

**MINI PROJECT REPORT**

**OLABODE KAOSARA OMOWUNMI**

[omowuniolabode5@gmail.com](mailto:omowuniolabode5@gmail.com)

**TASK 1**

**Downloading a Biological Sequence from NCBI and View/Edit It**

Getting a sequence data is a cornerstone of bioinformatics research which allows for local curation and preprocessing, which is often necessary for downstream analysis.

**Objective**: Download the human TNF gene sequence and view it using BioEdit.

A screenshot of a computer

Description automatically generated

Here is a pictorial representation of the Human TNF gene sequence downloaded from NCBI database and viewed with the BioEdit Software.

**TASK 2**

**Translation of the Human TNF Gene Sequence into Amino Acids**

This process is also known as protein translation, and it is crucial for understanding the protein coding potential of an amino acid sequence.

**Objective**: Translate the DNA sequence of the TNF gene into an amino acid sequence.

A screenshot of a computer

Description automatically generated

Here is a pictorial represntation of the translation process output.

The Frame 1 method of grouping sequence into codon was used to achieve this translation (amino acid generation). As evident, the first codon (three sets of amino acids) codes for S-Serine, the second codes for R- Arginine and so on.

**TASK 3**

**Finding ORFs (Open Reading Frames) in the Human TNF Gene Sequence**

Open Reading Frames (ORFs) are stretches of DNA that have the potential to code for proteins. Identifying these ORFs within a DNA sequence is crucial for indicating potential protein-coding regions. Once identified, ORFs can be translated into amino acid sequences, which can then be analyzed to predict protein structure, function, and potential interactions.

**Objective**: Identify the ORFs within the TNF gene sequence.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Pictorial representation of the ORFs between position 1 – 1,500

After identifying the ORFs in the Human TNF gene sequence, it was observed that the first ORF has 143 amino acids between position 178 – 606, and the protein sequences are listed right below it. Between position 1 – 1500, there are 8 ORFs.

**TASK 4**

**Analyzing Sequence Composition (Nucleotide or Amino Acid Frequencies)**

**Objective**: Analyze the nucleotide composition of the TNF gene sequence.

A screenshot of a computer

Description automatically generated

The frequencies of the nucleotides present in the gene sequence have been displayed with several information such as the length of the base pairs, the molecular weight of single and double stranded DNA molecule, GC and AT percentage ratio as well as the number of times each nucleotides appear. This information is crucial in drug discovery, and protein analysis.

A screenshot of a computer screen

Description automatically generated

It is evident in the graph that there is a higher GC content in the Human TNF gene sequence which means that there is a higher percentage of coding regions in the sequence.

**TASK 5**

**Identifying Transcription Factor Binding Sites (TFBS) Using the PROMO Tool**

TFBS are essential components of regulatory elements like promoters and enhancers. By locating these sites, researchers can gain insights into how gene expression is controlled and help uncover the underlying molecular mechanisms of action in diseases. TFBS can be potential targets for drug development.

**Objective**: Identify potential transcription factor binding sites in the TNF gene promoter region.

A screenshot of a computer

Description automatically generated

The TBSs are displayed with their locations, there is also an information on the percentage distribution of nucleotides contained in the gene sequence.

A screenshot of a computer

Description automatically generated

This display popped up after clicking on one of the TBSs to view the specific sequence of the binding site

A screenshot of a computer screen

Description automatically generated

Further information on where the transcription factors bind in the Human TNF gene sequence

**TASK 6**

**Searching for Functional Motifs in the Human TNF Genome or Transcriptome Using MEME Suite**

Identifying these motifs within a genome or transcriptome is crucial for understanding gene function, regulation, and protein structure.

**Objective**: Search for functional motifs in the TNF gene sequence using MEME Suite.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Meme suite came up with three motifs as instructed for the Human TNF gene sequence. Information such as the E-value, and the number of the binding sites for each motif (there are represented in colored bars under the motifs location tab) are displayed. For further information on each motif, the more icon is clicked, and the result is displayed below. It gave more information on the type of strand present, where the motif occurrence starts and other pertinent information.A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**TASK 7**

**Predict Coding/Non-Coding Regions in a Genome Using GENSCAN**

Coding regions (exons) are essential for protein synthesis, non-coding regions can reveal regulatory elements like promoters, enhancers, and silencers, which play crucial roles in gene expression. Identifying these regions helps locate genes within the vast expanse of genomic DNA.

**Objective**: Predict the coding and non-coding regions within the TNF gene sequence.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The output displayed gave information on the length of the basepairs which is 2815bp, the GC content of the sequence which is analogous to the previously derived GC content. There is another section giving information about the exons, the type of exons present, +/- strand, start and stop position of the exon, length of the exon, type of frame, phase, initiator signal score, terminal signal score, coding region sdcore, probability/quality of an exon.

**TASK 8**

**Converting Between Sequence File Formats Using BioEdit (FASTA to PHYLIP)**

**Objective**: Convert the TNF gene sequence from FASTA format to PHYLIP format

A screenshot of a computer

Description automatically generated

The Human TNF gene sequence file in FASTA file format was converted to a PHYLIP file format and the output was viewed in a notepad as displayed above.