BIO
CODE
CODE OF LIFE

LEARN PYTHON &
R FOR BIOINFORMATICS

RefSeq: NCBI Reference Sequence Database

NCBI's reference sequence (RefSeq) database (http://www.ncbi.nlm.nih.gov/RefSeq/) is a curated non-redundant collection of sequences representing genomes, transcripts and proteins. Annotations include coding regions, conserved domains, tRNAs, sequence tagged sites (STS), variation, references, gene and protein product names, and database cross-references.

Steps:

- From the RefSeq Access section click on the hyperlink *Human Genome Resources and Download.* It displays different reference datasets like Reference Genome Sequence, RefSeq Reference Genome Annotation, RefSeq Transcripts, RefSeq Proteins, ClinVar dbSNP and dbVar under two assemblies GRCh38 and GRCh37. It also shows the hyperlinks of different database.
- Click on the FASTA of GRCh38 of RefSeq.

- It starts downloading the non-redundant datasets of transcripts of this assembly.
- From the RefSeq homepage enter Homo sapiens in the query section. It shows all the non-redundant curated datasets.
- From the RefSeq Access section click on ftp RefSeq.
- It shows parent directories of reference sequences of different organisms.
- Click on the directory of Homo sapiens, it shows different datasets in Homo sapiens directory.
- Click on the RefSeq gene it displays all the genes available in non-redundant genomic datasets.
- From the Announcements, it shows the different non-redundant datasets, click on it.
- It takes to the datasets of different organisms.
- Click on the viral, it provides non-redundant datasets of viruses.
- Open the RefSeq Gene database, enter the LCT in the query section. Click on search.
- It shows LCT genes with different accession numbers from different organisms.

Summary:

In this video we discussed the RefSeq database of NCBI. We described the use of RefSeq database, its features and described the data it provides to retrieve. We also searched for different datasets. Then we described the use of ResSeq Gene database and searched for LCT gene to analyse the non-redundancy of results it provides. We also explored the recent Announcement of non-redundant datasets.