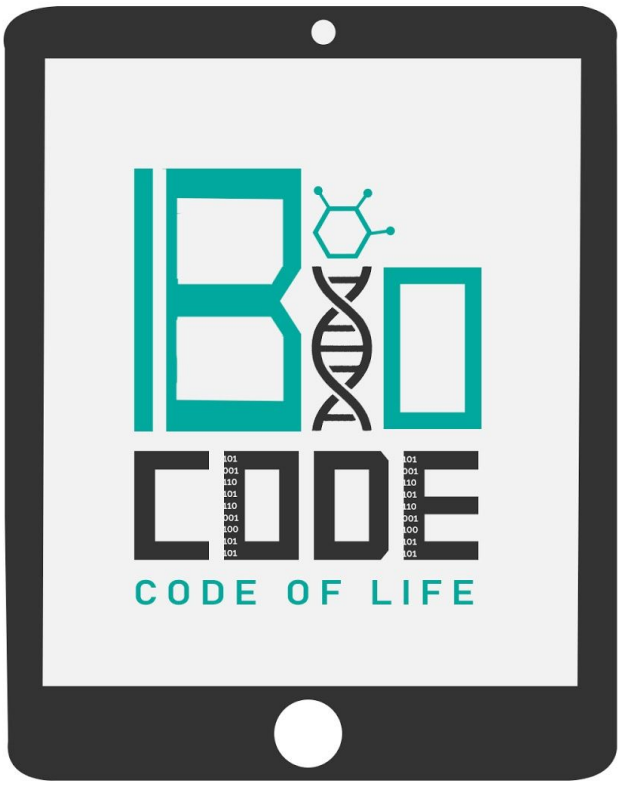

LEARN PYTHON & R FOR BIOINFORMATICS

Sequence analysis on NCBI

Introduction

In this video we perform the sequence analysis of LCT gene on NCBI. NCBI's sub-database GenBank allows to search a particular gene of our interest and provide sets of information linked to it in GenBank format and FASTA format for sequence.

Steps and Analysis Report:

- Enter the key word LCT in a query box and click search.
- From the NCBI result page click the hyperlink LCT-Lactase of Homo sapiens.
- It opens a GenBank page of it.
- First it shows the name of the gene as the title and below that it gives the NCBI reference sequence database accession number with its version.
- The major information that it gives about the LCT gene is LOCUS that is accession of gene on NCBI, total base pairs molecule type i.e. mRNA and its recent update/version.

- Then it tells the DEFINITION that is the name of the gene and molecule type, after that it tells VERSION that tells the total changes or updates in the sequence. Then it tells KEYWORDS.
- After that it tells the SOURCE of sequence that is Homo sapiens, then it tells REFERNCE with total base pairs.
- Then it shows the information about authors and title of the article in which the variant is detected.
- If we go to the FEATURES portion of the page it has the SOURCE section that tells total base pairs, organism, molecule type, taxon number, chromosome number and its position on that chromosome.
- From FEATURES gene section tell the total base pairs it synonyms and reference IDs to different resources.
- Exon section from FEATURES tells position of exon in a gene, misc feature tells the presence of stop codon in a sequence then CDS section tells the position of coding DNA. It also gives the EC numbers i.e. enzyme commission number, its product, protein id and references to other database and finally the amino acid sequence of CDS.
- It also tells the predicted position of signal peptide in a sequence, below that it shows the different exon regions in the gene and at the end of the report it gives a complete mRNA sequence.

Summary:

In this video we performed a sequence analysis on the mRNA sequence of LCT gene of Homo sapiens and retrieved various types of information related to it. We found out its total base pairs, different references to different databases, EC numbers, exons, enzyme product, predicted position of signal peptide, stop codon presence, amino acid sequence and a complete mRNA sequence of it.