BioinformHER: Module1, Mini project report

Sampurna Mukherjee

# Task 1: Downloading and visualizing TNF gene sequence

**Output:**

## A screenshot of a computer Description automatically generatedScreenshot:

## File generated:

**Fasta file** containing TNF gene sequence: TNF\_sequence.fasta

# Task 2: Translation of the TNF gene sequence

**Output:**

## **A screenshot of a computer Description automatically generated**Screenshot:

# Task 3: Finding ORFs in TNF gene

**Output:**

## A screenshot of a computer Description automatically generatedScreenshot:

**File generated: orf\_task3.txt**

## Interpretation:

* There are multiple open reading frames in this gene.
* The reading frames are present in 6 possible directions, three forward (shown as 1,2,3) and three reverse (shown as -1,-2,-3)
* Some reading frames encode shorter peptides whereas others code for longer proteins.
* Shorter peptides and longer proteins are present in both directions of DNA strand.
* No ORF consists of the entire DNA sequence, signifying non-coding regions might be present in this gene.

# Task 4: Nucleotide composition and GC content analysis of TNF gene

**Output:**

## Screenshot:

## Interpretation:

* GC content is higher than 50% in TNF gene – GC content is 52.81%
* The presence of more GC nucleotides and therefore less AT nucleotides might have functional implications influencing codon usage and codon bias. (Hershberg, 2016)

# Task 5: Searching for Transcription factor binding sites

**Output:**

## A screenshot of a computer Description automatically generatedScreenshot:

## Interpretation:

* Multiple transcription factor binding sites are present in TNF gene sequence.
* This signifies that TNF gene is highly regulated in terms of gene expression.
* The different transcription factors turn gene expression “on” or “off” based on different factors including cell signalling.
* Thus, this shows that TNF gene expression is modulated by multiple transcription factors depending on cell conditions.

# Task 6: Searching for functional motifs

**Output:**

## Screenshot:

## Interpretation:

Three key motifs has been found by the MEME algorithm using default settings. The location of occurrence of these motifs in the gene is shown in the Motif location panel – the first motif shown in red, the second in cyan and the third in green. All three motifs are primarily found in the forward strand of the gene, though the second motif is also present in two locations in the reverse strand, as shown in the motif location. The first motif is located at the start of the genomic sequence , the second is dispersed across the gene and the third is present at the end of the genomic sequence. Thus, these motifs are probably associated with transcription initation, continuation and termination respectively.

# Task 7: Predicting coding/non-coding areas of genome using GENSCAN

**Output:**

## Screenshots:

A blue screen with white text

Description automatically generated

**A screenshot of a computer program

Description automatically generated**

## Interpretation:

* 1 initial exon is found starting at 221 base pairs, followed by two internal exons present at 1013 base pair and 1246 base pair starting points followed by the terminal exon starting at 1595 base pairs.

Reference: https://www.nature.com/scitable/topicpage/rna-splicing-introns-exons-and-spliceosome-12375/

* The polyA transcription termination signal is present , starting at 2792 base pairs.
* All the exons and the polyA termination signal is present at the forward strand of TNF gene
* Thus, the transcription initiation site, splicing sites and termination sites are detected by this GENSCAN output.

# Task 8: Conversation from FASTA to Phylip format

**Output:**

## Screenshot:

**File generated: TNF\_sequence\_phylip.phy**