**INTRODUCTION TO BIOINFORMATICS - GLOSSARY**

This is an introductory course that covers the topics of big data bioinformatics and its uses in **basic research, healthcare**, and the **biotech** and **pharmaceutical industries**. In many ways, advances in this discipline have helped make major discoveries in recent years leading to new opportunities. That is why many biologists are starting to see data skills as essential to their studies and research. The topics are organized into **3 major segments**: **Introduction to Big Data: Bioinformatics, Bioinformatics in Healthcare, Translational Bioinformatics.** This course is designed to introduce undergraduate and graduate-level students in biology or related fields to the field of bioinformatics, or the intersection of informatics and biology, and the opportunities that come with the available big data for research and industry.

In this blog, we have listed all the terminologies and their meanings for your understanding, from the [**Introduction to Bioinformatics**](https://learn.omicslogic.com/courses/course/course-1-introduction-to-bioinformatics)course. Clicking on the lesson heading will redirect you back to the lesson where these terms are used.

[**L1- Introduction to Big Data Bioinformatics**](https://learn.omicslogic.com/Learn/course-1-introduction-to-bioinformatics/lesson/l1-introduction-to-big-data-bioinformatics)

In this lesson, you will get a survey, a broader overview of Bioinformatics, Big data, their applications and data driven approach.

**Genomics:**Genomics is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes. A genome is an organism's complete set of DNA, including all of its genes as well as its hierarchical, three-dimensional structural configuration.

**Personalized Medicine:** Personalized medicine is an emerging practice of medicine that uses an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease.

**T**[**ranscriptome**](https://edu.t-bio.info/glossary/transcriptome/)**sequencing:**The set of genes which are transcribed in any one condition is known as the transcriptome, and the process of determining the genetic codes contained in the transcriptome, and their relative proportions, is known as transcriptome sequencing.

**Proteomics:**Proteomics is the large-scale study of proteins. Proteins are vital parts of living organisms, with many functions. The proteome is the entire set of proteins produced or modified by an organism or system. Proteomics enables the identification of ever-increasing numbers of proteins.

**Metabolomics:**Metabolomics is the large-scale study of small molecules, commonly known as metabolites, within cells, biofluids, tissues or organisms. Collectively, these small molecules and their interactions within a biological system are known as the metabolome. Thus metabolomics best represents the molecular phenotype.

**Metagenomics:**Metagenomics enables the study of all microorganisms, regardless of whether they can be cultured or not, through the analysis of genomic data obtained directly from an environmental sample, providing knowledge of the species present, and allowing the extraction of information regarding the functionality of microbes.

**Big data:**Big data is a term that describes large, hard-to-manage volumes of data – both structured and unstructured – that inundate businesses on a day-to-day basis. ... Big data can be analyzed for insights that improve decisions and give confidence for making strategic business moves.

**Cloud computing:**Cloud computing is the delivery of different services through the Internet. These resources include tools and applications like data storage, servers, databases, networking, and software. As long as an electronic device has access to the web, it has access to the data and the software programs to run it.

**Machine learning:**Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

**Multi-omics:**Multi Omics, multi-omics, integrative omics, "panomics" or 'pan-omics' is a biological analysis approach in which the data sets are multiple "ones", such as the genome, proteome, transcriptome, epigenome, metabolome, and microbiome (i.e., a meta-genome and/or meta-transcriptome, depending upon how it is sequenced); in other words, *the use of multiple omics technologies to study life in a concerted way*. By combining these "ones", scientists can analyze complex biological big data to find novel associations between biological entities, pinpoint relevant biomarkers and build elaborate markers of disease and physiology. In doing so, multi omics integrate diverse omics data to find a coherently matching geno-pheno-enviro type relationship or association. The OmicTools service lists more than 99 softwares related to multi omic data analysis, as well as more than 99 databases on the topic.

**Supervised:**Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.

**Unsupervised:**Unsupervised learning refers to the use of artificial intelligence (AI) algorithms to identify patterns in data sets containing data points that are neither classified nor labeled. In other words, unsupervised learning allows the system to identify patterns within data sets on its own.

**Mutation:** A mutation is a change in a DNA sequence. Mutations can result from DNA copying mistakes made during cell division, exposure to ionizing radiation, exposure to chemicals called mutagens, or infection by viruses.

**Biomarkers:**A biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease. A biomarker may be used to see how well the body responds to a treatment for a disease or condition.

[**The Cancer Genome Atlas (TCGA)**](https://cancergenome.nih.gov/)**:** The Cancer Genome Atlas (TCGA) is a public funded project that aims to catalogue and discover major cancer-causing genomic alterations to create a comprehensive “atlas” of cancer genomic profiles.

**Personalized medicine:**Personalized medicine, also referred to as precision medicine, is a medical model that separates people into different groups—with medical decisions, practices, interventions and/or products being tailored to the individual patient based on their predicted response or risk of disease.

**Bioinformatics:** Bioinformatics is defined as the application of tools of computation and analysis to the capture and interpretation of biological data. It is an interdisciplinary field, which harnesses computer science, mathematics, physics, and biology.

**Exome:** The exome is the part of the genome composed of exons, the sequences which, when transcribed, remain within the mature RNA after introns are removed by RNA splicing and contribute to the final protein product encoded by that gene.

[**L2: Bioinformatics in Healthcare**](https://learn.omicslogic.com/Learn/course-1-introduction-to-bioinformatics/lesson/l2-bioinformatics-in-healthcare)

In this lesson, you will get an overview of examples from the healthcare area, where bioinformatics can be applied.

**Molecular diagnostics:**Molecular diagnostics is a collection of techniques used to analyze biological markers in the genome and proteome, and how their cells express their genes as proteins, applying molecular biology to medical testing. They overlap with clinical chemistry (medical tests on bodily fluids).

**S**[**ingle nucleotide polymorphisms**](https://edu.t-bio.info/glossary/single-nucleotide-polymorphism/)**:**A single nucleotide polymorphism, or SNP (pronounced "snip"), is a variation at a single position in a DNA sequence among individuals. If a SNP occurs within a gene, then the gene is described as having more than one allele. In these cases, SNPs may lead to variations in the amino acid sequence.

[**Allele**](https://edu.t-bio.info/glossary/allele/)**:** An allele is a variant form of a gene. Some genes have a variety of different forms, which are located at the same position, or genetic locus, on a chromosome. Alleles contribute to the organism's phenotype, which is the outward appearance of the organism. Some alleles are dominant or recessive.

[**L3: Translational Bioinformatics**](https://learn.omicslogic.com/Learn/course-1-introduction-to-bioinformatics/lesson/l3-translational-bioinformatics)

In this lesson, you will learn about other fields, where bioinformatics can be applied like Environment, Agriculture, etc. Besides, you will get an overview on the analysis platform "[T-Bioinfo Server](https://server.t-bio.info/)".

[**Sequencing**](https://edu.t-bio.info/glossary/next-generation-sequencing/)**:**DNA sequencing is a laboratory technique used to determine the exact sequence of bases (A, C, G, and T) in a DNA molecule. The DNA base sequence carries the information a cell needs to assemble protein and RNA molecules. DNA sequence information is important to scientists investigating the functions of genes.

**Taxonomic:**Taxonomy is the practice and science of categorization or classification. In a wider, more general sense, it may refer to a categorisation of things or concepts, as well as to the principles underlying such a categorisation. Taxonomy organizes taxonomic units known as "taxa" (singular "taxon")."

**M**[**icrobiome**](https://edu.t-bio.info/glossary/microbiome/)**:**The microbiome consists of microbes that are both helpful and potentially harmful. Most are symbiotic (where both the human body and microbiota benefit) and some, in smaller numbers, are pathogenic (promoting disease). In a healthy body, pathogenic and symbiotic microbiota coexist without problems.

**T**[**raits**](https://edu.t-bio.info/glossary/trait/)**:**A quality that forms part of your character or personality

**Agrigenomics:** Agricultural genomics uses technology to study and classify the genes of plants in order to develop new, specific varieties of crops. These new crops can help address various food problems around the world and can improve the productivity and sustainability of many plant varieties and livestock productions.

**Bacillus thuringiensis:**Bacillus thuringiensis (Bt) is a species of bacteria that lives in soil. It makes proteins that are toxic to some insects when eaten, but not others. The proteins are not toxic to humans because, like all mammals, we cannot activate them. Bt is not toxic to non-target wildlife.

**Biopharmaceuticals:** A biopharmaceutical, also known as a biologic(al) medical product, or biologic, is any pharmaceutical drug product manufactured in, extracted from, or semi synthesized from biological sources. Different from totally synthesized pharmaceuticals, they include vaccines, whole blood, blood components, allergenics, somatic cells, gene therapies, tissues, recombinant therapeutic protein, and living medicines used in cell therapy. Biologics can be composed of sugars, proteins, nucleic acids, or complex combinations of these substances, or may be living cells or tissues. They (or their precursors or components) are isolated from living sources—human, animal, plant, fungal, or microbial. They can be used in both human and animal medicine.

**RNAi:**RNA interference (RNAi) is a biological process in which RNA molecules are involved in sequence-specific suppression of gene expression by double-stranded RNA, through translational or transcriptional repression. ... RNAi is now known as precise, efficient, stable and better than antisense therapy for gene suppression.

**CRISPR:** CRISPR is a molecular tool that can act within the genome of a living organism to add, remove, or exchange a DNA sequence. In other words, it enables scientists to delete or rewrite mutations in DNA. ... This precision is why CRISPR is called "gene editing" rather than "gene therapy."

**Metagenomics:** Metagenomics enables the study of all microorganisms, regardless of whether they can be cultured or not, through the analysis of genomic data obtained directly from an environmental sample, providing knowledge of the species present, and allowing the extraction of information regarding the functionality of microbes.

**Genetics:**Genetics is the study of heredity in general and of genes in particular. Genetics forms one of the central pillars of biology and overlaps with many other areas, such as agriculture, medicine, and biotechnology.

**Ascertainment bias:**Ascertainment bias happens when the results of your study are skewed due to factors you didn’t account for, like a researcher’s knowledge of which patients are getting which treatments in clinical trials or poor [Data Collection Methods](https://www.statisticshowto.com/data-collection-methods/) that lead to non-[representative samples](https://www.statisticshowto.com/representative-sample/" \t "_blank).

**Microsatellites:**A microsatellite is a tract of repetitive DNA in which certain DNA motifs (ranging in length from one to six or more base pairs) are repeated, typically 5–50 times. Microsatellites occur at thousands of locations within an organism's genome.

**American Gut Project:**The project is an active research effort in which scientists aim to work with citizen scientists, as well as academic and industry researchers, to understand the life associated with humans and to advance microbiome science.

[**Trans**](https://www.nature.com/subjects/transcriptomics)**c**[**riptomic**](https://www.nature.com/subjects/transcriptomics)**:** Transcriptomics allows identification of genes and pathways that respond to and counteract biotic and abiotic environmental stresses. The non-targeted nature of transcriptomics allows the identification of novel transcriptional networks in complex systems.

**E**[**pigenomic**](https://www.nature.com/subjects/epigenomics?WT.ac=search_subjects_epigenomics)**:**Epigenomics is the study of the complete set of epigenetic modifications on the genetic material of a cell, known as the epigenome. The field is analogous to genomics and proteomics, which are the study of the genome and proteome of a cell.

[**Metagenomic**](https://www.nature.com/subjects/metagenomics?WT.ac=search_subjects_metagenomics)**:** Metagenomics is the study of a collection of genetic material (genomes) from a mixed community of organisms. Metagenomics usually refers to the study of microbial communities.

**M**[**etabolomi**](https://www.nature.com/subjects/metabolomics?WT.ac=search_subjects_metabolomics)**c:** Metabolomics is defined as the systematic study of all chemical processes concerning metabolites, providing characteristic chemical fingerprints that specific cellular processes yield, by means of the study of their small-molecule metabolite profiles.

**P**[**roteomic**](https://www.nature.com/subjects/proteomics?WT.ac=search_subjects_proteomics)**:**Proteomics is the large-scale study of proteins. Proteins are vital parts of living organisms, with many functions. The proteome is the entire set of proteins produced or modified by an organism or system. Proteomics enables the identification of ever-increasing numbers of proteins.

**S**[**tructural biology**](https://www.nature.com/subjects/structural-biology?WT.ac=search_subjects_structural_biology)**:**Structural biology is the study of how biological molecules are built. Using a variety of imaging techniques, scientists view molecules in three dimensions to see how they are assembled, how they function, and how they interact.