**Introduction to off-target prediction pipeline**

The pipeline is developed for the prediction of potential off-target sites of a gRNA in Human genome and the prediction of gRNA efficiency to ensure the high gRNA efficiency in experimental conditions. All the scripts of off-target prediction pipeline are written and implemented in python. For the implementation of off-target prediction pipeline in user system requires various python packages and tools which are given below.

**Requirements:**

* FASTA36
* ViennaRNA package
* SeqIO, SeqUtils, Seq module of Bio package
* Biopython
* Python packages: Pandas, Numpy, sci-kit learn, joblib
* Human reference genome (Hg37)

This pipeline can be implemented using a following commands after installation of above given packages in your **Linux-based system**. Off-target prediction pipeline takes 3 inputs which are user input file containing gRNA sequence/s, reference genome and the output file name. All the file names should be input in the command with filename as well as their location.

**Usage:**

**For Potential off-target site prediction:**

Script required for the implementation of potential off-targets prediction pipeline is Fasta36\_SeqSearch.py. which takes three arguments as given:

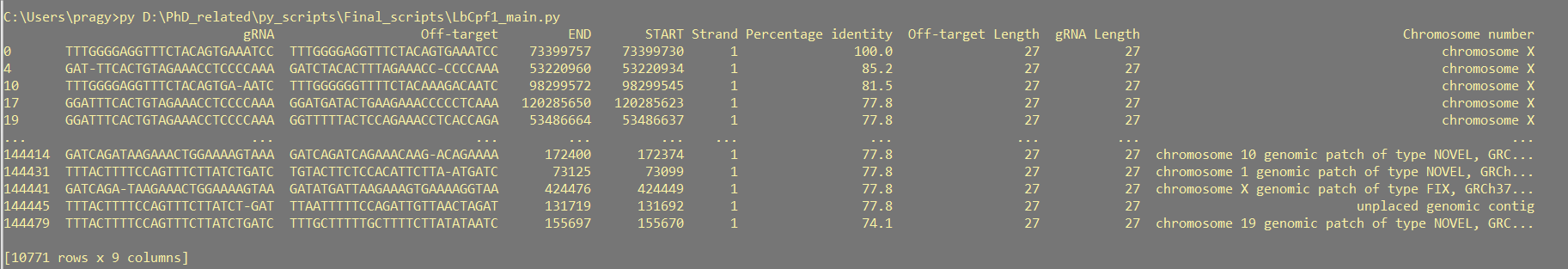
>> cd \LocationOfOff-targetPredictionPipeline\off-target\_prediction\_pipeline\

>> Python3 Fasta36\_SeqSearch.py input\_filename.fasta Reference\_genome.fna Output\_file.csv

Time taken by the pipeline for the prediction of Potential off-target sites – **5-10 minutes**

**Output:**

Command output will also be summarised in the terminal screen as shown below. Detailed output will be stored in the file which can be accessed by the user in the given folder location. Output file stores the potential off-target sites with other features such as input gRNA sequence, predicted off-target sequence, start and end of the sequence in Human genome, strand, chromosome number, length of off-target sequence, length of gRNA sequence and percent identity between gRNA and predicted off-target sequences.



**For target efficiency prediction:**

Script required for the implementation of target efficiency prediction pipeline are LbCpf1\_main.py and AsCpf1\_main.py for LbCpf1 and AsCpf1 species respectively. which also takes three arguments similar to Fasta36\_SeqSearch.py implementation as shown below.

>> cd \LocationOfOff-targetPredictionPipeline\off-target\_prediction\_pipeline\

**For the prediction of target efficiency for AsCpf1 species**

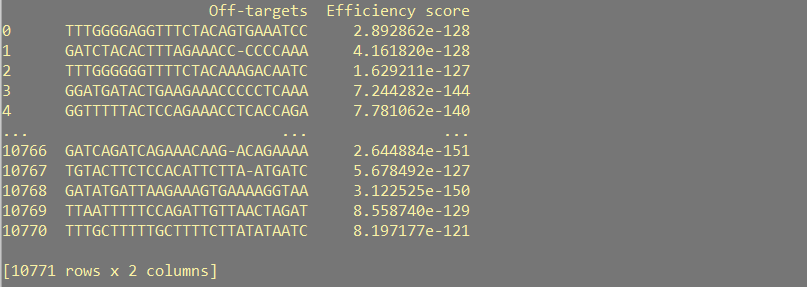
>> Python3 AsCpf1\_main.py input\_filename.fasta Reference\_genome.fna Output\_file.csv

Or

**For the prediction of target efficiency for LbCpf1 species**

>> Python3 LbCpf1\_main.py input\_filename.fasta Reference\_genome.fna Output\_file.csv

Time taken by the pipeline for the prediction of target efficiency – **between 5-8 minutes**



Note: Working in a off-target prediction pipeline folder for all the implementations is recommended.