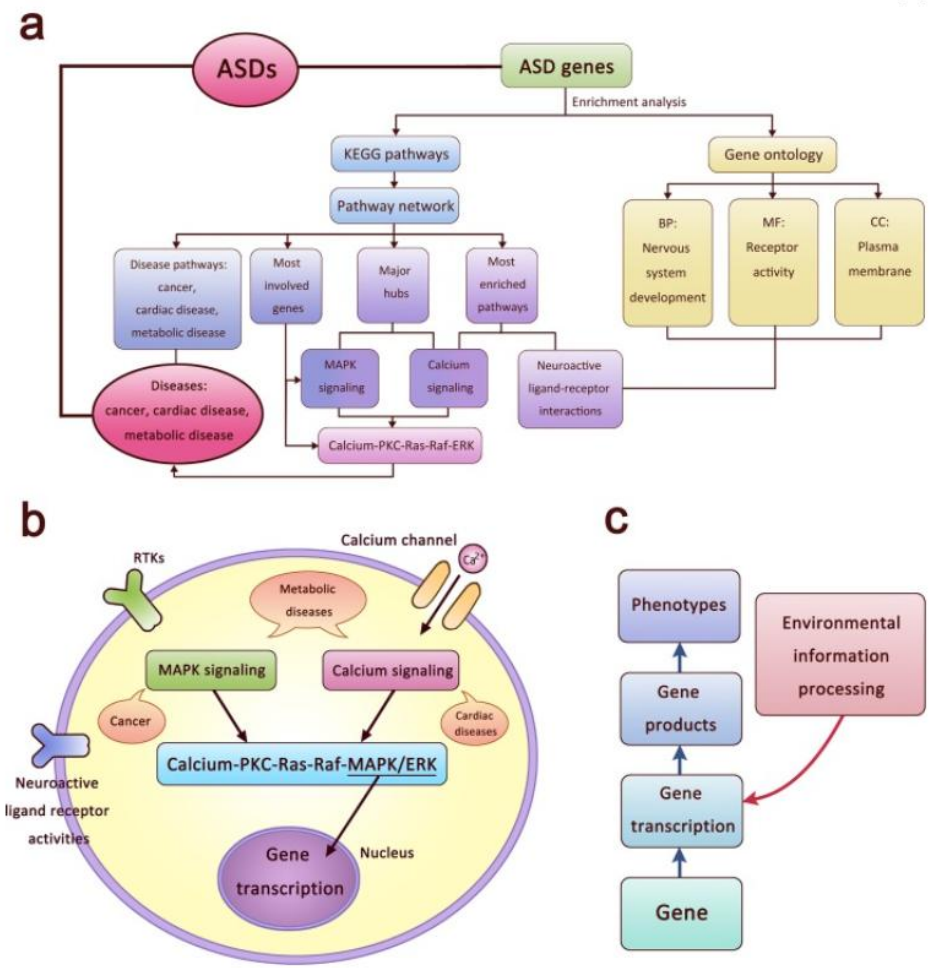


Proteomics: Analysing Proteins and Their Functions

Protein Networks and Molecular Pathways: Networks: Interaction networks, signaling pathways

Pathways: Cellular processes, disease mechanisms





أكاديمية كاوست
KAUST ACADEMY

Lunch Break

KAUST

أكاديمية كاوست
KAUST ACADEMY

Day 1: Introduction to Bioinformatics and Biological Resources

Afternoon Session

KAUST

Outline Day 1

Afternoon (2-5 pm):

Biological Resources:

- Types of biological databases/resources: Sequence, structure, functional.
- Overview of genomic, transcriptomic,, and proteomic databases/resources.

Practical Session:

- Primary Option: Query databases like GenBank, PDB, and KEGG (web-based).
- Alternative: Access and query these databases using Google Golab.

Data and Information

Data
Raw, unorganized facts that need to be processed to gain meaning.
Example: Exam scores of ind1iv1idual students.

Information
When data is processed, organised, structured, or presented in a given context to make it useful.
Example: The average ,exam score of a cl1ass..

Knowledge
Information that is understood, interpreted and applied based on experience or learning. It reflects insights gained from analyzing and understanding information.
Examples: Knowing how to improve student performance based on exam score trends

Biological Data

1. Sequence Data
Description: Linear sequences of nucleotides (DNA, RNA) or amino adds (proteins).
Examples: DNA sequences (A, T, G, C), RNA sequences (A, U, G, C), and protein sequences (chains of amino acids).
Use Cases: Genome annotation, gene expression studies, protein synthesis understanding.
2. Structural Data
Description: Information about the 3D arrangement of biological molecules, particularly proteins.
Examples: Primary, secondary, tertiary, and quaternary structures of proteins, often visualized in 3D models.
Use Cases: Drug design, understanding protein interactions, and enzyme functionality.
3. Functional Data
Description: Data related to the function or activity of genes and proteins.
Examples: Gene ontology, metabolic pathways, enzyme kinetics.
Use Cases: Functional genomics, pathway analysis, metabolic modeling.

Biological Data

4. Expression Data
Description: Information about gene or protein expression levels, often measured through experiments.
Examples: RNA-seq data, microarray data, proteomics data (mass spectrometry).
Use Cases: Identifying differentially expressed genes or proteins in conditions like disease.
5. Interaction Data
Description: Describes interactions between biomolecules (e.g., protein-protein, protein-DNA, or protein-RNA interactions).
Examples: Protein interaction networks, gene regulatory networks, signaling1pathways.
Use Cases: Understanding cellular processes, pathway reconstruction, systems biology.
6. Phenotypic Data
Description: Data that captures observable traits or characteristics of an organism.
Examples: Disease phenotypes, clinical trial data, morphological traits.
Use Cases: Genome-wide association studies (GWAS), personalised medicine, trait mapping.

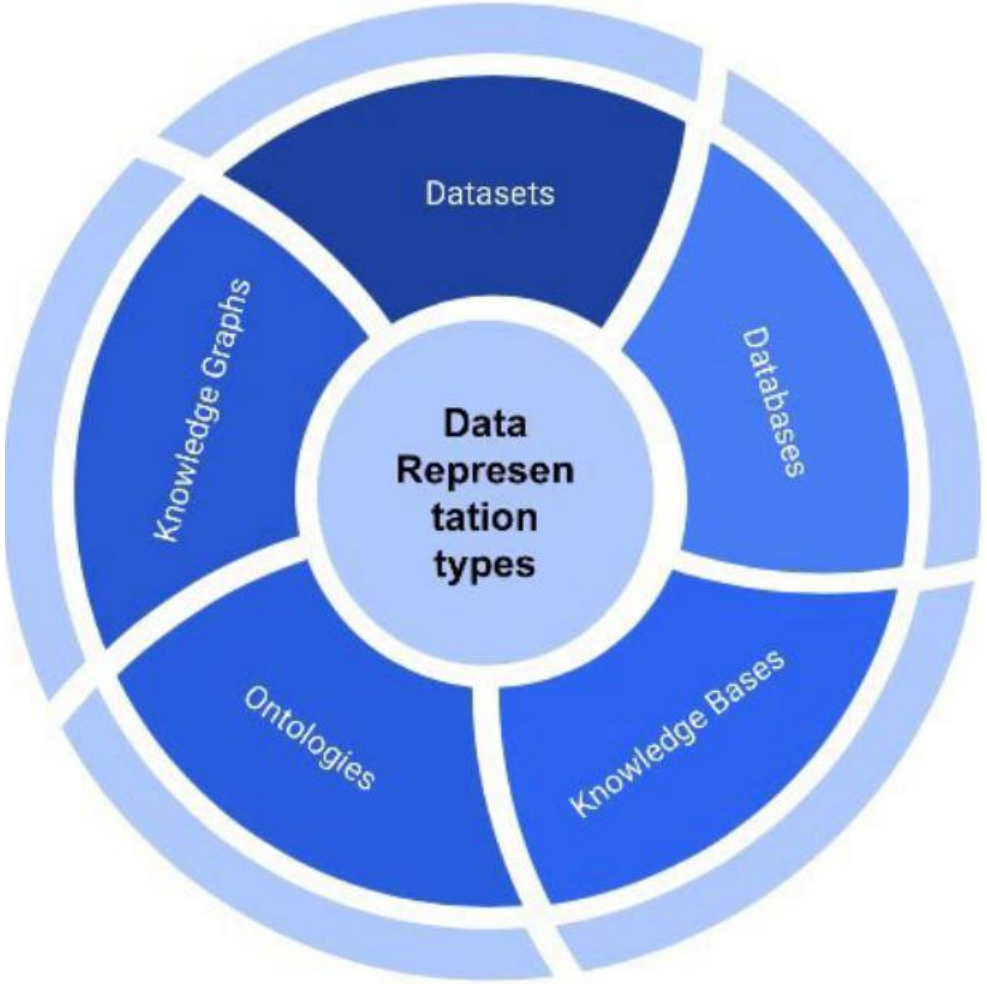
Biological Data

7. Time-Series Data
Description: Data collected at regular intervals over time, often used to study changes in biological systems.
Examples: Heart rate, EEG, gene expression over time.
Use Cases: Longitudinal studies, circadian rhythm research, real-time monitoring of biological processes.
8. Textual Data
Description: Written descriptions and annotations associated with biological data.
Examples: Clinical notes, literature reviews, gene annotations.
Use Cases: Mining biological literature, natural language processing (NLP) applications., clinical research.
9. Image Data
Description: Visual representations of biological structures or experimental results.
Examples: Microscopy images, MRI scans, electrophoresis gels.
Use Cases: Cell imaging, diagnostic imaging, protein visualization

Biological Data Representation Types

Ontologies: Formalized representations of knowledge with structured concepts and their relationships.

Knowledge Graphs: Graph-based representations that map relationships between biological entities and their attributes.

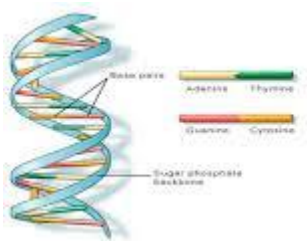


Datasets: Raw collections of biological data

Databases: Structured repositories

Knowledge Bases: Integrated collections of curated information with annotations and additional context

Biological Datasets



```
CTGGGGCTTTACTGATGTCATACCGTCTTGCACGGGGATAGAAT
ATTTTCTGAAAGTTACAGACTTCGATTAAAAAGATCGGACTGCG
TTTTTCGACGTGTCAAGGACTCAAGGGAATAGTTTGGCGGGAGC
CGATAAAATTCAACTACTGGTTTCGGCCTAATAGGTCACGTTTT
CCCTGGGTGTTCTATGATAAGTCCTGCTTTATAACACGGGGCGG
ATCCAAGCGCCCGCTAATTCTGTTCTGTTAATGTTCATACCAAT
AGCCCAGTCGCAAGGGTCTGCTGCTGTTGTCGACGCCTCATGTT
GGTTAAGGCGTGTGATCGACGATGCAGGTATACATCGGCTCGGA
TCGCGGTTCGGCGCGTAGTTGAGTGCGATAACCCAACCGGTGGC
AGACAACCTAACTAATAGTCTCTAACGGGGAATTACCTTTACCA
CAATGATATCGCCCACAGAAAGTAGGGTCTCAGGTATCGCATAC
GACAGTAGAGAGCTATTGTGTAATTCAGGCTCAGCATTCATCGA
```

Description: Raw data collected from experiments; observations, or simulations, usually in a structured format like tables.

Examples:

- Genomic sequences: FASTA, FASTQ files containing DNA or RNA sequences
- Transcriptomic data: RNA-seq data matrices
- Proteomic data: Mass spectrometry-based proteomics datasets.
- Images

Applications:

Data mining

Statistical analysis

Comparative studies

Dataset	Number of Attributes	Number of Classes	Class 1, 2 Samples
ALL-AML Leukemia	7129	2 (ALL vs. AML)	47 25
Ovarian Cancer	15154	2 (Cancer .vs. Normal)	162 91
Lung Cancer	12533	2 (MPM vs. ADCA)	31 150
Colon Cancer	2000	2 (positive vs. negative)	22 40
Prostate Cancer	12600	2 (tumor vs. normal)	77 59
Breast Cancer	24481	2 (relapse vs. non-relapse)	46 51
CNS Tumor	7129	2 (class 1 vs. class 2)	21 39

Biological Datasets

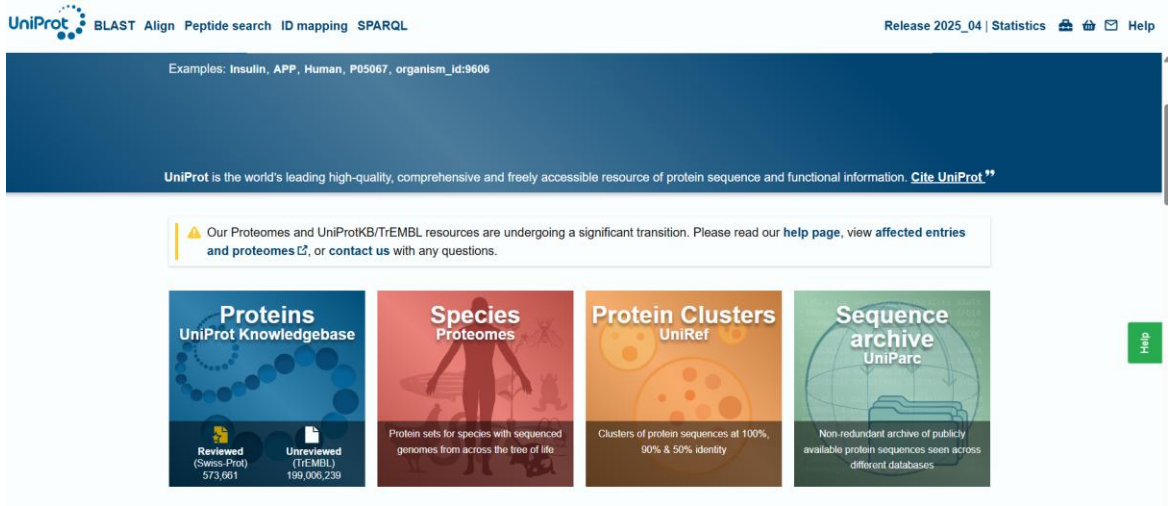
Description: Structured collections of biological data curated for easy retrieval, usually organized and searchable by specific
criteria.

Examples:

- NCBI GenBank: DNA and RNA sequences
- UniProt: Protein sequences and functional information
- PDB (Protein Data Bank): 3D structures of proteins and nucleic acids

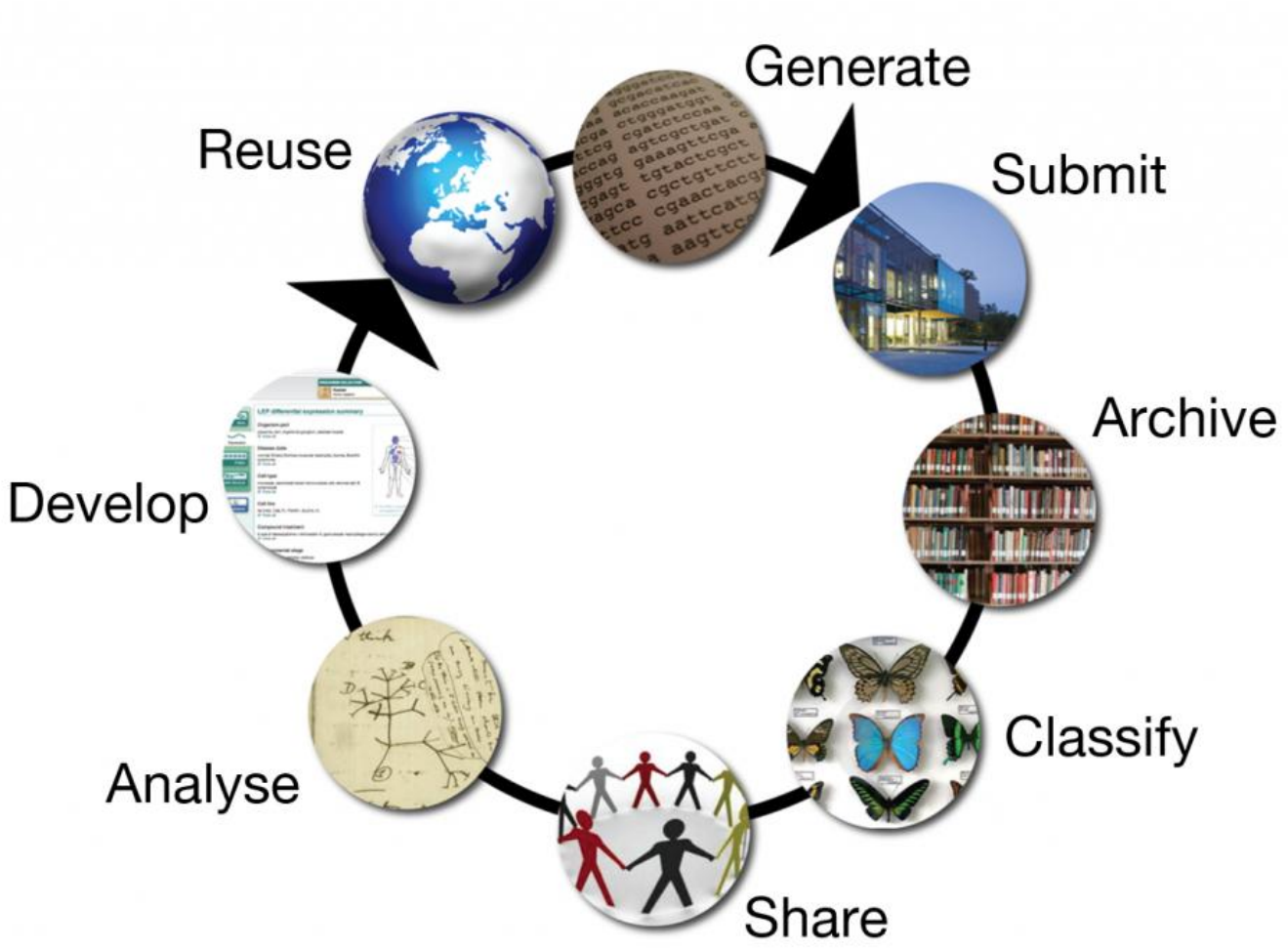
Applications:

Data retrieve
Sequence alignment
Structural prediction



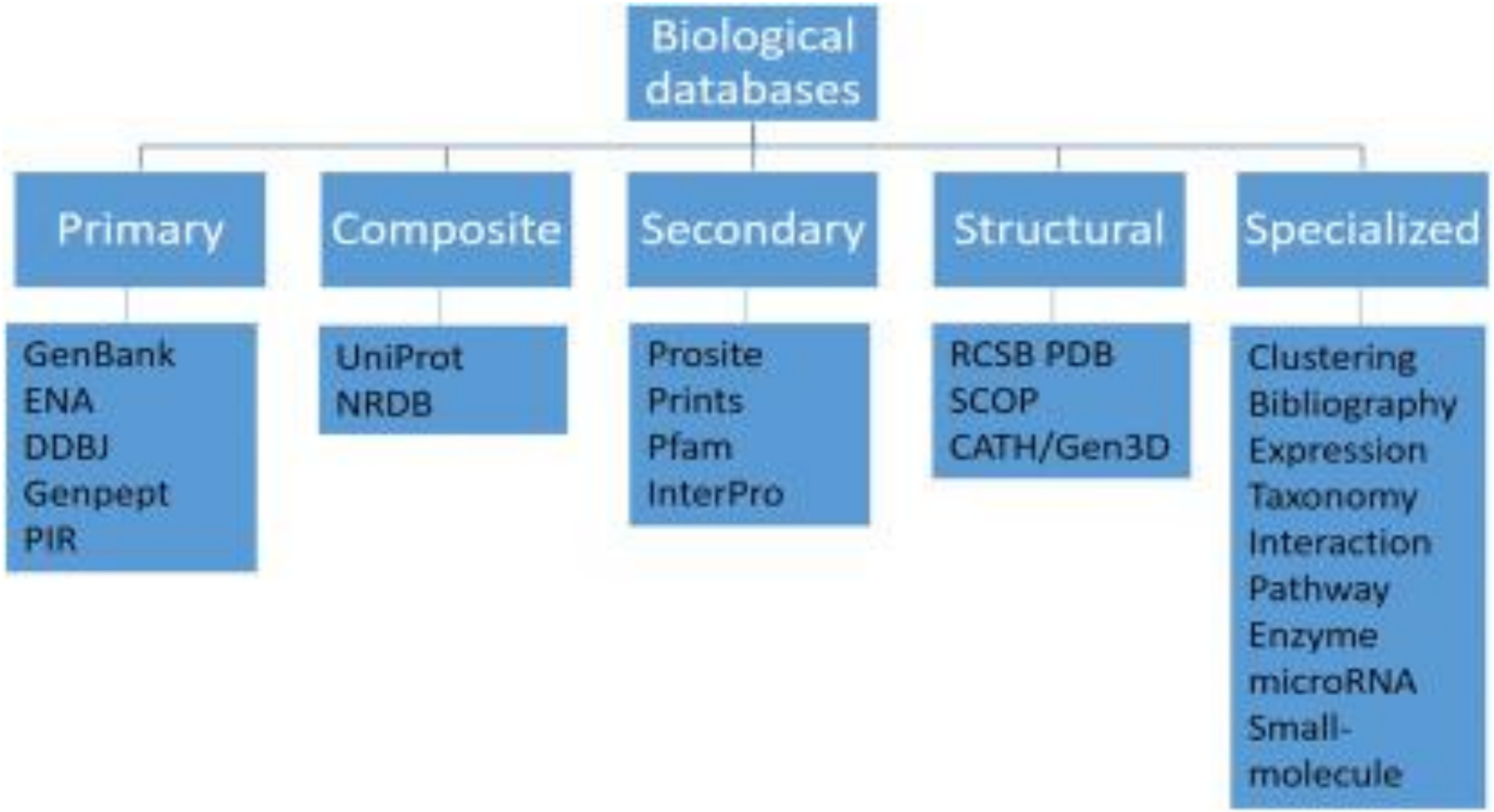
The Role of Public Databases

Bioinformatics centres of excellences:
The EMBL-European Bioinformatics Institute (EMBL EBI)
The US National Centre for Biotechrmlogy Information (NCBI)
The National Institute of Genetics in Japan (NIG).



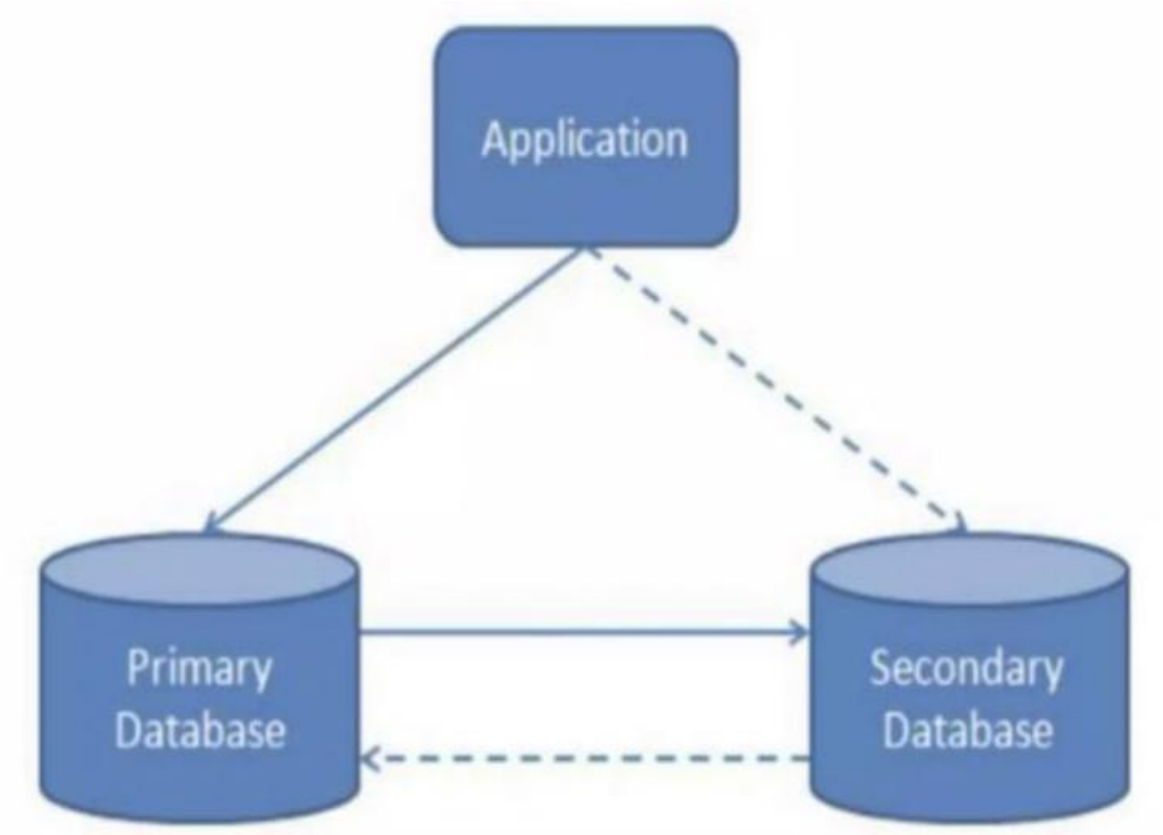
Types of Biological Databases

On the basis of nature of data



Types of Biological Databases

Based on the manner of storage



Primary Databases

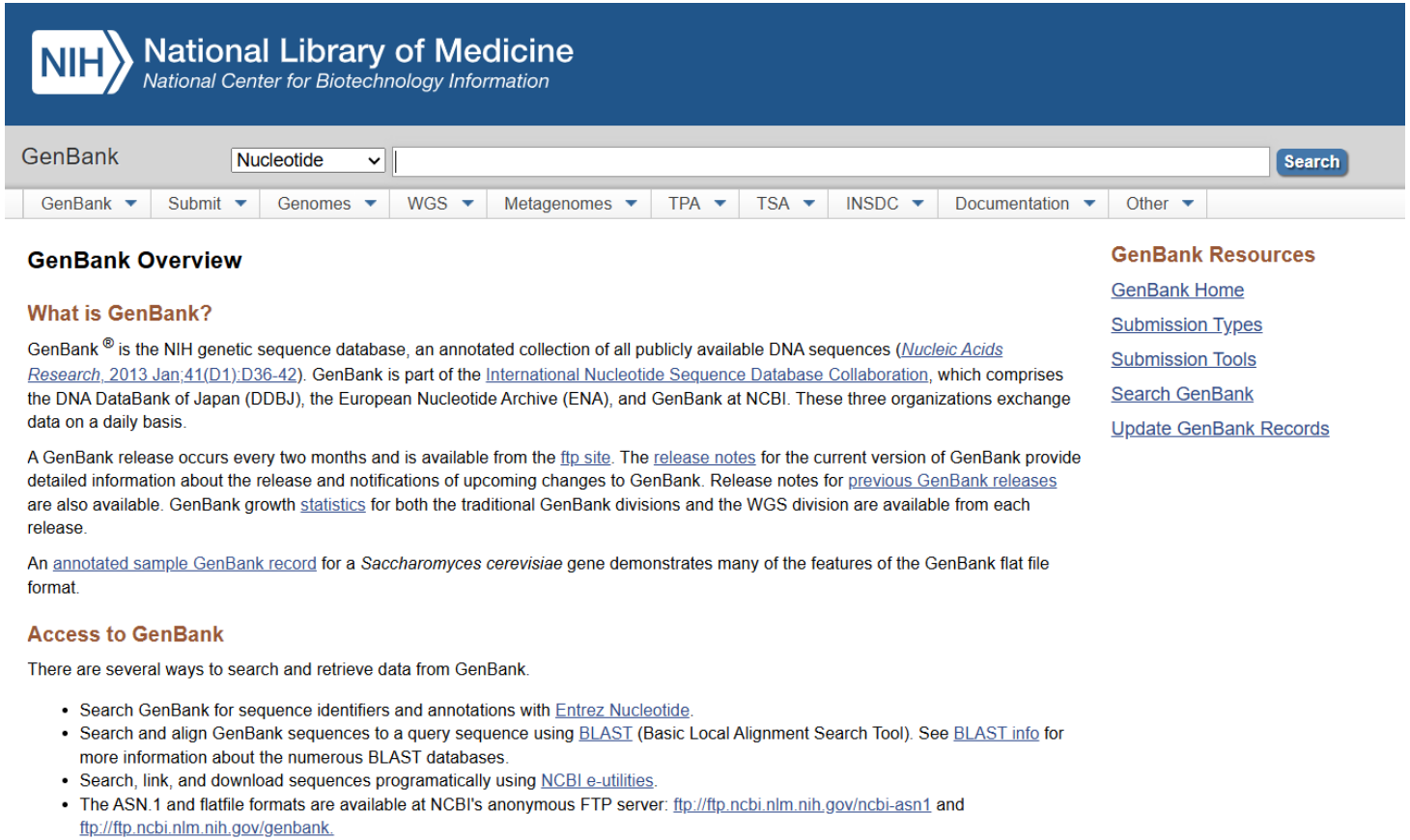
- Biomolecular data are stored in the original form
- Experimental results are submitted directly by researchers into these databases, serving as archival repositories.
- Contents controlled by the submitter.
- Once an accession number is assigned to the data, it cannot be changed or modified.

Examples:

- GenBank
- EMBL-EBI (ENA (European Nucleotide Archive), PRIDE, Array Express, Bioimage Archive)
- DDBJ
- PDB

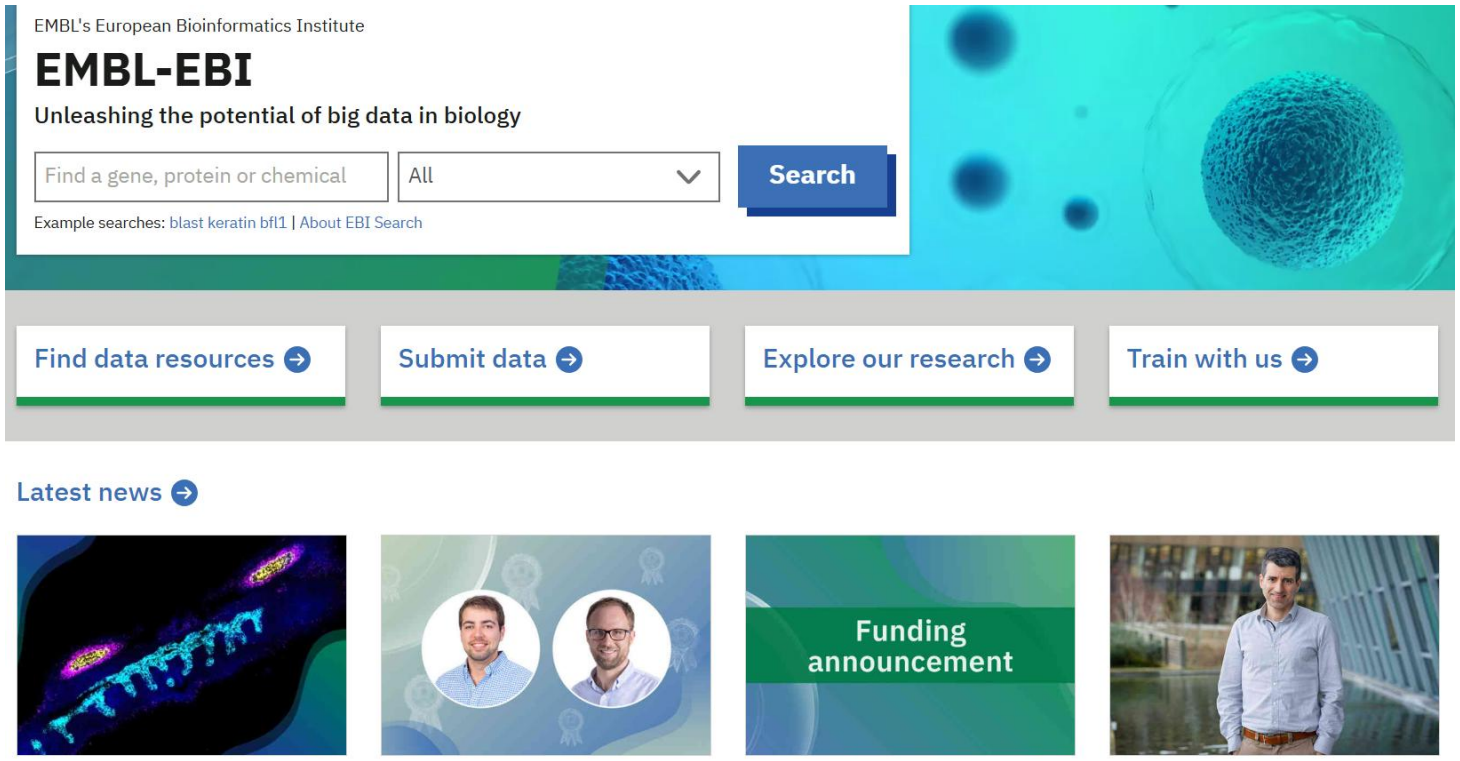
GeneBank

- A comprehensive genetic sequence, database hosted by NCBI.
- Provides publicly available DNA sequences from various species, contributing to global research.
- Part of the international Nucleotide Sequence Database Collaboration (INSDC) alongside DDBJ and EMBL-EBI.



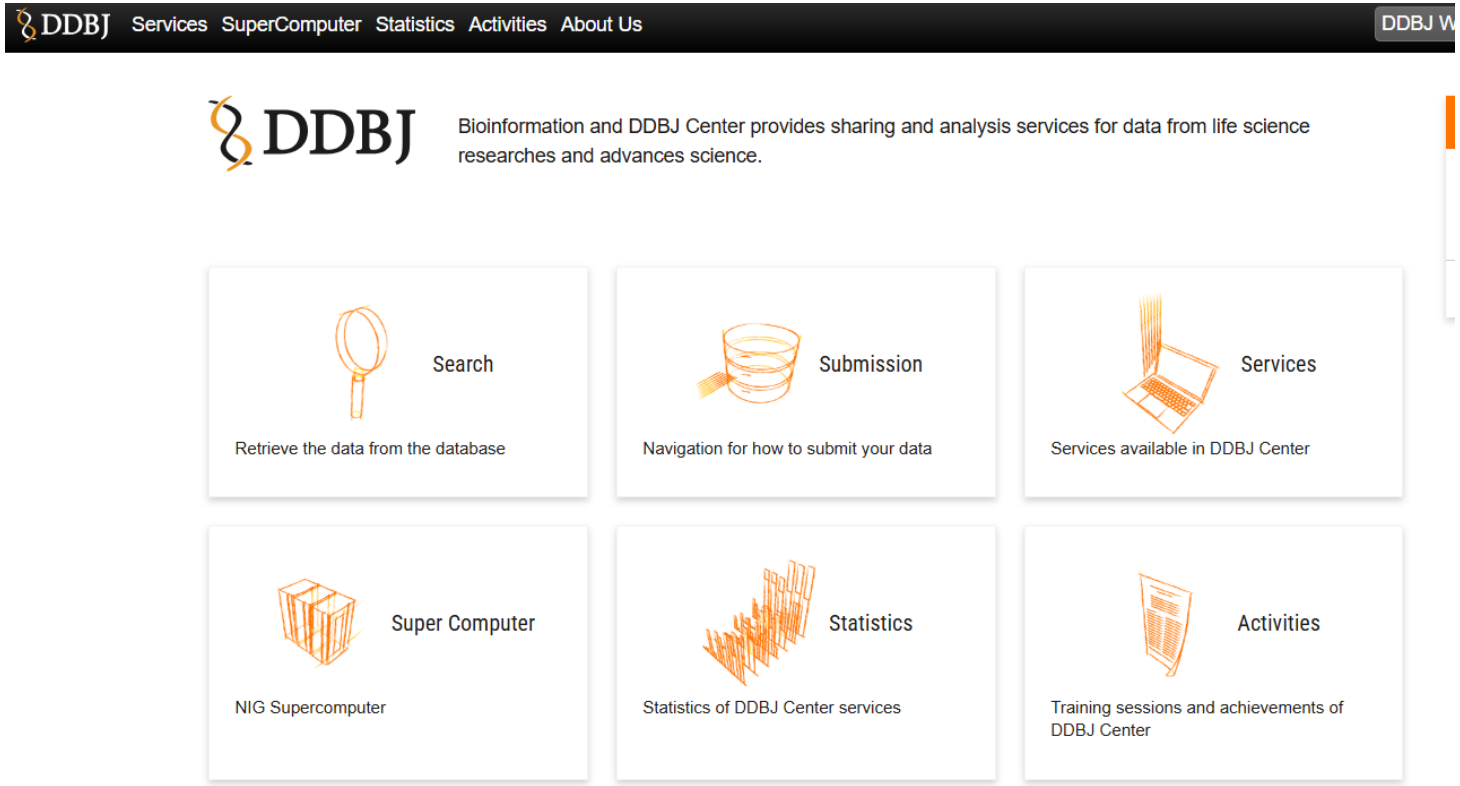
European Molecular Biology Laboratory

- European Bioinformatics Institute (EBI) is part of the EMBL and provides bioinformatics services and research
- Hosts a variety of biological databases
  - **ENA (European Nucleotide Archive):** Raw DNA and RNA sequences,
  - **PRIDE**: Raw proteomics data (mass spectrometry)
  - **ArrayExpress**: Raw transcriptomics data
  - (microarray or RNA-seq datasets)
  - **BioImage Archive**: Raw biological images from various experimental techniques.



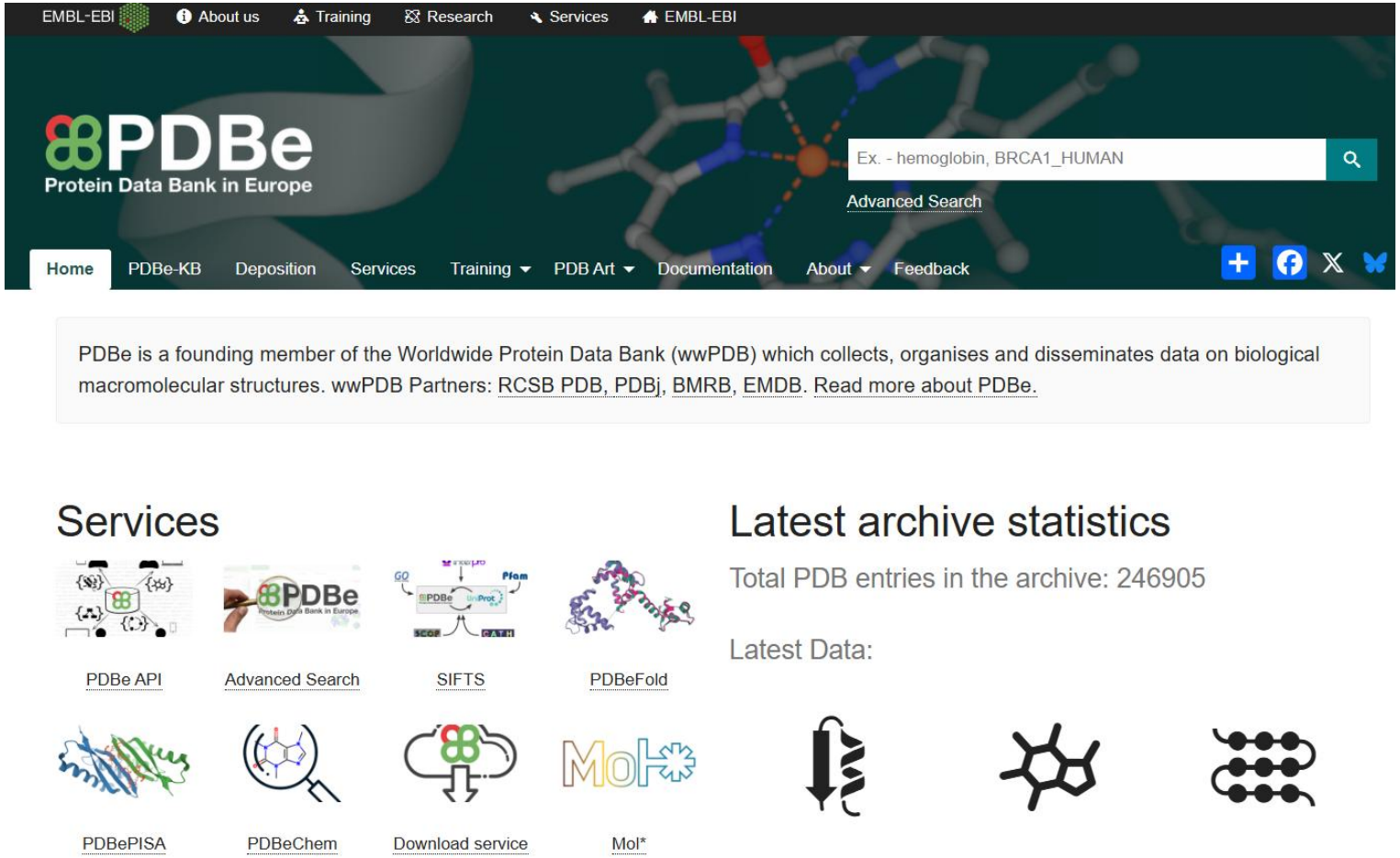
DDBJ

- The DNA Data Bank of Japan (DDBJ) is a biological database that collects DNA sequences.
- DDBJ is a member of the International Nucleotide Sequenoe Database Collaboration (INSDC).



Protein Data Bank (PDB)

- Contains 3D structures of proteins, nucleic acids, and complexes from experimental data.
- **Global Collaboration:**
  Managed by RCSB PDB (USA), PDBe (EMBL-EBL UK), and PDBj (Japan).

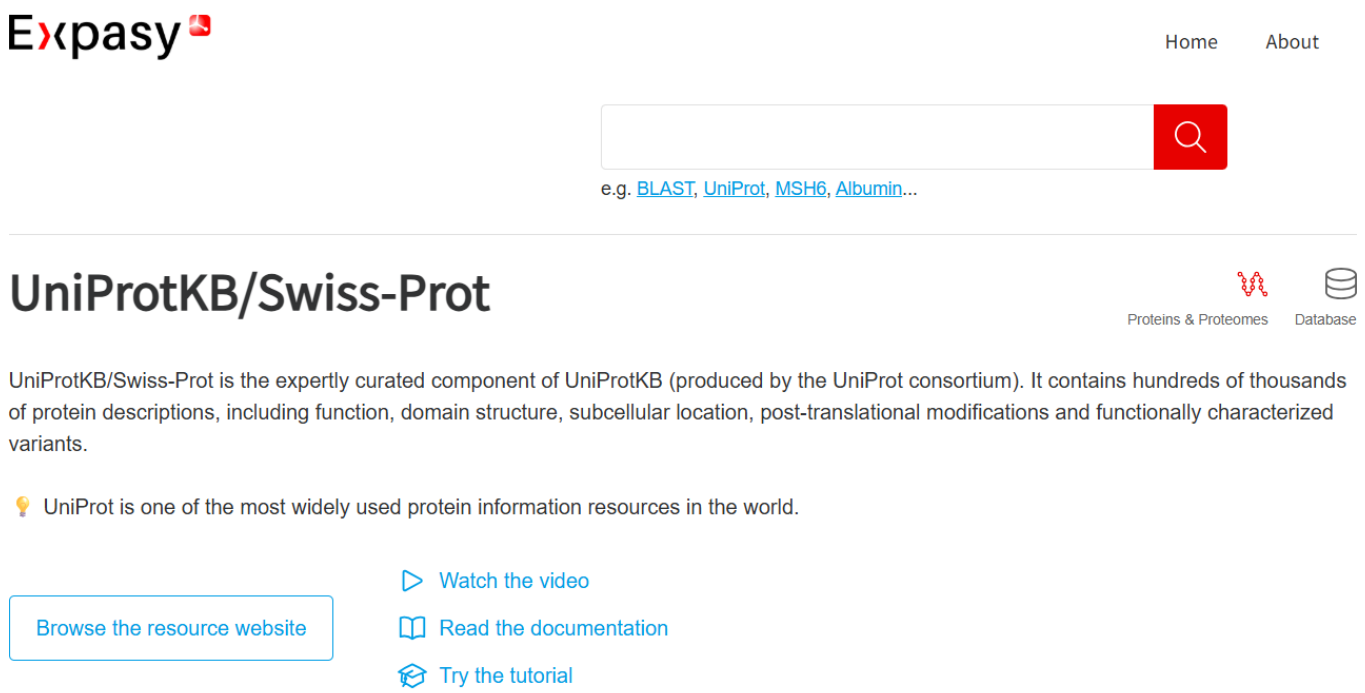


Secondary Databases

- Derives data from analysed primary data
- Contains manually created or automatically generated data
- Contents of secondary db controlled by a third party
  - EMBL's secondary databases :
  - UniProt: A protein sequence and function database that curates and annotates protein sequences based on raw data from primary sources.
  - InterPro: Classifies proteins into families and predicts domains based on protein sequences.
  - Pfam: Provides protein family annotations using hidden Markov models (HMMs) derived from primary protein sequence data.
  - Reactome: A curated pathway database detailing biological processes.
- SIB
  Swiss-Prat
  PROSITE
- NCBI
  PubMed

Swiss-Prot

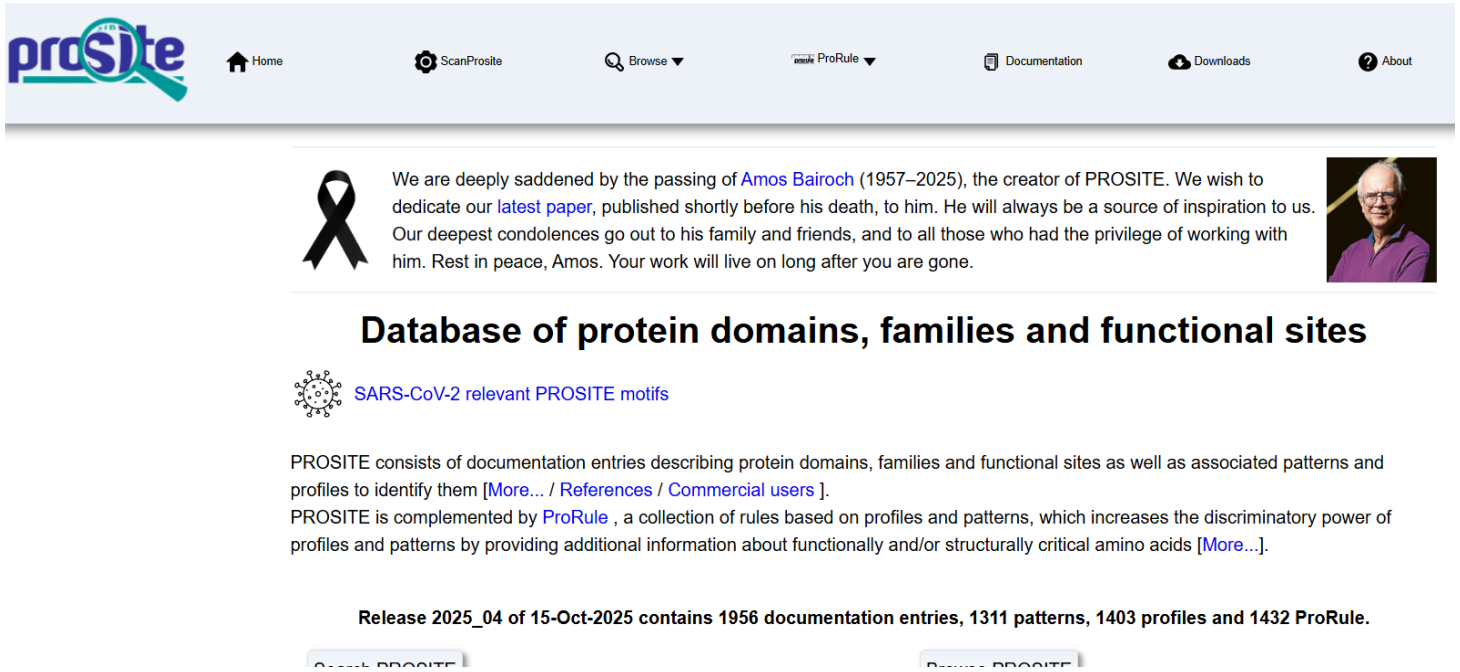
- SWISS-PROT is a curated protein sequence database.
- Aims to provide a high level of annotation, including:
  - Protein function descriptions
  - Domain structures
  - Post-translational modifications
  - Variants
- Swiss -Prat was created in 1986 by Amos Bairoch in collaboration with the Swiss Institute of Bioinformatics.



PROSITE

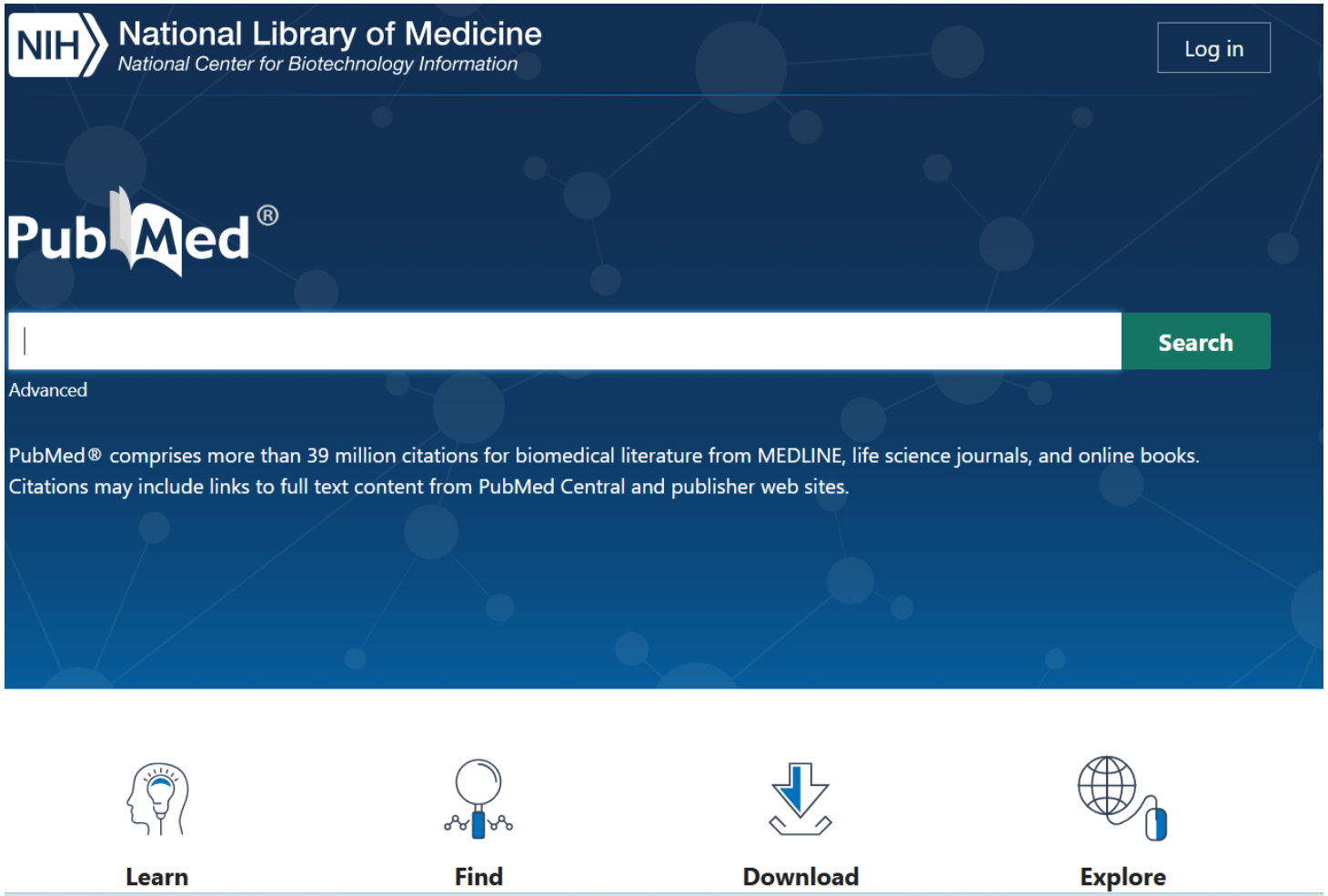
- Protein database

- Integrated with Swiss-Prot protein annotation
- Contains entries describing:
  - Protein families
  - Domains
  - Functional sites
  - Amino acid patterns and profiles
  - Integrated with Swiss-Prot protein annotation



PubMed

- Complementary Resource: Enhances primary literature searches with MEDLINE's extensive biomedical references.
- Supplementary Access: Links to full text articles and resources beyond primary databases.
- Integrated Citations: Provides citation tracking and related articles, complementing primary data sources.
- Rich Metadata: Offers detailed abstracts and indexing to support in-depth research.



Biological Knowledge Bases

Description: Enhanced databases that include relationships and context beyond simple data storage, often integrating various types of biological data.

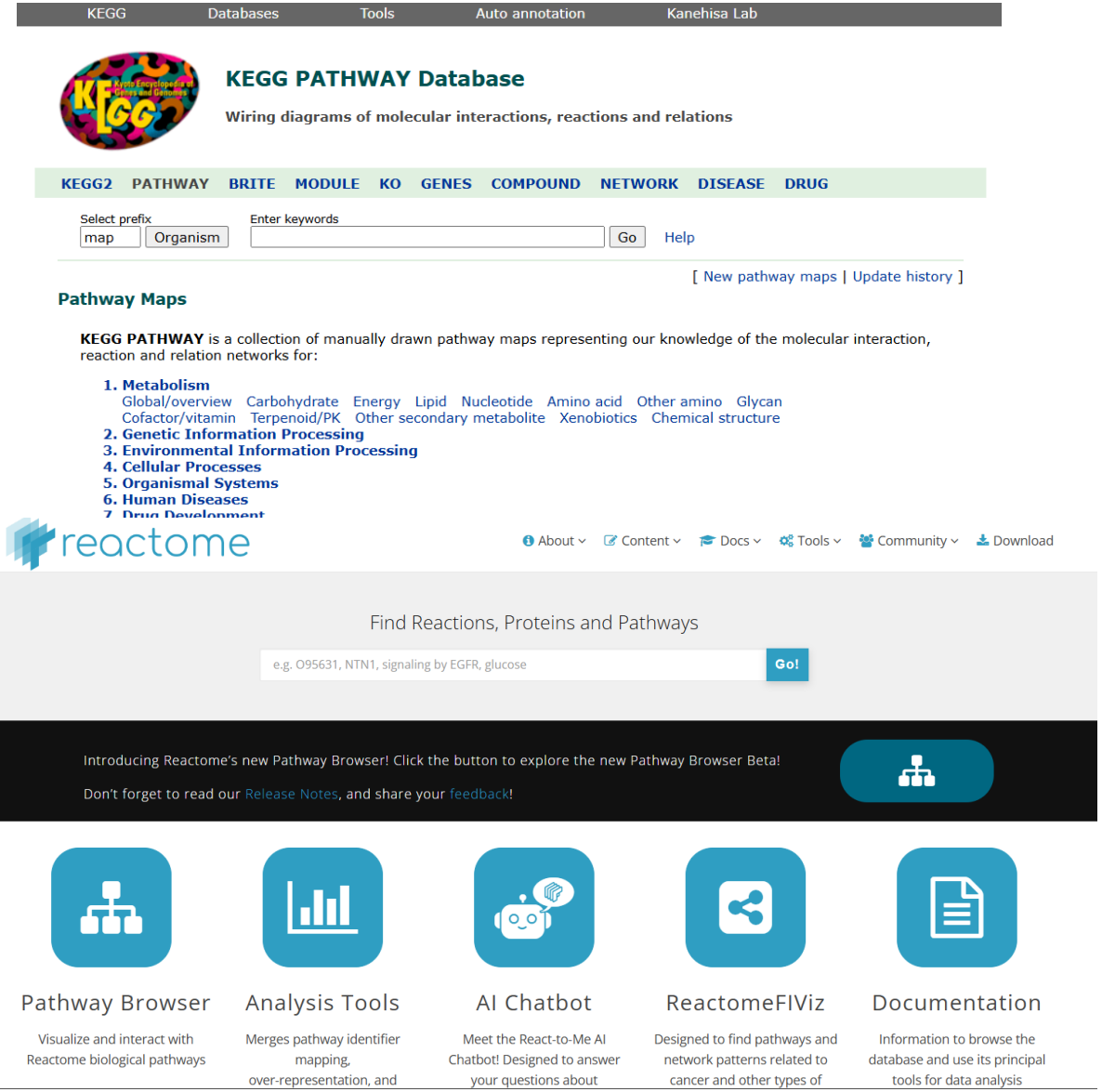
Examples:

KEGG: Pathway and network information

Reactome: Integrates information about molecular interactions, pathways, and reactions

Applications:

- Pathway analysis
- Functional annotation
- Biological interpretation



Biomedical Ontologies

Description: Structured vocabularies that describe categories and relationships in a specific domain

Examples:

- Gene Ontology (GO): Terms for gene product attributes
- Disease Ontology: Classification of diseases

Applications:

- Annotation
- Data integration
- Semantic search
- Natural Language Processing

Ontology Languages

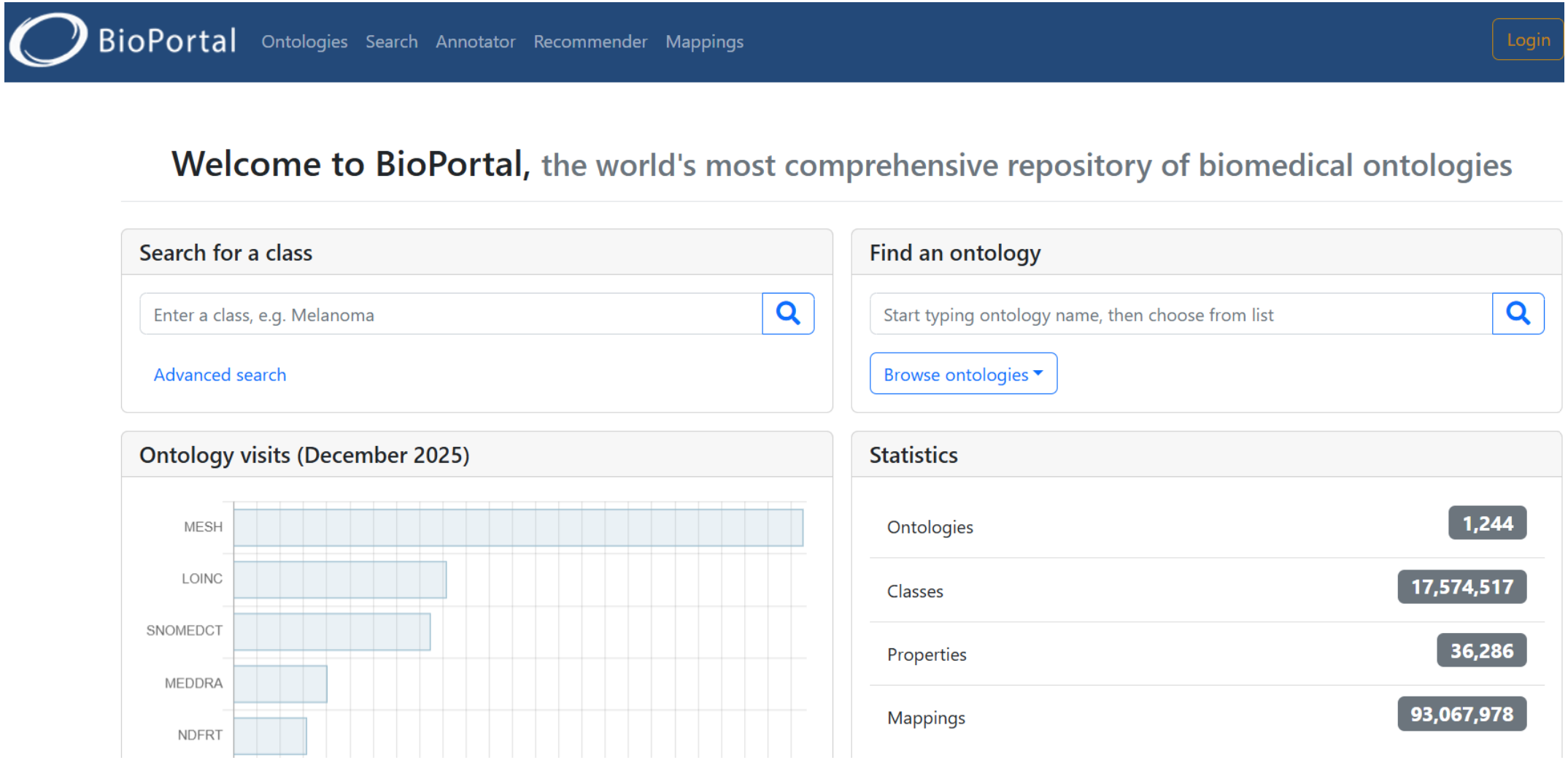
1. OWL (Web Ontology Language)
Standard for creating complex, web-based ontologies
Highly expressive, supports logical rules and reasoning.
Based on Description Logic.
Formats: RDF/XML, OWL/XML, Turtle.

2. OBO (Open Biomedical Ontologies Format)
Simpler format for biomedical ontologies.
Lightweight, easier to read/write.
Focuses on terminologies and basic relationships.
Format: Plain-text.


```
[Term]
id: GO:0002927
name: archaeosine-tRNA biosynthetic process
namespace: biological_process
def: "The chemical reactions and pathways involved in the biosyn
[GOC:hjd, PMID:20129918]
comment: Archaeosine (7-formamidino-7-deazaguanosine) is a struc
guanine transglycosylase (ArcTGT) which catalyzes the exchange o
is_a: GO:0006400 ! tRNA modification
is_a: GO:0009058 ! biosynthetic process
```

KAUST

NCBO - The National Centre for Biomedical Ontology



KAUST

NCBO

أكاديمية كاوست
KAUST ACADEMY

BioPortal Ontologies Search Annotator Recommender Mappings Login Support

Human Disease Ontology

Last uploaded: December 24, 2025

Summary Classes Properties Notes Mappings Widgets

Details

Acronym	DOID
Visibility	Public
Description	Creating a comprehensive classification of human diseases organized by etiology., The Disease Ontology has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts.
Status	Production
Format	OWL
Categories	Human, Neurological Disorder, Neurologic Disease, Health, Biomedical Resources
Groups	OBO Foundry
Bibliographic reference	https://disease-ontology.org/community/publications
Contact	Lynn Schriml (lynn.schriml@gmail.com)

Metrics

Classes	19,318
Individuals	0
Properties	47
Maximum depth	13
Maximum number of children	1,703
Average number of children	4
Classes with a single child	2,122
Classes with more than 25 children	91
Classes with no definition	8,252

Visits

KAUST

ZOOMA

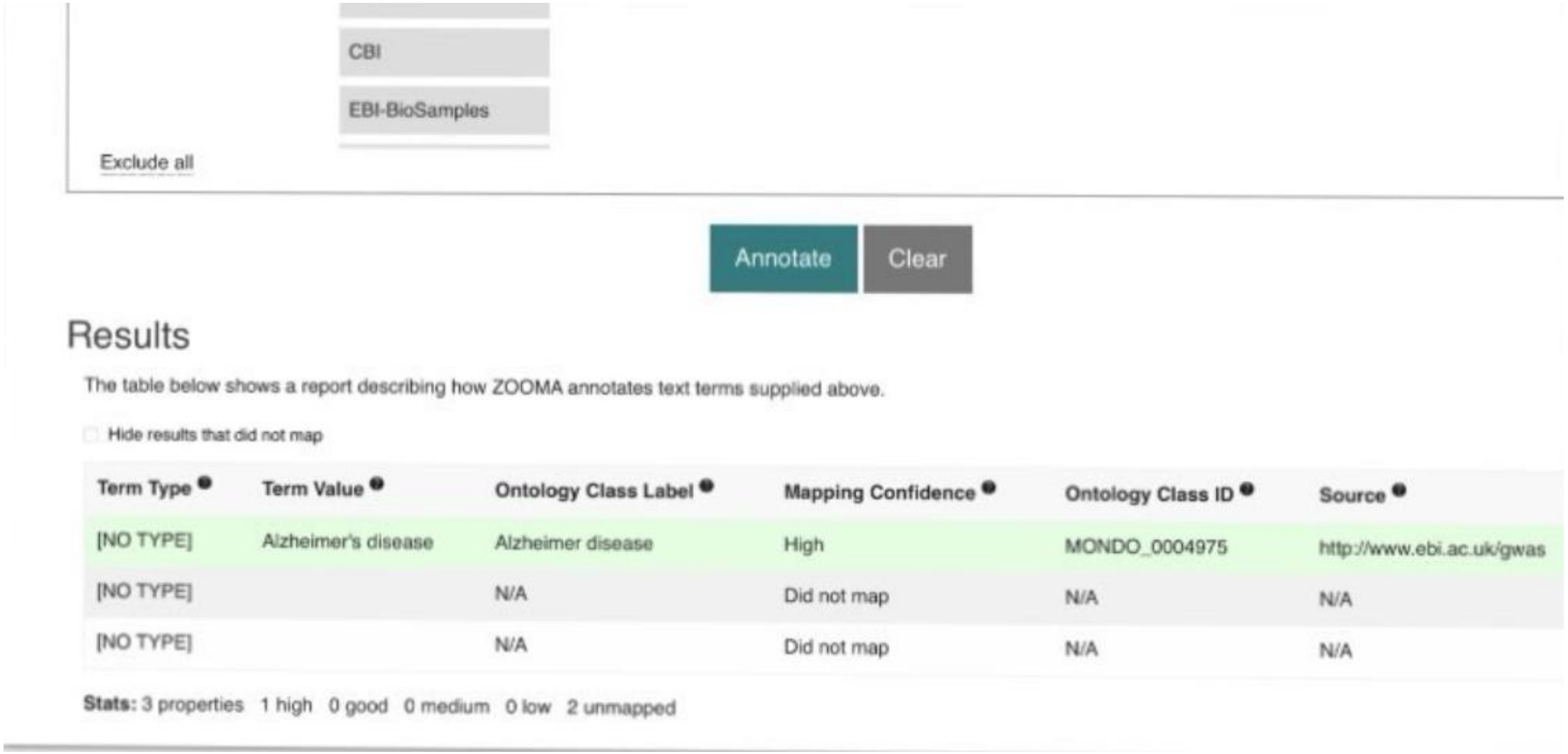
أكاديمية كاوست
KAUST ACADEMY

Query

Use the text box to find possible ontology mappings for free text terms in the ZOOMA repository of curated annotation knowledge. You can add one term (e.g. 'Homo sapiens') per line. If you also have a type for your term (e.g. 'organism'), put this after the term, separated by a tab. If you are new to ZOOMA, take a look at our getting started guide.

Show me some examples...

ZOOMA


CBI
EBI-BioSamples
Exclude all
Annotate
Clear
Results
The table below shows a report describing how ZOOMA annotates text terms supplied above.
Hide results that did not map
Term Type
Term Value
Ontology Class Label
Mapping Confidence
Ontology Class ID
Source
[NO TYPE]
Alzheimer's disease
Alzheimer disease
High
MONDO_0004975
http://www.ebi.ac.uk/gwas
[NO TYPE]
N/A
Did not map
N/A
N/A
[NO TYPE]
N/A
Did not map
N/A
N/A
Stats: 3 properties 1 high 0 good 0 medium 0 low 2 unmapped

Biological Knowledge Graphs

Definition:

Graph-based representation of biological entities and their relationships

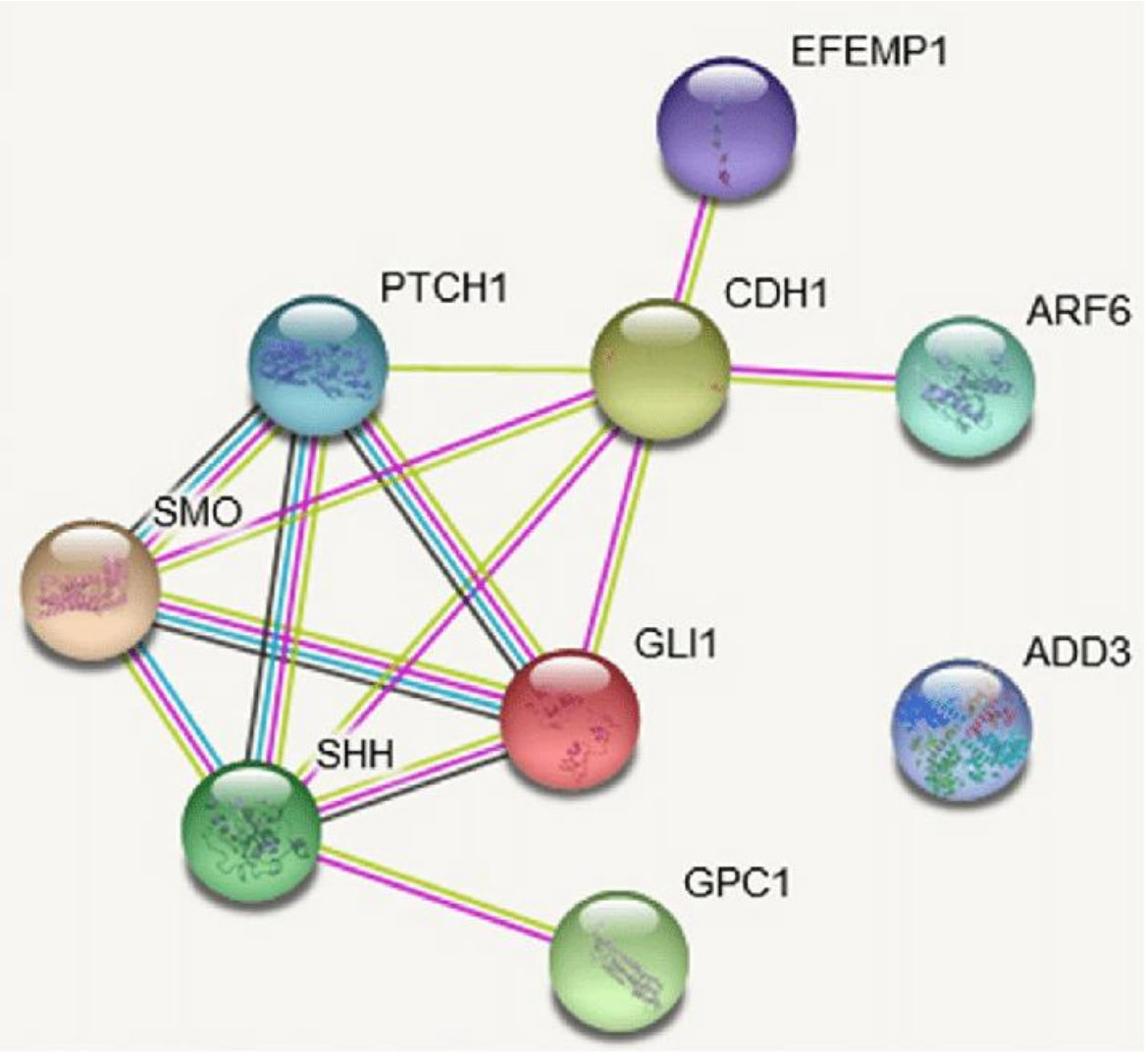
Purpose:

Integrate diverse biological data into a unified model for enhanced understanding

Components:

Nodes: Represent biological entities such as genes, proteins, pathways

Edges: Represent relationships or interactions between entities



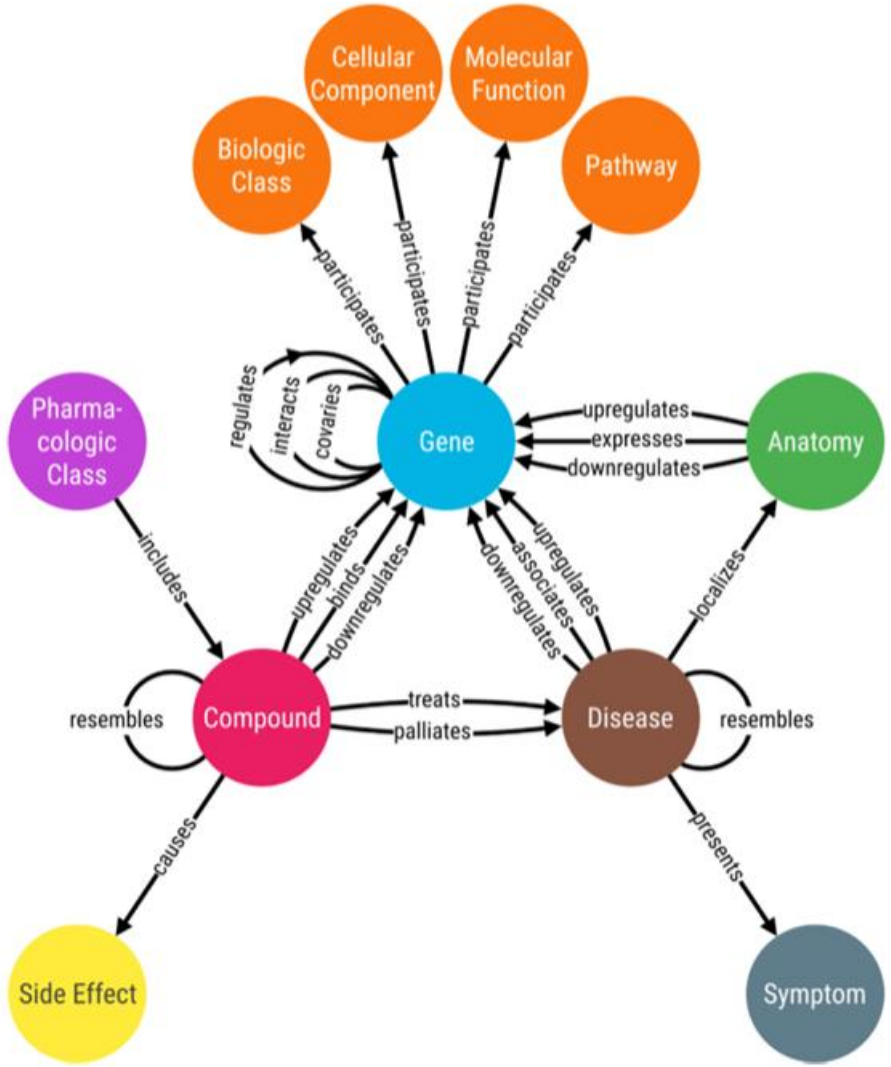
STRING PPI Network

Biological Knowledge Graphs

Applications:
Disease Research: identifies connections between genes, proteins and diseases.

Drug Discovery: Facilitates target identification and drug repurposing by linking drug targets with biological pathways.

Personalized Medicine: Helps in understanding patient-specific biological profile and predicting treatment outcomes.



Building Knowledge Graphs

Data Collection

Graph Construction

Data Processing and Integration

Storage

Practical: Access Biomedical Resources