**BIOINFORMATICS MINI PROJECT: Comprehensive Sequence Analysis of the Human TNF Gene**

**TASK 1:** Download the human TNF gene sequence from NCBI and view/edit it

**Tool(s) used:** NCBI and BioEdit

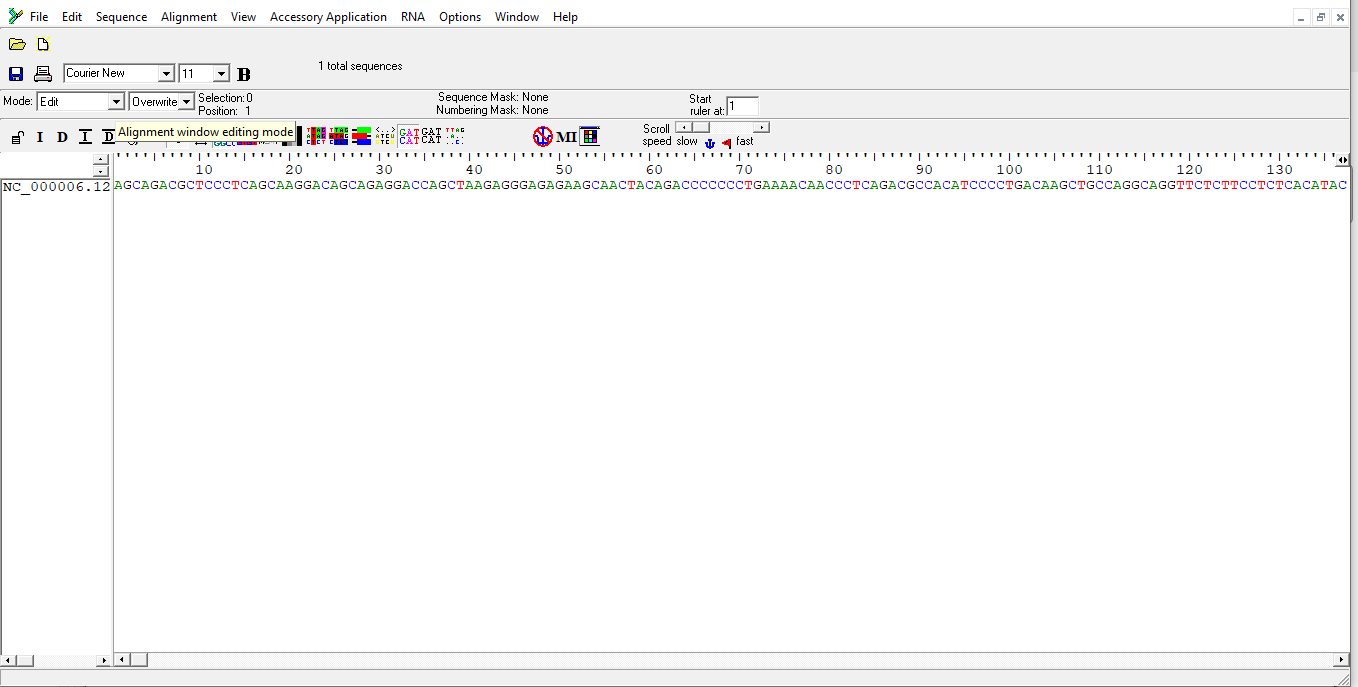
**Output:**

Figure 1: Human TNF gene sequence on BioEdit

**TASK 2**: Translate the DNA sequence of the TNF gene into an amino acid sequence

**Tool(s) used**: BioEdit

**Output:**

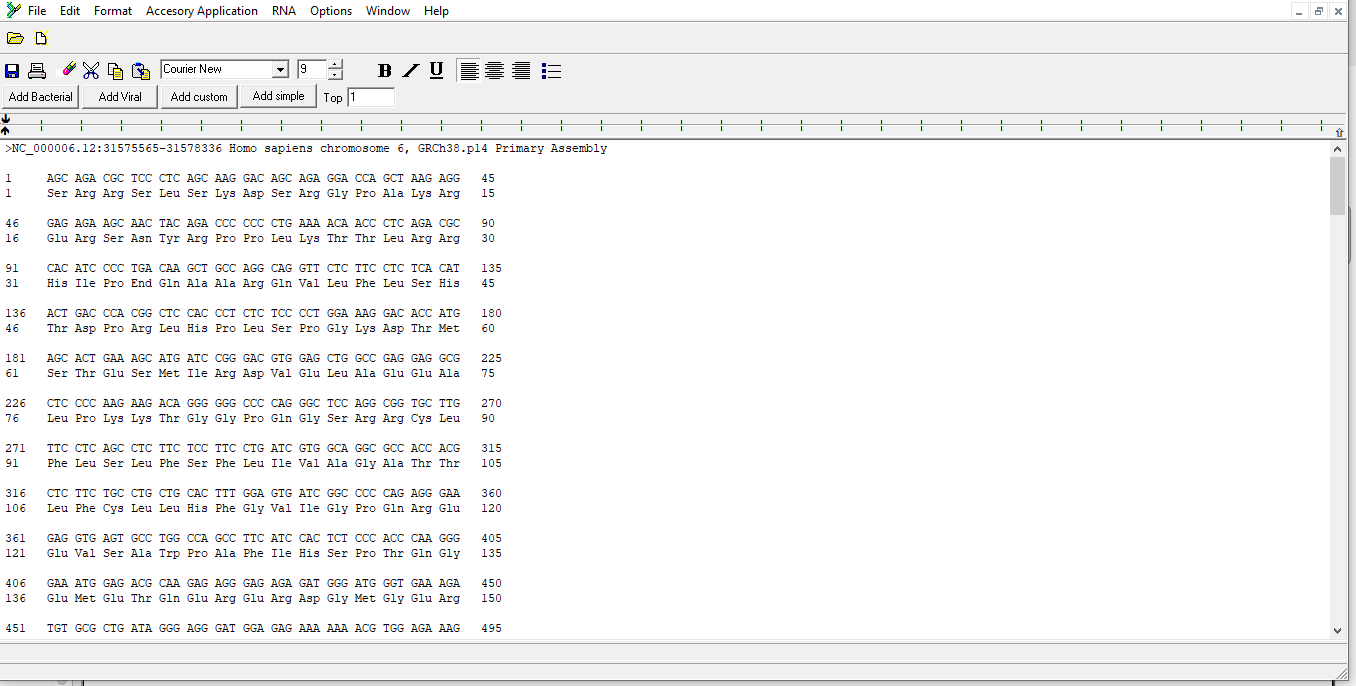
****

Figure 2: Translation of the human TNF gene sequence into amino acid sequence

**TASK 3:** Identify the ORFs (Open Reading Frames) within the TNF gene sequence

**Tool(s) used:** BioEdit

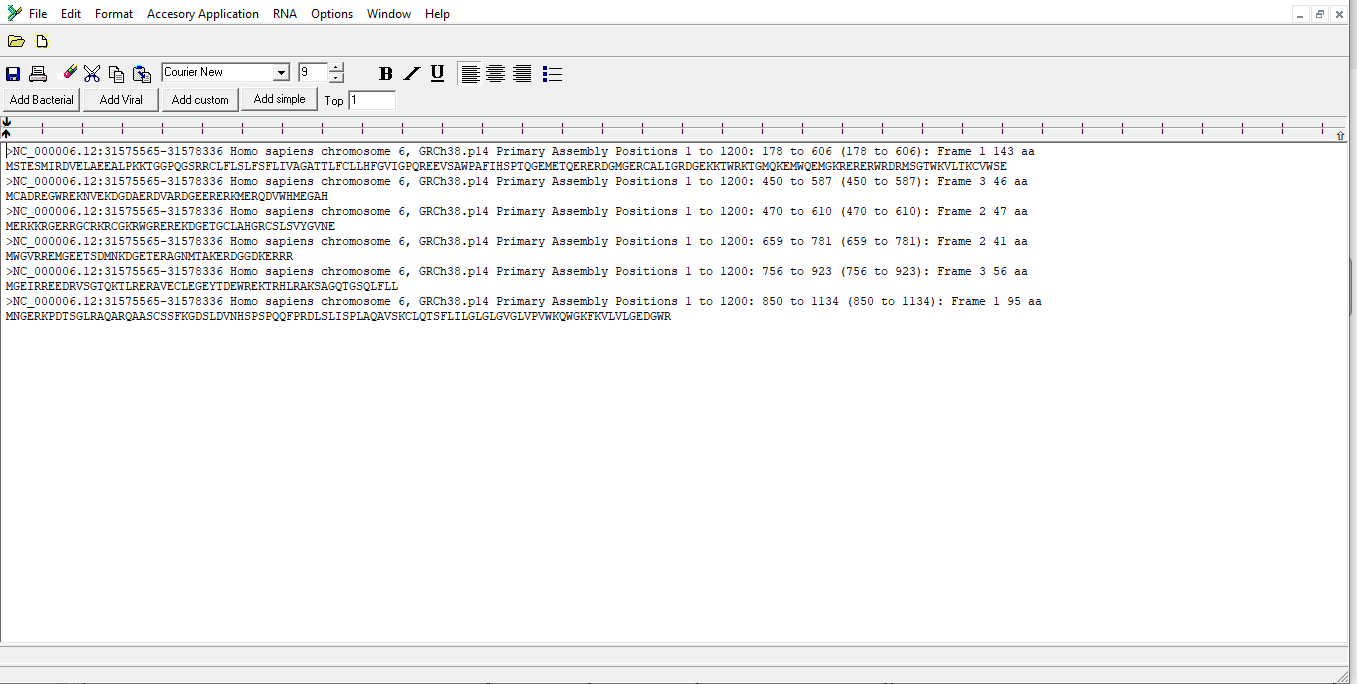
**Output:**

Figure 3: Identification of ORFs in the TNF gene sequence

**Result interpretation:**

- Figure 3 displays 6 open reading frames (ORFs) ranging from positions 1 to 1200.

- Each ORF is annotated with its start and stop positions, frame designation, and corresponding protein translations.

- The first ORF spans positions 178 to 606, producing a protein of 143 amino acids. Below, the amino acid sequence translation for that specific ORF is shown.

- Identifying ORFs is crucial because it helps researchers predict the locations of genes within a DNA sequence, including those that may play roles in biological processes and disease development.

**Task 4:** Analyze the nucleotide composition of the TNF gene sequence

**Tool(s) used:** BioEdit

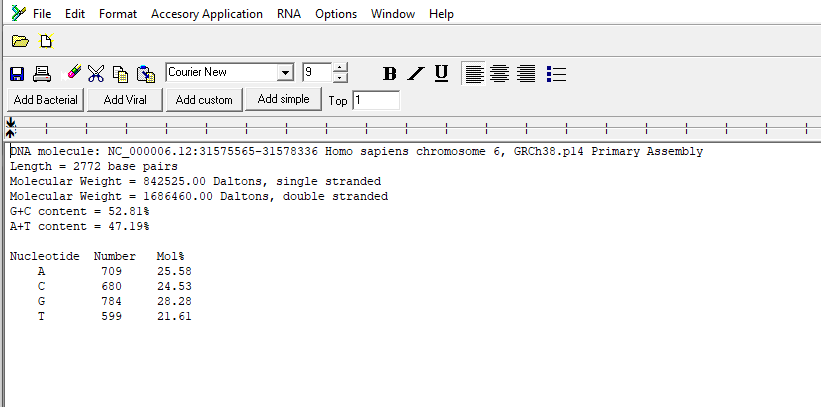
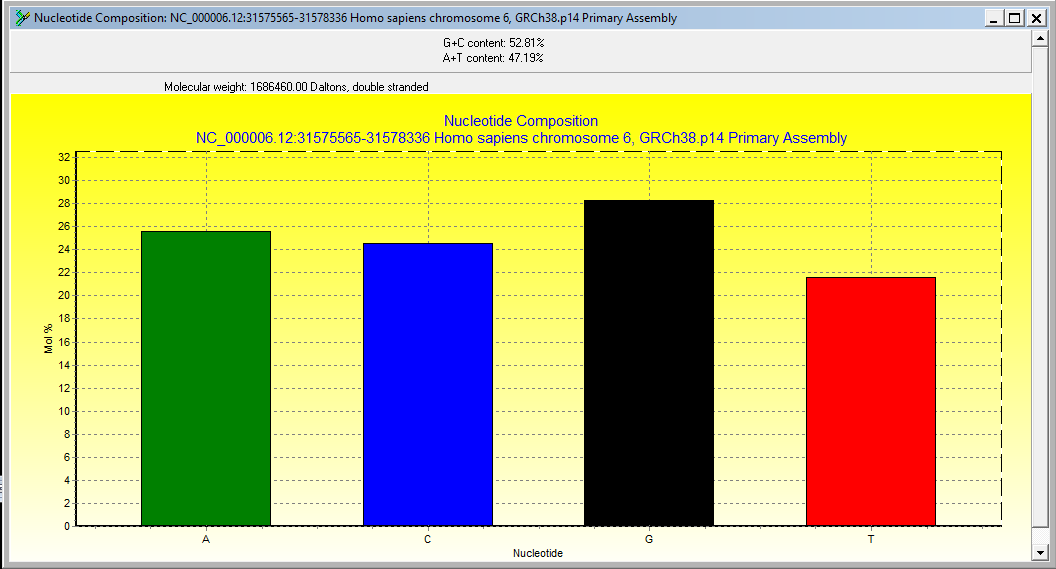
**Output:**

Figure 4: Analysis of nucleotide composition in the TNF gene sequence

Figure 5: Analysis of nucleotide composition in the TNF gene sequence (Bar graph)

**Result interpretation:**

- Figures 4 and 5 illustrate the nucleotide frequency, molecular weight, and GC/AT content of the TNF gene sequence.

- The observed nucleotide counts are 709 adenine (A), 680 cytosine (C), 784 guanine (G), and 599 thymine (T).

- The sequence has an overall GC content of 52.81%.

- This value is particularly relevant because GC content influences the thermal stability of DNA, with higher GC levels generally conferring greater stability.

**Task 5:** Identify potential transcription factor binding sites in the TNF gene promoter region **Tool(s) used:** TFBSPred

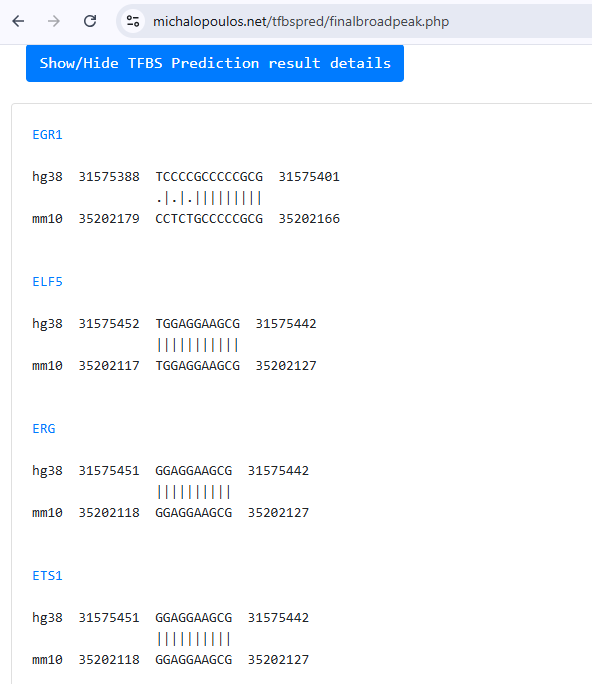
**Output:**

Figure 6: Identification of potential transcription factor binding sites in the TNF gene sequence

**Result interpretation:**

- Figure 6 shows the predicted transcription factor binding sites in the TNF gene sequence using TFBSPred. Binding sites were found for factors such as EGR1, ELF5, ERG, and ETS1, with additional factors like PATZ1 - although not included in this figure.

- Each of these sites could contribute to the regulation of TNF gene expression by allowing specific transcription factors to bind and control activity. These predictions highlight possible regulatory elements that may influence TNF function in health and disease.

**Task 6:** Search for functional motifs in the TNF gene sequence using MEME Suite

**Tool(s) used:** MEME Suite

**Output:**

Figure 7: Identification of functional motifs in the TNF gene sequence

**Result interpretation:**  
- Figure 7 shows three discovered motifs, along with their E-values, sites, lengths, and positions in the TNF gene sequence.

- The E-value indicates how statistically significant a motif is. Lower E-values indicate greater significance, while values above 0.05 are usually considered not significant and are shown in grey by MEME.

- In this result, all three motifs have E-values above 0.05, so they are non-significant and displayed in grey.

**Task 7:** Predict the coding and non-coding regions within the TNF gene sequence

**Tool(s) used:** GENSCAN

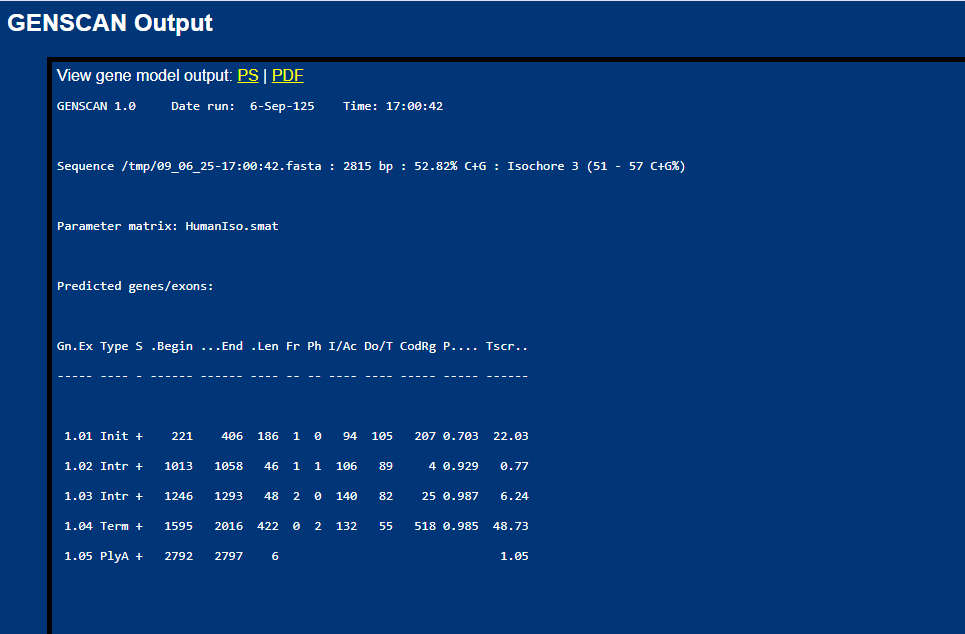
**Output:**

Figure 8: Identification of coding and non-coding regions within the TNF gene sequence

**Result interpretation:**  
- The output in figure 8 shows details for each predicted gene/exon, including sequence length, reading frame, type, strand orientation, and several specific scores (I/AC, Do/T, etc).

- In total, four exons were predicted: two internal (Intr), one initial (Init), one terminal (Term), followed by one predicted polyA site (PlyA). All of these are located on the positive (+) strand, as indicated in the third column (S); and most show high probability values.

- It should be noted that predicted exons with higher probabilities are generally more reliable than those with lower ones.

**Task 8:** Convert the TNF gene sequence from FASTA format to PHYLIP format

**Tool(s) used:** BioEdit

**Output:** Included within the upload