

Introduction to Bioinformatics

Bioinformatics, as related to genetics and genomics, is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences. Scientists and clinicians use databases that organize and index such biological information to increase our understanding of health and disease and, in certain cases, as part of medical care. In short, Application of computer science, statistics, and math to analyze biological data.

Role: Bioinformatics. The role of bioinformatics in biological research can be compared with the role of data analysis in the age of information and the Internet. In earlier days, the primary challenge was getting to the information. Advances in reading DNA sequences have lowered that barrier substantially. Going forward, the challenge is how to understand and interpret the information that has been collected. Because the data sets are large, whether you're talking about information about website visits or the human genome, computer-based methods are the default approach. In the end, bioinformatics work with human genomes seeks to discover practical insights about human health and biology with all its complexity.

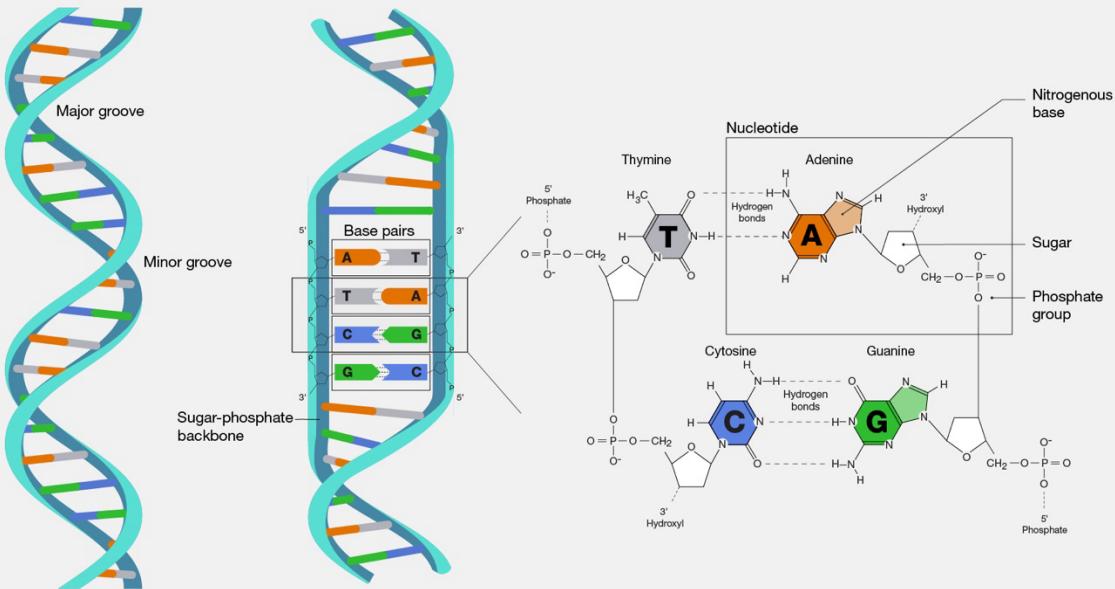
Importance:

- ◆ Human Genome Project: 3 billion base pairs needed computers.
- ◆ Modern medicine: Cancer genomics, personalized therapy.
- ◆ Agriculture: Drought-resistant crops.
- ◆ Vaccine invention: Corona virus

Think of DNA as a huge **book** written with only four letters (A, T, G, C). Bioinformatics is like **Google Search + Grammar + Translator** for that book. [<https://youtu.be/L9NriBoubWE>]

DNA: Is there a more amazing molecule than DNA? It makes each of us who we are. The more scientists understand it, the more we all understand ourselves, one another, and the world around us. For example, did you know that we are all far more alike than we are different? In fact, the DNA from any two people is 99.9% identical, with that shared blueprint guiding our development and forming a common thread across the world. The differing 0.1% contains variations that influence our uniqueness, which when combined with our environmental and social contexts give us our abilities, our health, our behavior. How can one, single molecule contain so much mystery and wonder? We are only beginning to understand the answer to that question, which is what makes the study of DNA so exciting.

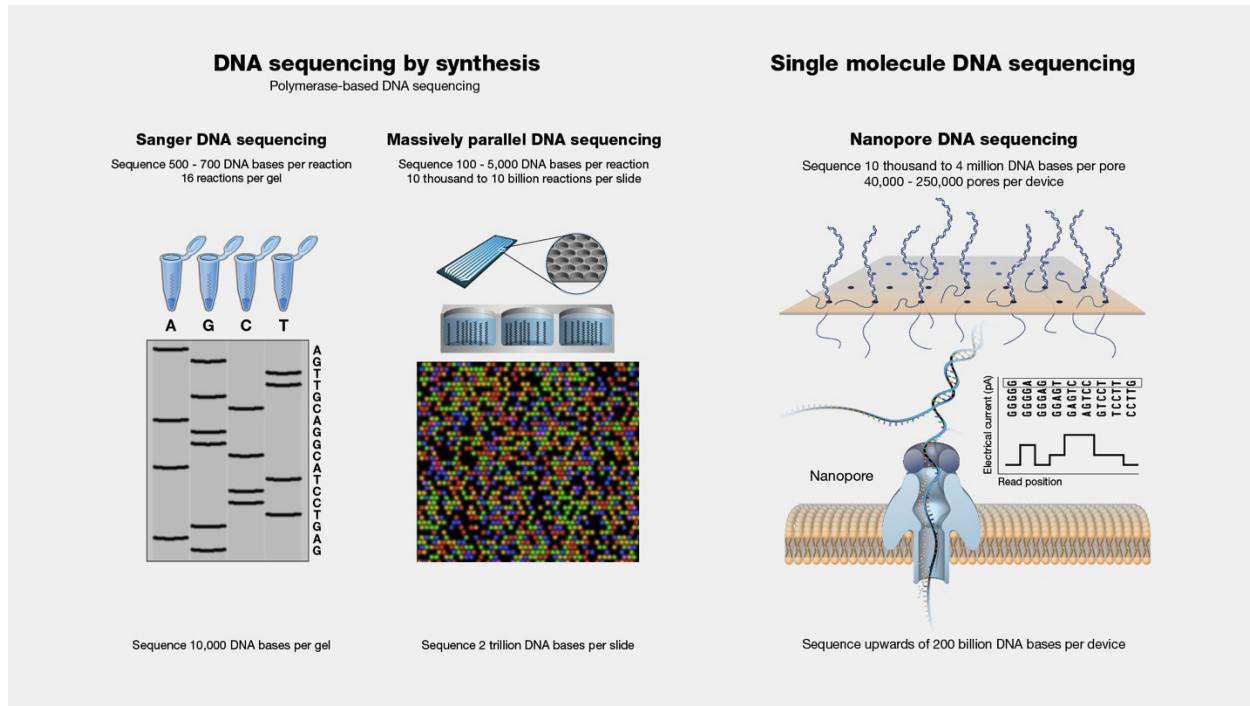
Deoxyribonucleic acid (DNA)



Deoxyribonucleic acid (abbreviated DNA) is the molecule that carries genetic information for the development and functioning of an organism. DNA is made of two linked strands that wind around each other to resemble a twisted ladder — a shape known as a double helix. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are connected by chemical bonds between the bases: adenine bonds with thymine, and cytosine bonds with guanine. The sequence of the bases along DNA's backbone encodes biological information, such as the instructions for making a protein or RNA molecule.

GENOME

The genome is the entire set of DNA instructions found in a cell. In humans, the genome consists of 23 pairs of chromosomes located in the cell's nucleus, as well as a small chromosome in the cell's mitochondria. A genome contains all the information needed for an individual to develop and function.



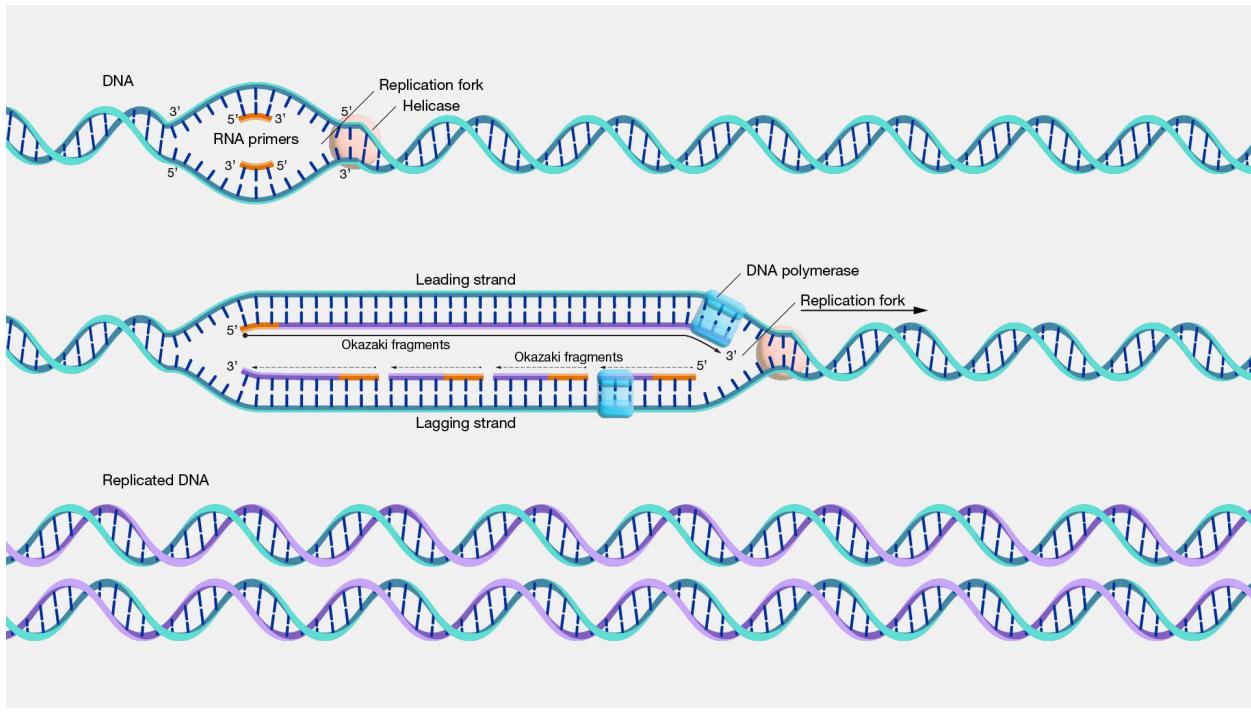
DNA sequencing refers to the general laboratory technique for determining the exact sequence of nucleotides, or bases, in a DNA molecule. The sequence of the bases (often referred to by the first letters of their chemical names: A, T, C, and G) encodes the biological information that cells use to develop and operate. Establishing the sequence of DNA is key to understanding the function of genes and other parts of the genome. There are now several different methods available for DNA sequencing, each with its own characteristics, and the development of additional methods represents an active area of genomics research.

Central Dogma of Biology

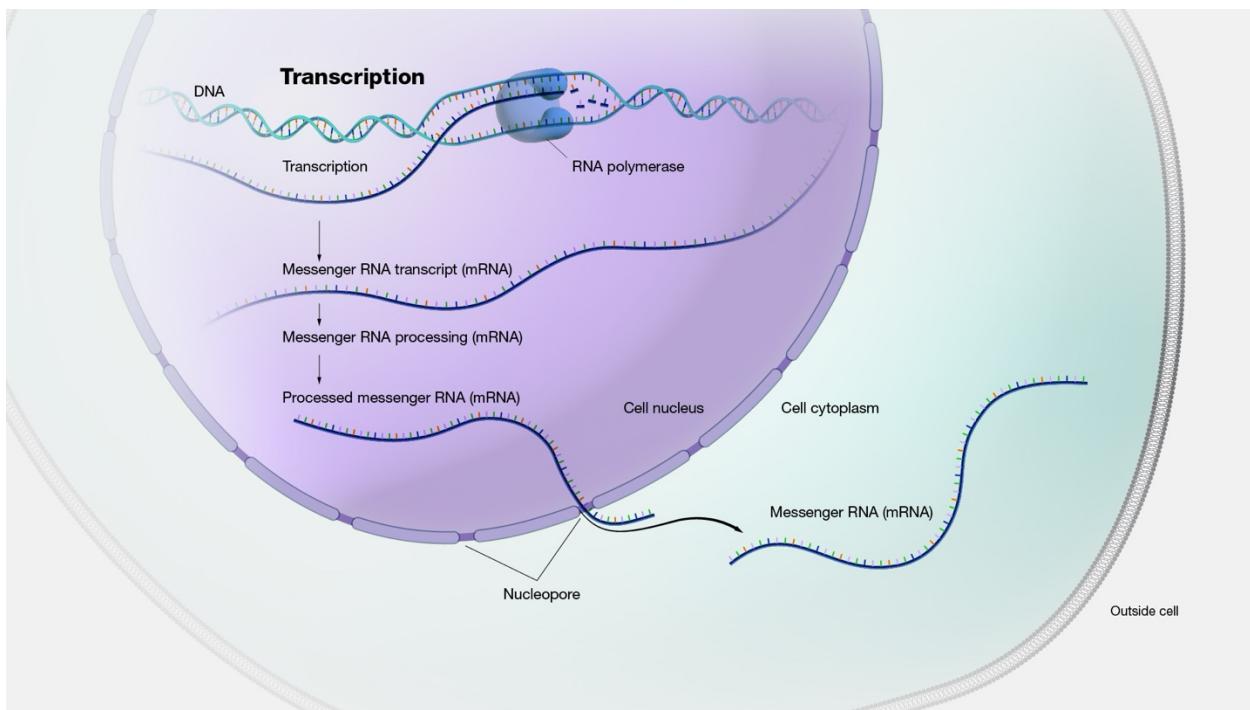
- ◆ DNA **stores** information.
- ◆ RNA **transfers** information.
- ◆ Protein **performs** the function.

DNA (Replication) → RNA (Transcription) → Protein (Translation)

DNA Replication is the process by which the genome's DNA is copied in cells. Before a cell divides, it must first copy (or replicate) its entire genome so that each resulting daughter cell ends up with its own complete genome.



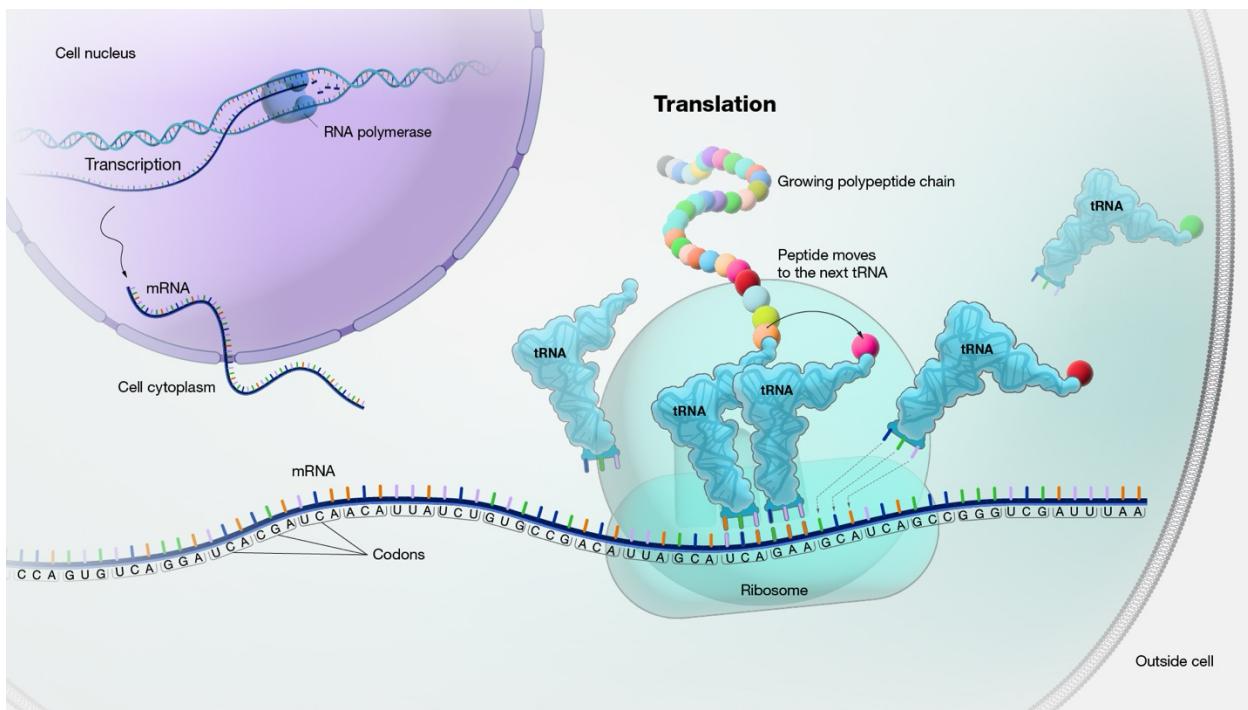
Transcription, as related to genomics, is the process of making an RNA copy of a gene's DNA sequence. This copy, called messenger RNA (mRNA), carries the gene's protein information encoded in DNA. In humans and other complex organisms, mRNA moves from the cell nucleus to the cell cytoplasm (watery interior), where it is used for synthesizing the encoded protein.
[<https://youtu.be/kHlgG035Id8>]



Example: Insulin hormone.

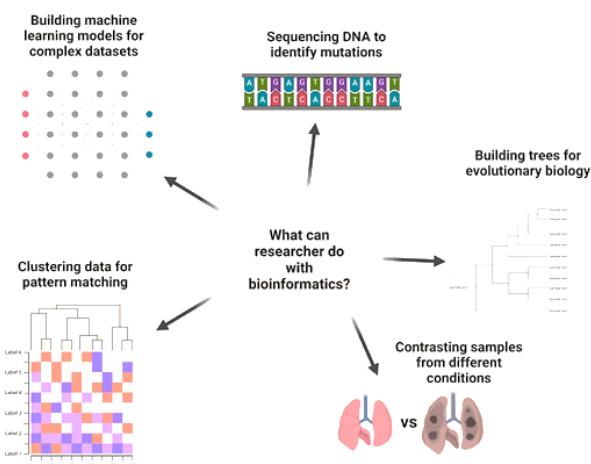
Gene (DNA) → mRNA → Insulin protein → Controls blood sugar.

Translation, as related to genomics, is the process through which information encoded in messenger RNA (mRNA) directs the addition of amino acids during protein synthesis. Translation takes place on ribosomes in the cell cytoplasm, where mRNA is read and translated into the string of amino acid chains that make up the synthesized protein. [<https://youtu.be/oCp9IK6iBTo>]



Quick Review Questions

1. What is bioinformatics in one sentence?
2. What are the three main stages of the central dogma?
3. Which database can you use to find nucleotide sequences?



Lab Session: Introduction to Databases

Objective: Explore NCBI and GenBank.

1. Open [NCBI](#).
2. Search for a gene (example: BRCA1 human gene).
3. Show FASTA format (plain sequence of A, T, G, C).
4. Show GenBank format (includes annotations).

Task: Find the sequence length of the BRCA1 gene from NCBI.

References: <https://www.genome.gov/genetics-glossary/Bioinformatics>

**Cell Structure (DNA, RNA, TRANSCRIPTION, TRANSLATION, REPLICATION)

**Cell Chemistry (PROTEIN, AMINO ACID)

**VIRUS