Bioinformatics is divided into five areas:

1. **Functional genomics**. How do genes and intergenic segments contribute to metabolic pathways (aka gene expression patterns)? Examples include new species
2. **Structural genomics**. Understand 3D structures of proteins encoded by the genes.
3. **Comparative genomics**. Compare complete genetic material of one organism with another to better understand how species evolved and the role of coding and non-coding genes
4. **DNA microarrays**. A lab technique to find if a DNA of an individual has gene has mutation
5. **Medical informatics**. Converge medical and CS tools to improve patient care.

Three Branches of Bioinformatics:

1. Computational biology. Data-based solutions in bioinformatics
2. Genetics. Study and variation of inherited characteristics
3. Genomics. Branch of molecular biology concerned with structure, function, evolution, and mapping of genomes.

Biopython is a library of bioinformatics tools that you can use for predictive models and analytics.

Building Blocks of Bioinformatics

Tissue. Section of an organ that consists of a largely homogeneous population oof cell types. Organs are multifunctional and therefore cells are specialized to perform different functions.

Why homogeneous cell type? Finding a section of organ with homogeneous cell types ensures that the gene expression profiles extracted from the cells will accurately represent the class of cells that makeup the tissue.

Cell. Basic unit of a living organism. It's a product of the expression of its genes. A cell contains DNA, RNA, small molecules of nutrients and metabolites.

Gene. Specific segment of DNA on a chromosome that tells the cell how to function. Basic physical and functional unit of heredity. Some genes act as instructions to make molecules called proteins. And others (non-coding) assist in regulating the gene expression.

Gene Expression. The process by which information encoded in the gene is used to assemble a protein molecule. It's a combination of transcription and translation.

Transcription. The process by which cells make an RNA copy (called messenger RNA or mRNA) of a piece of DNA. This mRNA carries the genetic information needed to make proteins in a cell.

Translation. Translating the sequence of mRNA into a sequence of amino acids during protein synthesis

ACGT. A DNA molecule has four types of bases. They are Adenine (A), Cytosine (C), Guanine (G), and Thymine (T).

Base Pairs. DNA molecule consists of two strands wound around each other with each strand held together by bonds between the four bases ACGT. These bases form specific pairs (A with T and G with C) called base pairs.

Human Genome. A human genome has **three billion** base pairs of DNA distributed across 23 chromosomes.

Chromosome. Made up of proteins and DNA and organized into genes. Chromosomes ensure that DNA is copied exactly during a cell division. Human chromosomes range from 50 million to 300 million base pairs.

5' and 3' in DNA. Each end of DNA has a number. One end is referred to as 5' (five prime) and the other end is 3' (three prime). The 5' and 3' represent # of carbon atoms in deoxyribose sugar molecule to which a phosphate group bonds.

KOH-don (Condon). A Condon signals start/stop of translation. It is a sequence of three consecutive nucleotides in a DNA or RNA molecule that codes for specific amino acids.

DNA vs RNA (Key Differences).
DNA: Several million Base Pairs, fairly stable, self-replicating
RNA: several thousand Base Pairs, reactive, synthesized by transcription

Gene vs. Genome. Gene is a specific segment of DNA that tells Cells how to function. A genome is the entirety of the genetic material inside an organism. A human genome consists of 20K-25K genes.

Human Genome. A human genome consists of 20K to 25K genes. Every person has two copies of a gene (one inherited from each parent). Most of the genes are common across all people. Less than 1% of the genes are slightly different in different people and these small differences contribute to each person's unique features.

Diseases. Variations in genes (from less than 1% that differ) can influence health, appearance, and risk of developing certain diseases. Example. Sickle cell disease.

Trends in Bioinformatics.
Single-cell RNA-sequencing (scRNA-seq). It captures global state of all mRNA expressions within a tissue up to a single-cell resolution. The scRNA-seq assists in understanding

1. How does a tissue composition and function change during development of a disease?
2. The heterogeneity and complexity of RNA transcripts (coding for proteins or non-coding for other functions) within individual cells.
3. Composition of different cell types and functions within a tissue.

RNA-seq provides "average" of the expression profiles of 1000s of cells.

Disease. Genetic variation may contribute to a disease largely through misregulation of gene expression. A gene expression implies the information encoded in a gene is turned into a function. This occurs via transcription of RNA molecules that code for proteins or non-coding RNA molecules for other functions within a cell. Goals of scRNA-Seq:

1. Measure the distribution of expression levels for each gene across a population of cells.
2. Measure transcriptional differences across and within groups of cells
3. Resolve single-cell heterogeneity.

Applications: analysis of cancer evolution and mechanisms of therapeutic resistance.