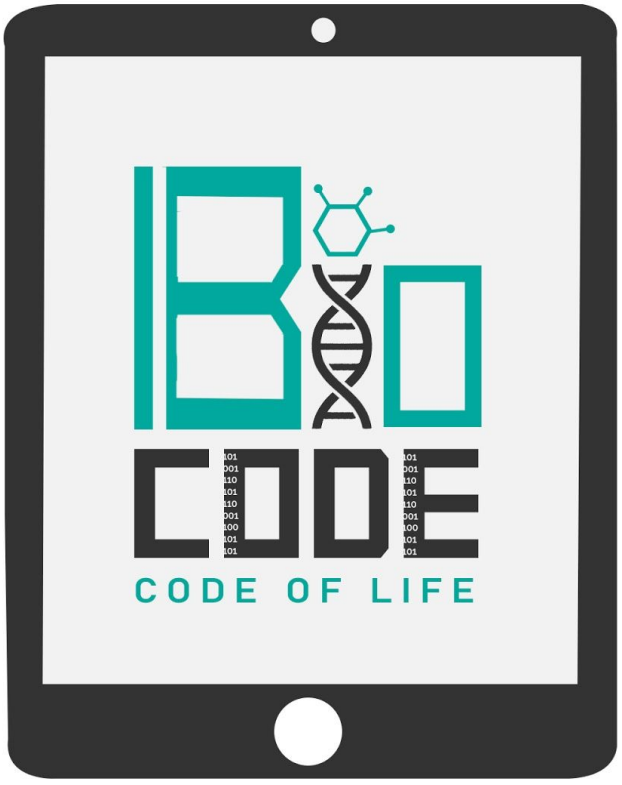

LEARN PYTHON & R FOR BIOINFORMATICS

National Center for Biotechnology (NCBI)

Introduction:

The **National Center for Biotechnology Information** (**NCBI**) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). It is primary database, primary databases are archival in nature. They consist of experimentally derived data such as nucleotide sequence, protein sequence or macromolecular structure. Experimental results are submitted directly into the database by researchers.

In this video we provide an introduction on NCBI. The NCBI is a host database, contains series of 32 sub-databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Each sort of database contains different data. This video consists step-wise information regarding the functionality of NCBI, 32 sub-databases of NCBI and a sequence analysis on gene LCT.

Steps:

- Homepage of NCBI contains various sub-databases from which we can retrieve data of our interest and query box to search.
- There are total 32 sub-databases, include Biocollections, Assembly, Protein, Gene, Nucleotide, BioProject, OMIM, Books and various others having different data.
- In the query box type **LCT,** a key word for a gene that makes an enzyme **Lactase.**
- Search it through all the databases.
- Search result shows the reference of key word LCT in all databases. Mentions Homo sapiens as the most probable interest of search,
- NCBI divides the results in different categories of sub-databases.
- Categories are Literature having books, articles, research papers in which LCT is mentioned, Gene providing gene from different organisms, Geo DataSets Geo profile for gene expression analysis, HomoloGene for homology analysis and PopSet for gene population analysis in set of organisms. Other categories Proteins for proteomics, Chemicals for finding substances compounds and their properties, Genetics for genotype-phenotype relationships, Genomes for genomic information.
- From Gennome category click on Nucleotide database showing total 7,856 reference nucleotide sequences of LCT gene in different organisms.
- It shows a most probable search interest as LCT gene of Homo sapiens and below that there's list of 20 reference sequences out of 7,865 reference sequences of LCT gene on first page that can be extend up to 200 reference per page.

Summary:

NCBI is a hosting database of total 32 sub-databases, having different datasets that are collected by researchers from all the world. In this video we discussed the functionality of NCBI and the search categories of NCBI. We performed a sequence analysis on a gene called LCT and searched it through all the databases and found the references in different database and then moved to Nucleotide database from Genome category where we watched the reference sequences of LCT gene.