Comprehensive Sequence Analysis of the Human TNF Gene

Project Title: Comprehensive Sequence Analysis of the Human TNF Gene

Project Overview

In this mini project, I performed a series of bioinformatics tasks using the human TNF gene as my sequence of interest. Started from downloading the sequence then translating it, finding ORFs, analyzing sequence composition, identifying transcription factor binding sites. Searching for functional motifs. Predicting coding or non-coding regions and converting sequence file formats. I used tools such as NCBI BLAST, BioEdit, PROMO, MEME Suite and GENSCAN in this project.

Task 1: Download a Biological Sequence from NCBI and View/Edit It

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

Steps

- Accessed the NCBI homepage at NCBI.
- Searched for the human TNF gene using the term 'human TNF gene.'
- Located the correct sequence record (e.g., 'Homo sapiens TNF').
- Downloaded the sequence in FASTA format.
- Since the sequence contained several versions, I selected the most acceptable version in scientific field (NCBI Reference Sequence: NC_000006.12)
- Opened the sequence in BioEdit

Output

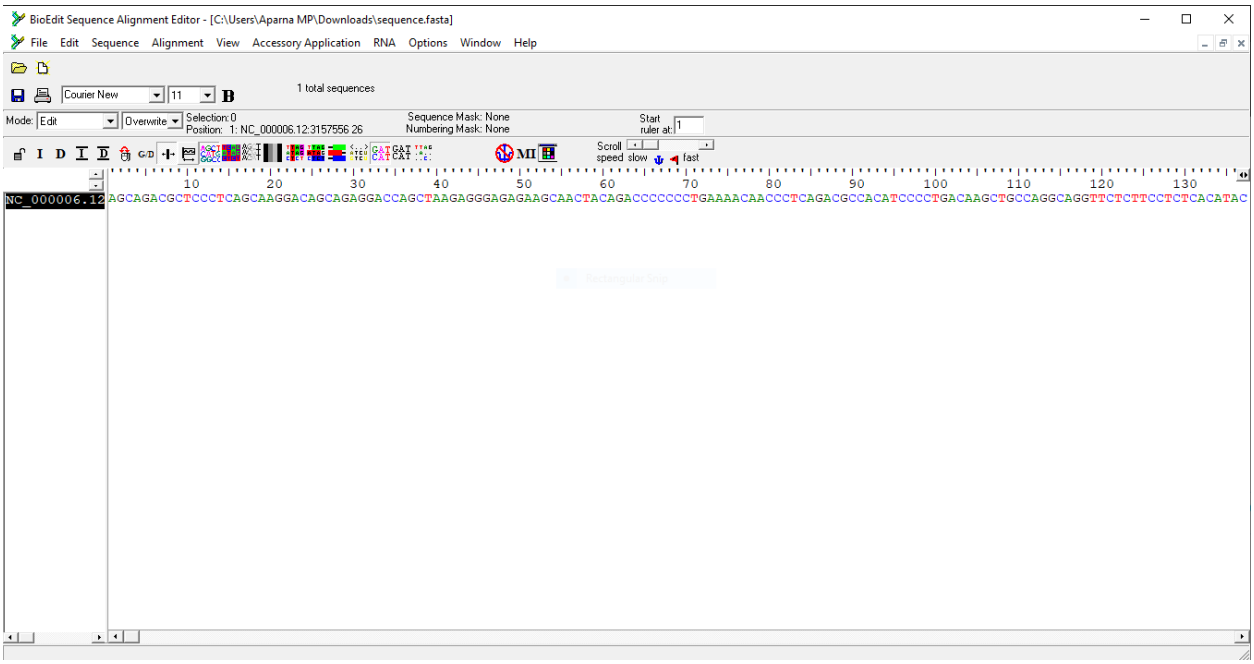


Figure 1: Screenshot of TNF Sequence in BioEdit

Task 2: Generate a Translation of a DNA or RNA Sequence into Amino Acids

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

Steps

- Opened the downloaded TNF gene sequence in BioEdit.
- Used the 'Translate' feature in BioEdit, to generate the amino acid sequence.

Output:

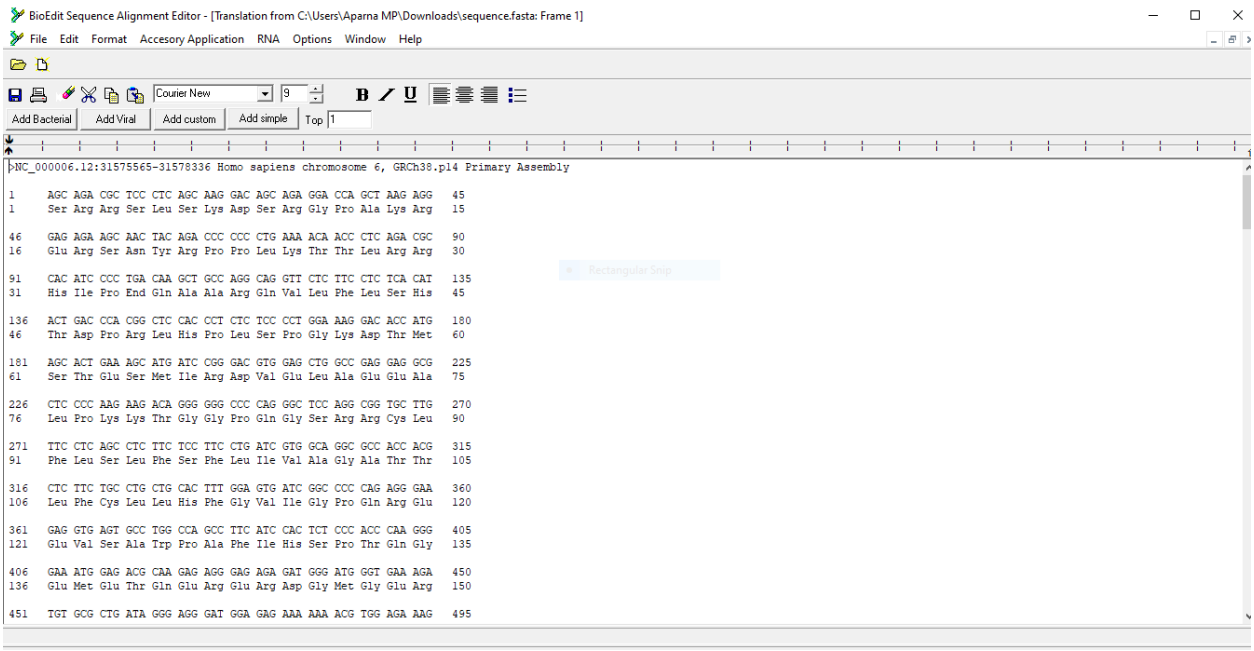


Figure 2: TNF translation in BioEdit

Task 3: Finding ORFs (Open Reading Frames) in a DNA or RNA Sequence

Objective: Identifying the ORFs within the TNF gene sequence

Steps

- Used BioEdit's ORF Finder tool to find ORFs in the TNF gene sequence.
- Recorded the start and stop positions, lengths, and protein translations of the ORFs.

Output

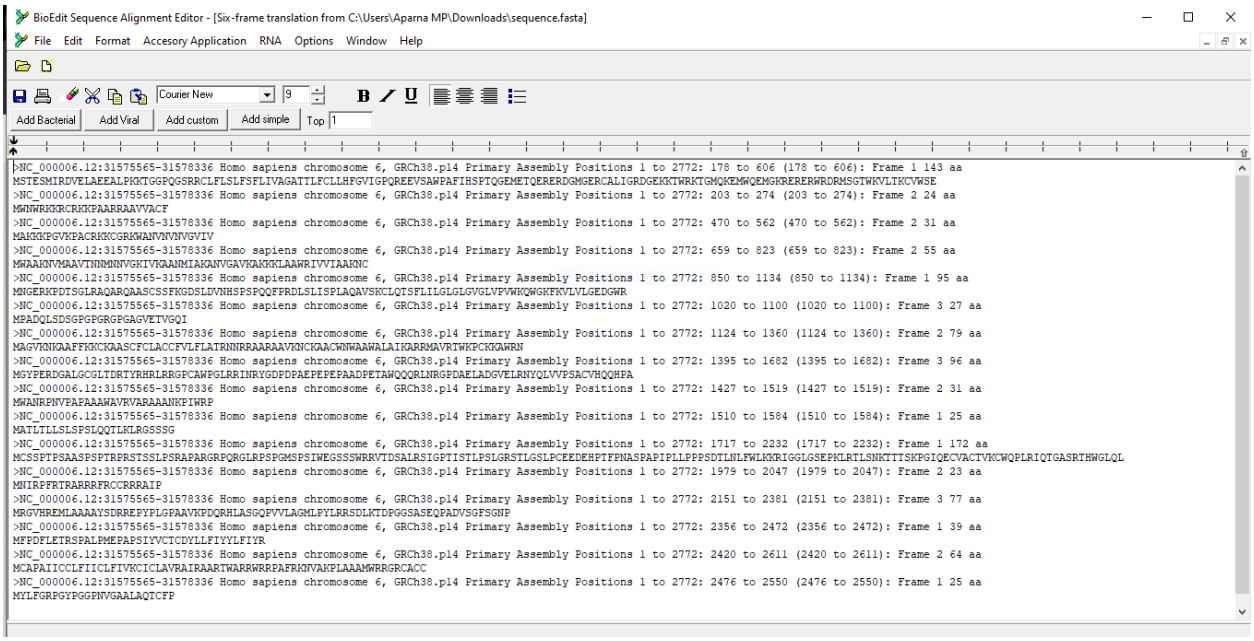


Figure 3: ORFs of TNF in BioEdit

Interpretation of Task 3

- From the ORFs, I selected 4 ORFs which have the highest length. They might be potential candidates for protein coding.
- Performed SMART BLAST analysis for the ORFs and found out 3 protein sequences were selected out from the database;

1. >NC_000006.12:31575565-31578336 TNF [organism=Homo sapiens] [GeneID=7124] [chromosome=6]: 178 to 606: Frame 1 143 aa
2. >NC_000006.12:31575565-31578336 TNF [organism=Homo sapiens] [GeneID=7124] [chromosome=6]: 850 to 1134: Frame 1 95 aa
3. >NC_000006.12:31575565-31578336 TNF [organism=Homo sapiens] [GeneID=7124] [chromosome=6]: 1395 to 1682: Frame 3 96 aa

- The lengthiest ORF was not found to be functional in BLAST analysis, which may suggest that Length of codons doesn't always guarantees protein coding genes.

Task 4: Analyze Sequence Composition (Nucleotide or Amino Acid Frequencies)

Objective: Analyze the nucleotide composition of the TNF gene sequence

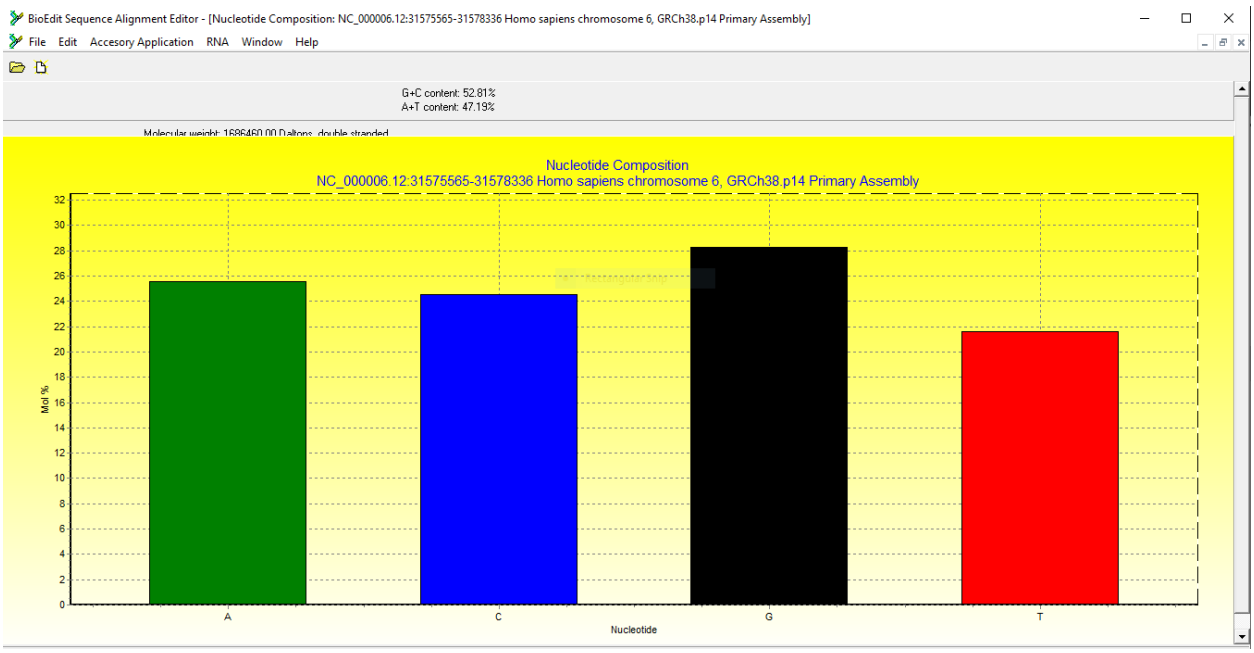
Steps

• Used BioEdit to analyze the sequence composition of the TNF gene.

• Calculated the frequencies of each nucleotide and the overall GC content.

• Interpreted the results and saved the analysis.

Output



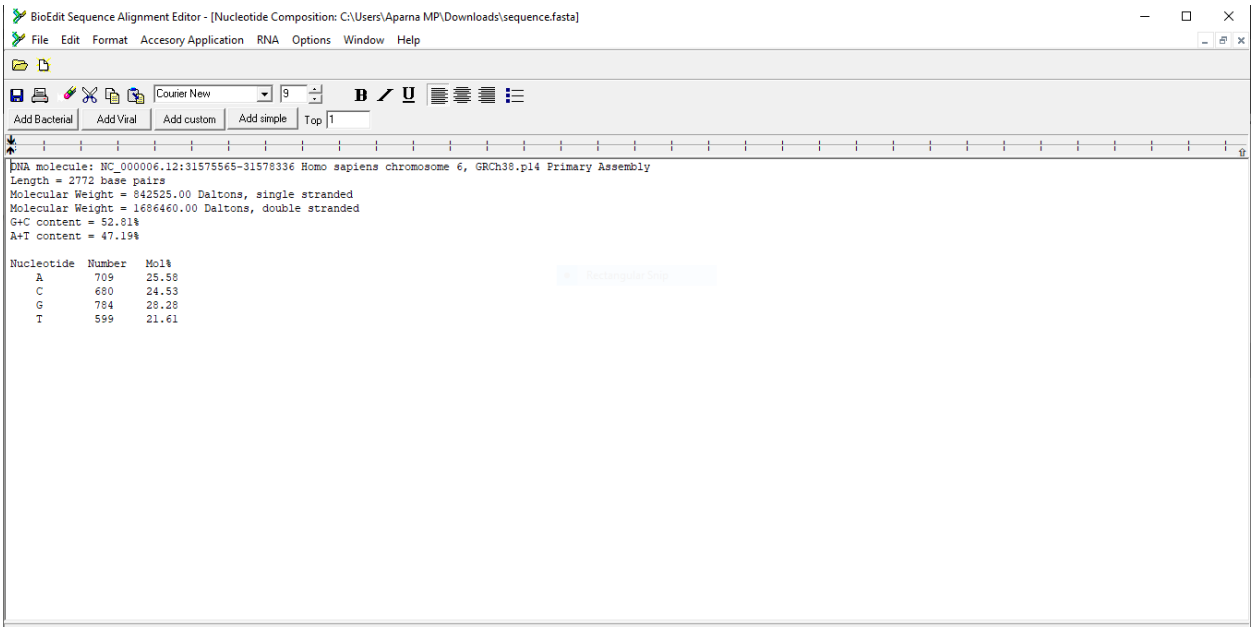


Figure 4: Nucleotide composition of TNF gene (BioEdit)

Interpretation of Task 4

This sequence has higher G+C content. This has advantages as well as disadvantages based on the context and its level. The advantages are genomic stability, protection against exonuclease degradation of mRNA, regulation of gene expression by secondary structures. The disadvantages may include PCR/ cloning challenges due to increased thermal stability, secondary structure causing genomic instability, transcriptional and translational hindrance.

Task 5: Identify Transcription Factor Binding Sites Using the PROMO Tool

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

Steps

- Accessed the PROMO tool at PROMO.
- Selected '*Homo sapiens*' as the species.
- Inputted the promoter region of the TNF gene or use the entire gene sequence.
- Identified potential transcription factor binding sites.

Output

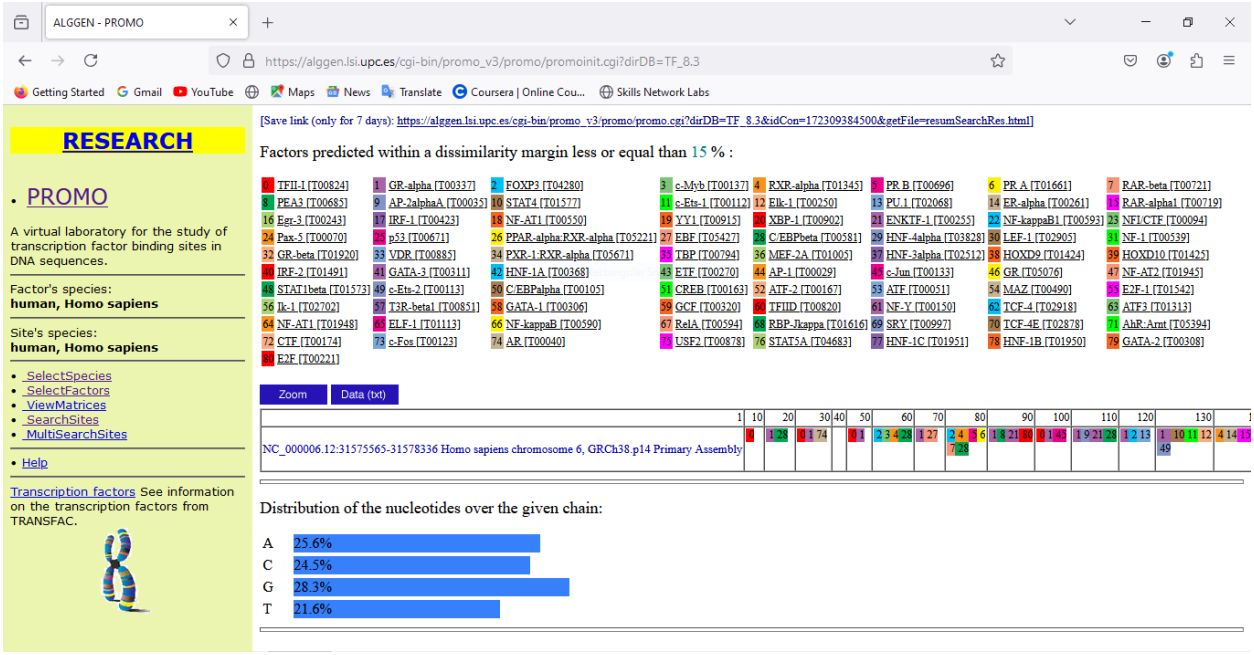


Figure 5: Transcription factor binding sites (PROMO)

Interpretation of Task 5

- Found many transcriptional binding sites along the sequence.

Task 6: Search for Functional Motifs in a Genome or Transcriptome Using MEME Suite

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

Steps

• Accessed the MEME Suite at MEME Suite.

• Uploaded the TNF gene sequence in FASTA format.

• Used the default settings to search for motifs.

• Interpreted and saved the results of the motif search.

Output

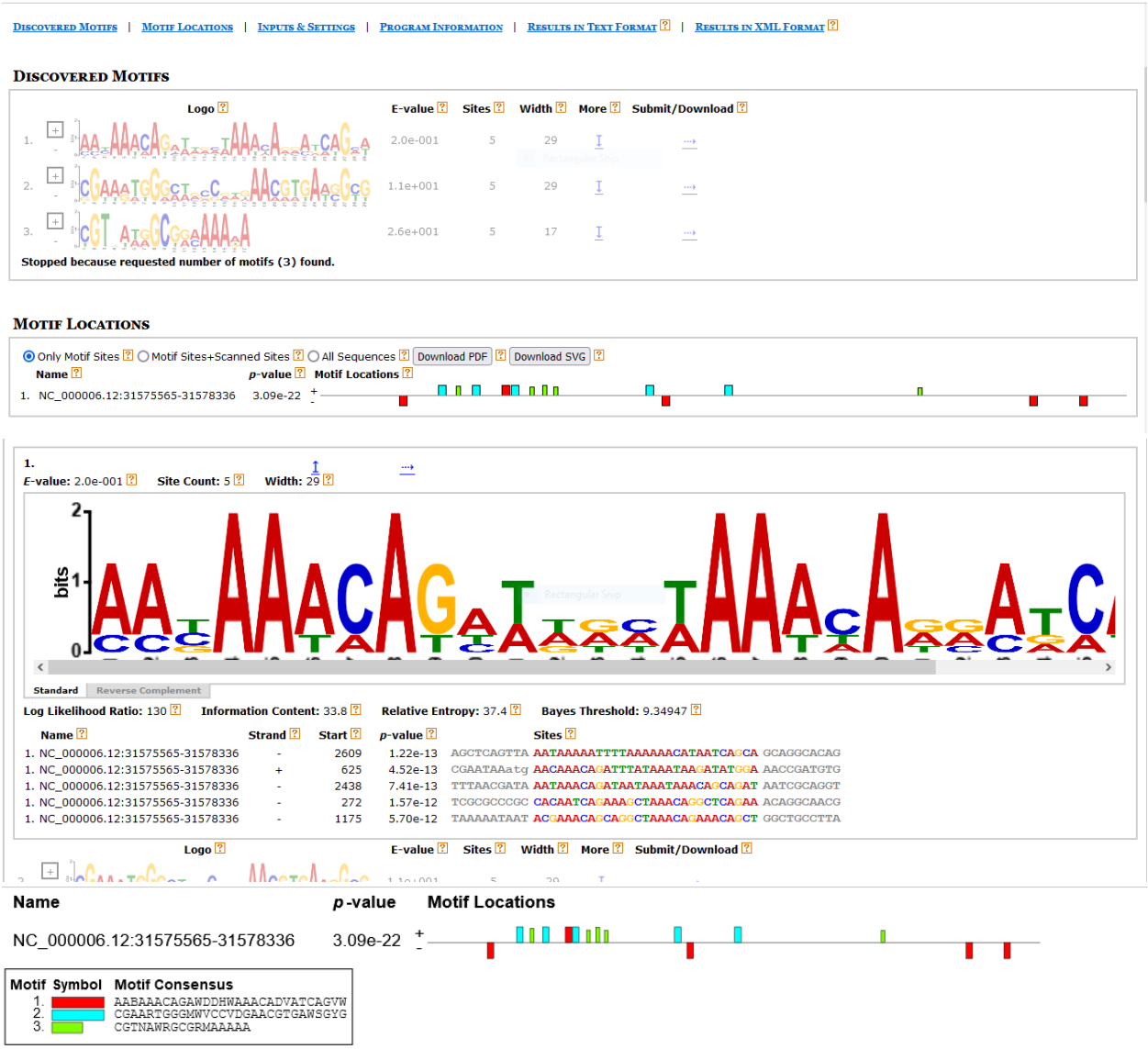


Figure 6: Motifs of TNF (MEME Suite)

Interpretation of Task 6

- MEME suite output provided 3 motifs spanning the sequence.
- The overall low p-value suggests that these motifs were significant.
- The motifs clustering in the region (approximately 433 – 818) suggest higher biological activity in that region.

Task 7: Predict Coding/Non-Coding Regions in a Genome Using GENSCAN

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

Steps

- Accessed the GENSCAN tool.
- Inputted the TNF gene sequence in the appropriate format.
- Ran the analysis to predict coding and non-coding regions.
- Saved and interpreted the results.

Output

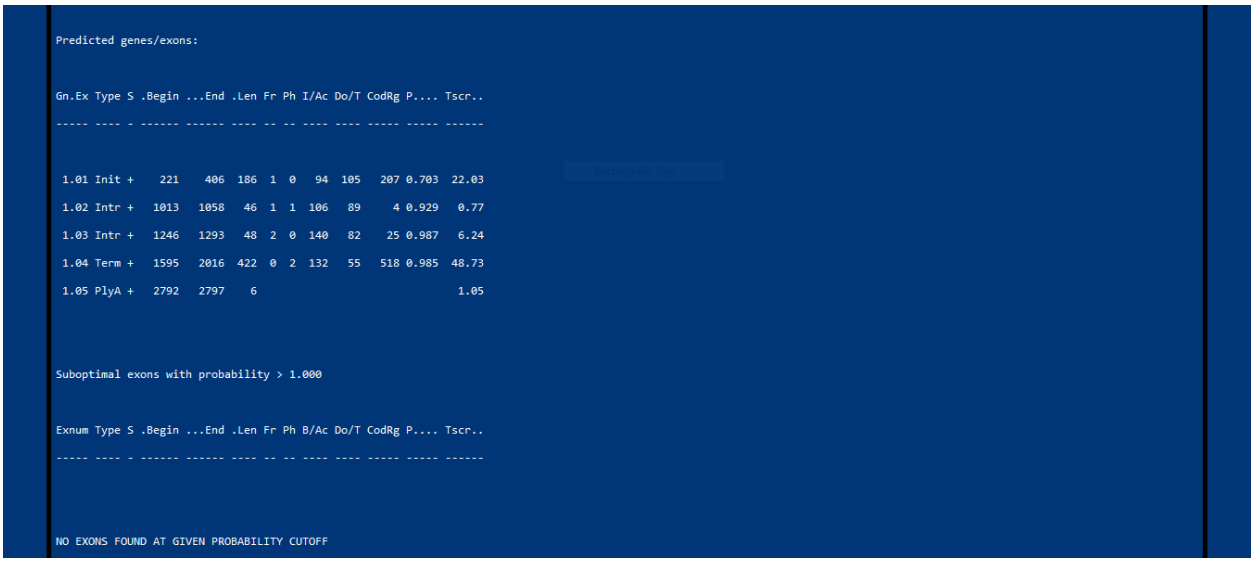


Figure 7 : GENSCAN output of TNF

Interpretation of Task 7

Based on the GENSCAN output, a gene has been predicted with five exons on the positive strand. The first exon is an initial exon starting at position 221 and ending at position 406 with a length of 186 base pairs. There are two internal exons, one spanning position 1013 to 1058 and another from 1246 to 1293. The gene ends with a terminal exon from position 1595 to 2016, followed by a polyadenylation signal at positions 2792 to 2797. All predicted exons have high coding region and transcript scores, indicating strong confidence in the gene predictions.

Task 8: Convert Between Sequence File Formats Using BioEdit (FASTA to PHYLIP)

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

Steps

• Opened the TNF gene sequence in BioEdit.

• Used the 'Save As...' feature to convert the file to PHYLIP format.

• Verified the conversion by opening the PHYLIP file in a text editor.

Discussion

This project provides a comprehensive analysis of the human TNF gene, offering valuable insights into its structure, function, and regulation. The identification of multiple ORFs within the gene, along with the translation of these sequences, suggests potential protein-coding regions that may play a critical role in the TNF gene's biological function. Although the longest ORF was not validated as functional, the presence of other significant ORFs supports the gene's involvement in protein synthesis.

The analysis of sequence composition revealed a higher GC content, which, while contributing to genomic stability, may also pose challenges in experimental procedures like PCR. The identification of transcription factor binding sites within the TNF promoter region underscores the gene's regulatory complexity, which is crucial for its role in immune responses and inflammation.

Furthermore, the discovery of functional motifs within the TNF gene highlights regions of potential biological activity, which could be pivotal in understanding the gene's contribution to disease processes. These motifs, particularly those clustered in specific regions, may serve as targets for future research aimed at uncovering the gene's involvement in autoimmune disorders or other inflammatory conditions.

In conclusion, this project not only enhances our understanding of the TNF gene but also opens avenues for further investigation into its role in health and disease. Future studies could focus on experimental validation of the predicted motifs and transcription factor

binding sites, as well as exploring the gene's variants in different populations to better understand its role in disease susceptibility.

Reference

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology, 215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2. NCBI BLAST
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98. BioEdit
- Farré, D., Roset, R., Huerta, M., Adsuara, J. E., Roselló, L., Albà, M. M., & Messeguer, X. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Research, 31*(13), 3651-3653. doi:10.1093/nar/gkg638. PROMO
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... & Noble, W. S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research, 37*(suppl_2), W202-W208. doi:10.1093/nar/gkp335. MEME Suite
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology, 268*(1), 78-94. doi:10.1006/jmbi.1997.0951. GENSCAN