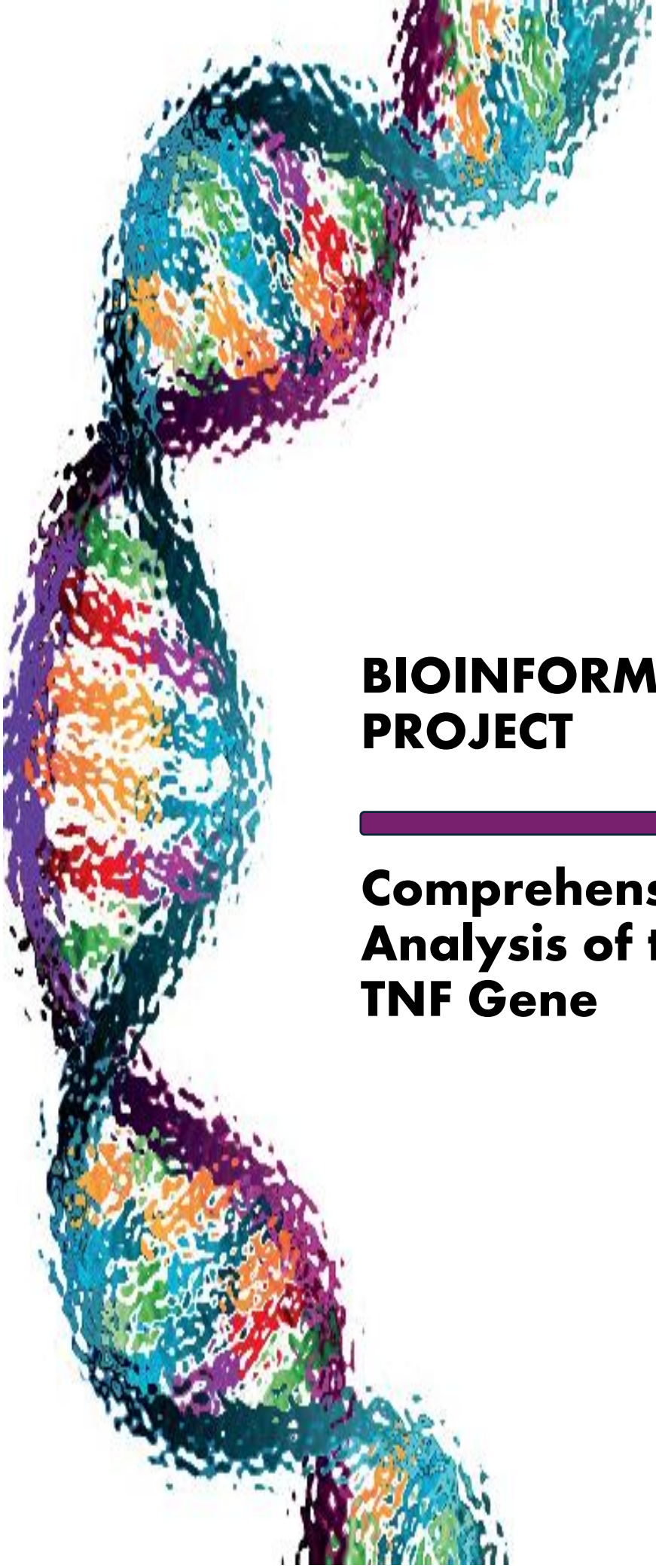

BioinformHER

BIOINFORMATICS MINI PROJECT

Comprehensive Sequence Analysis of the Human TNF Gene

BY: ONYIDO CHIOMA

DATE: 09-08-2024

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

Tool(s) used: NCBI and BioEdit

Output:

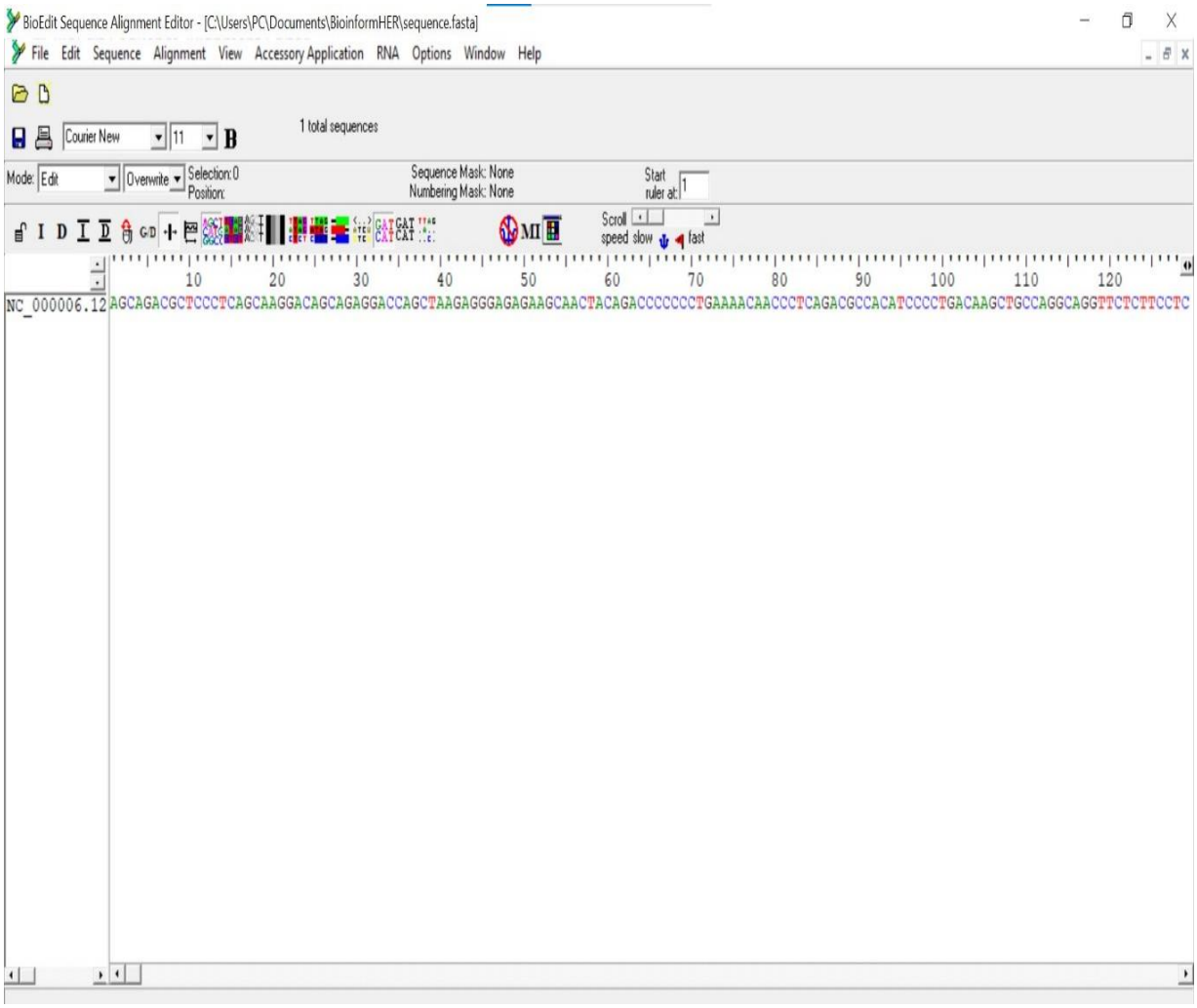

Figure 1: Human TNF gene sequence on BioEdit

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

Tool(s) used: BioEdit

Output:

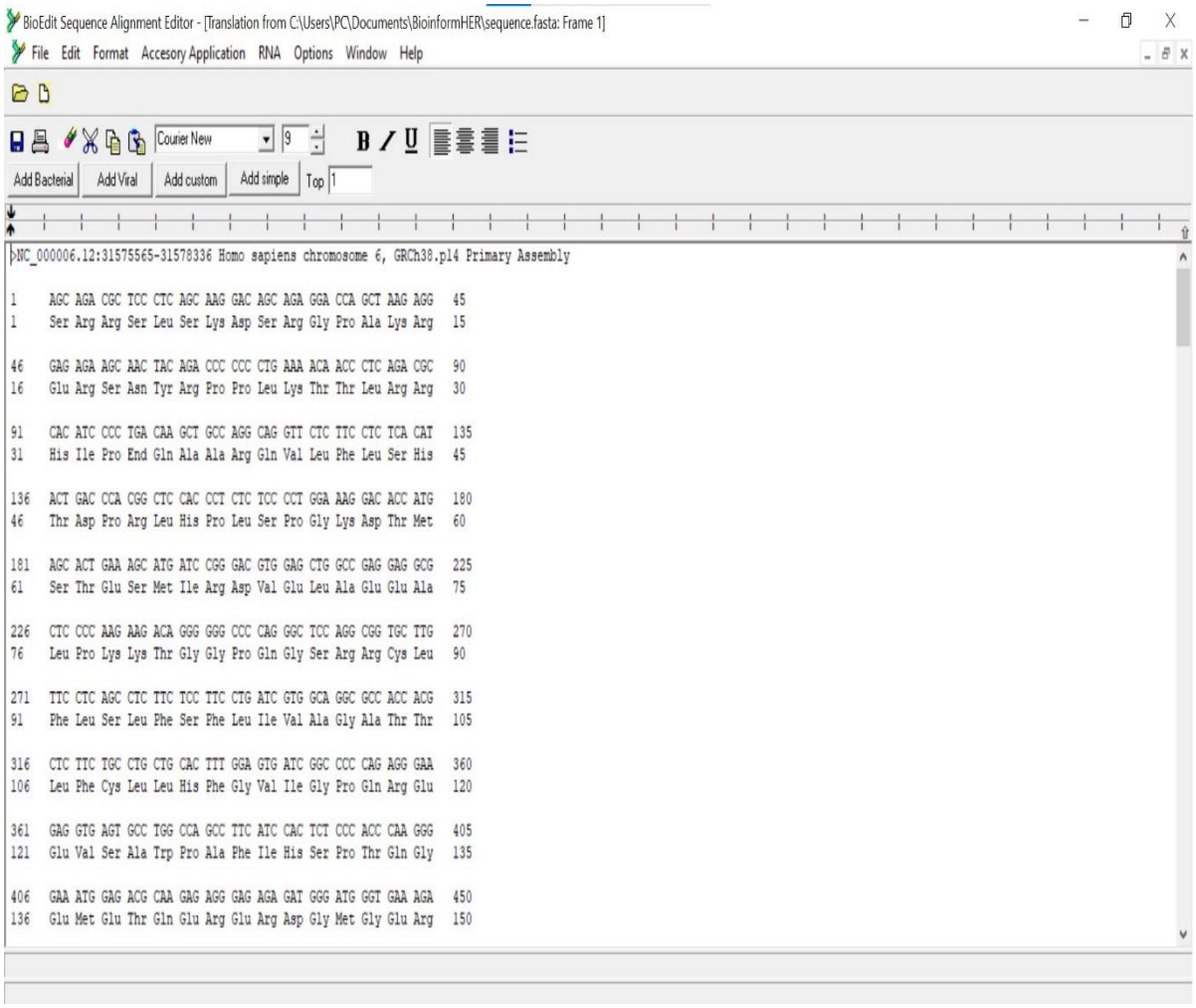

Figure 2: Translation of the human TNF gene sequence into corresponding amino acid sequence using BioEdit

TASK 3: Identify the ORFs within the TNF gene sequence

Tool(s) used: BioEdit

Output:

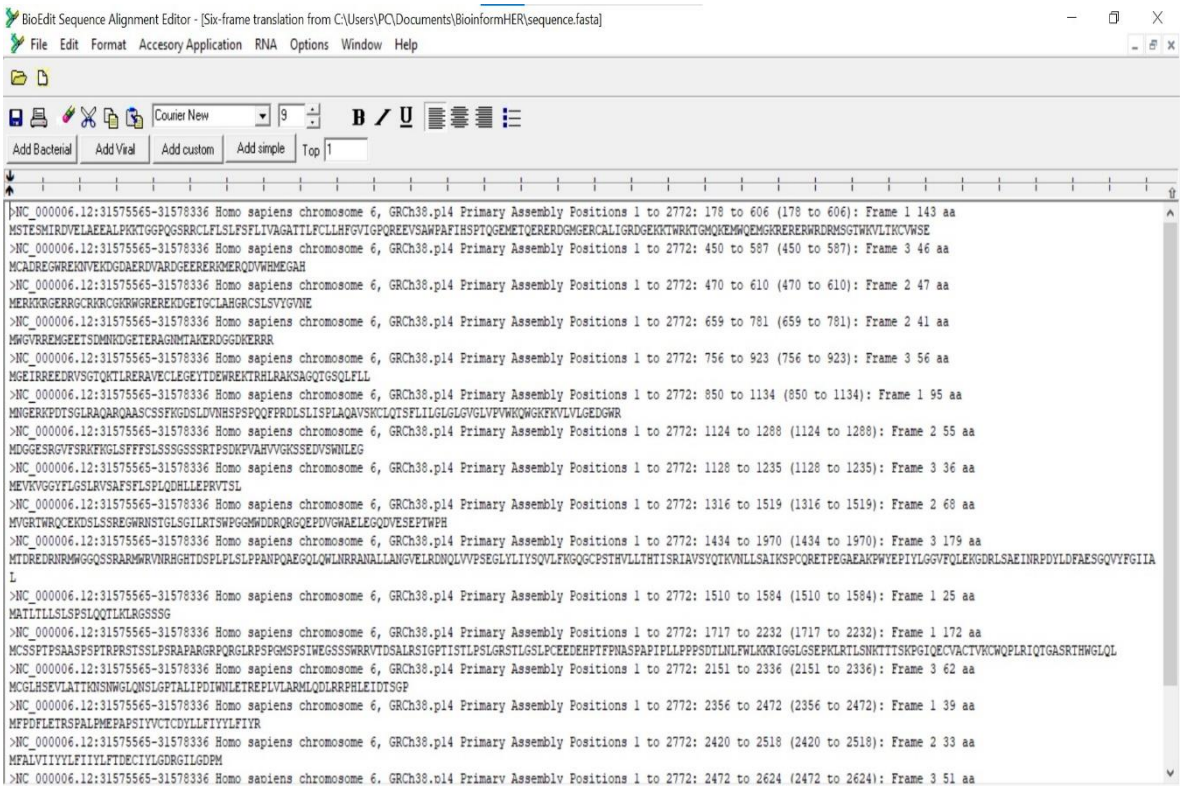

Figure 3: Identification of ORFs in the TNF gene sequence using BioEdit

Result interpretation:

Figure 3 displays 17 open reading frames (ORFs) ranging from positions 1 to 2772.

Each ORFs are annotated with its start and stop positions, lengths, and corresponding protein translations.

For example, in this case, the first ORF spans positions 178 to 606, resulting in a protein of 143 amino acids.

Just below that, you'll find the amino acid sequence translation for this specific ORF.

Identifying ORFs is important because it allows researchers to predict where genes (that may be involved in tumor development, for instance) are located within a DNA sequence.

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

Tool(s) used: BioEdit

Output:

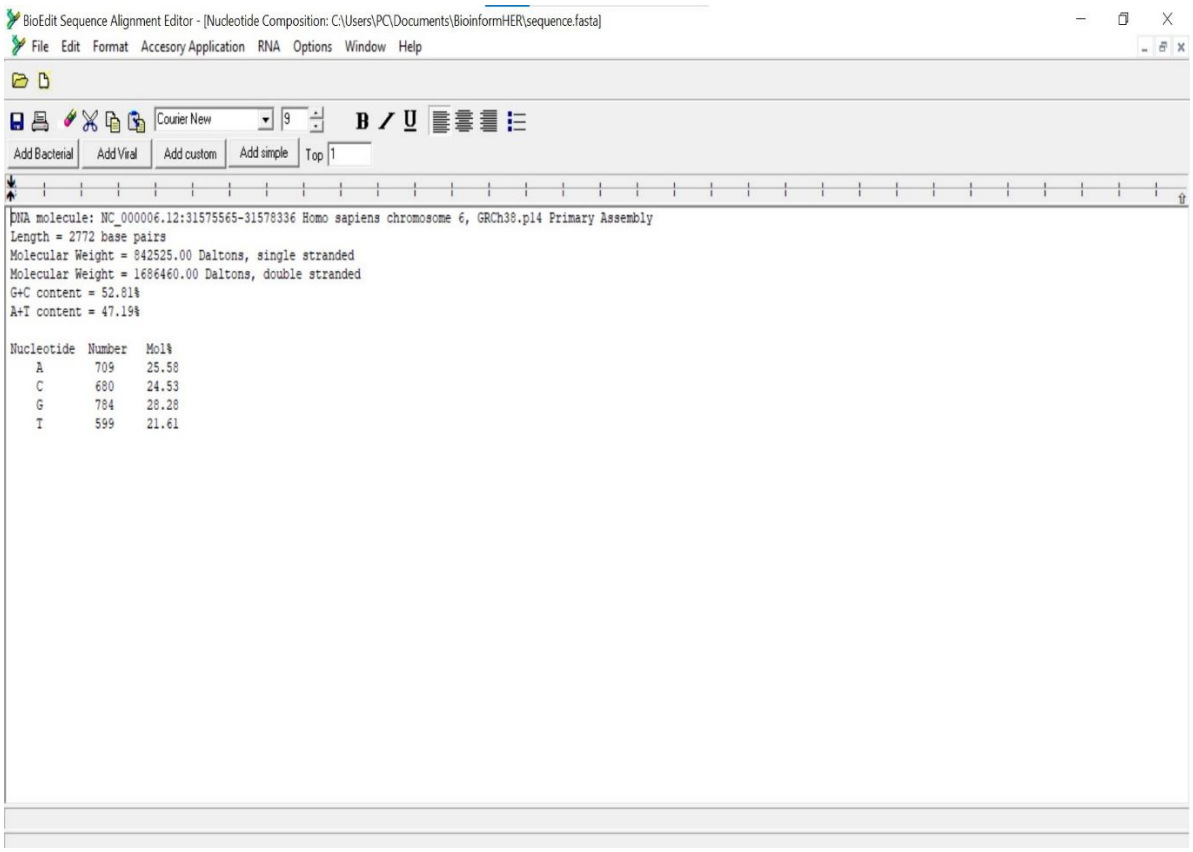

Figure 4: Analysis of nucleotide composition in the TNF gene sequence using BioEdit

Figure 5: Nucleotide composition bar graph of the TNF gene sequence

Result interpretation:

Figures 4 and 5 display the frequency, molecular weight, GC/AT content of each nucleotide in the TNF gene sequence.

The frequencies of adenine (A), cytosine (C), guanine (G) and thymine (T) are 709, 680, 784 and 599, respectively.

The overall GC content of the sequence is 52.81%.

Knowing the GC content is important for assessing the thermal stability of the sequence, as higher GC content typically means greater stability.

Task 5: Identify potential transcription factor binding sites in the TNF gene promoter region

Tool(s) used: PROMO

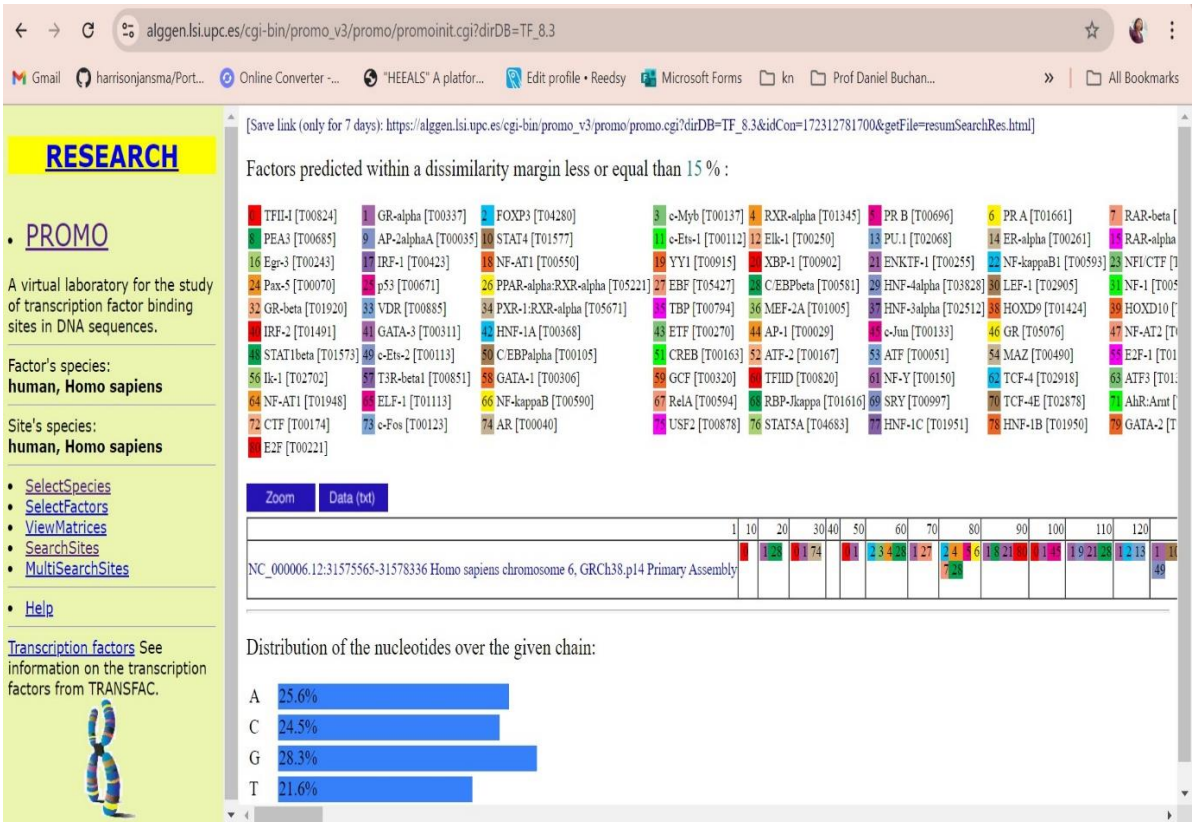

Figure 6: Identification of potential transcription factor binding sites in the TNF gene sequence using PROMO

Result interpretation:

Figure 6 shows a list of 80 potential transcription factor binding sites identified in the TNF gene sequence using PROMO.

The data shows which transcription factors can potentially bind to our sequence.

Each identified binding site could play a role in controlling TNF gene expression.
For example, after looking up the first predicted factor (TFII-I [T00824]) in the TRANSFAC database, I found that TFII-I promotes the formation of Phox1-dependent signal-responsive complexes on the serum response element (SRE).

Overall, this information gives insights into how gene regulation could contribute to diseases and potentially be used for developing therapeutic strategies.

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

Tool(s) used: MEME Suite

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

Result interpretation:

Figure 7 provides details of 3 discovered motifs, including the E-value (representing the statistical significance of each motif), along with the sites, lengths, and locations within the input sequence. The lower the E-value, the more significant the motif. Typically, E-values larger than 0.05 are considered non-significant and are usually greyed out in MEME. For example, in this case, all the motifs have E-values above 0.05 and are thus greyed out.

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

Tool(s) used: GENSCAN

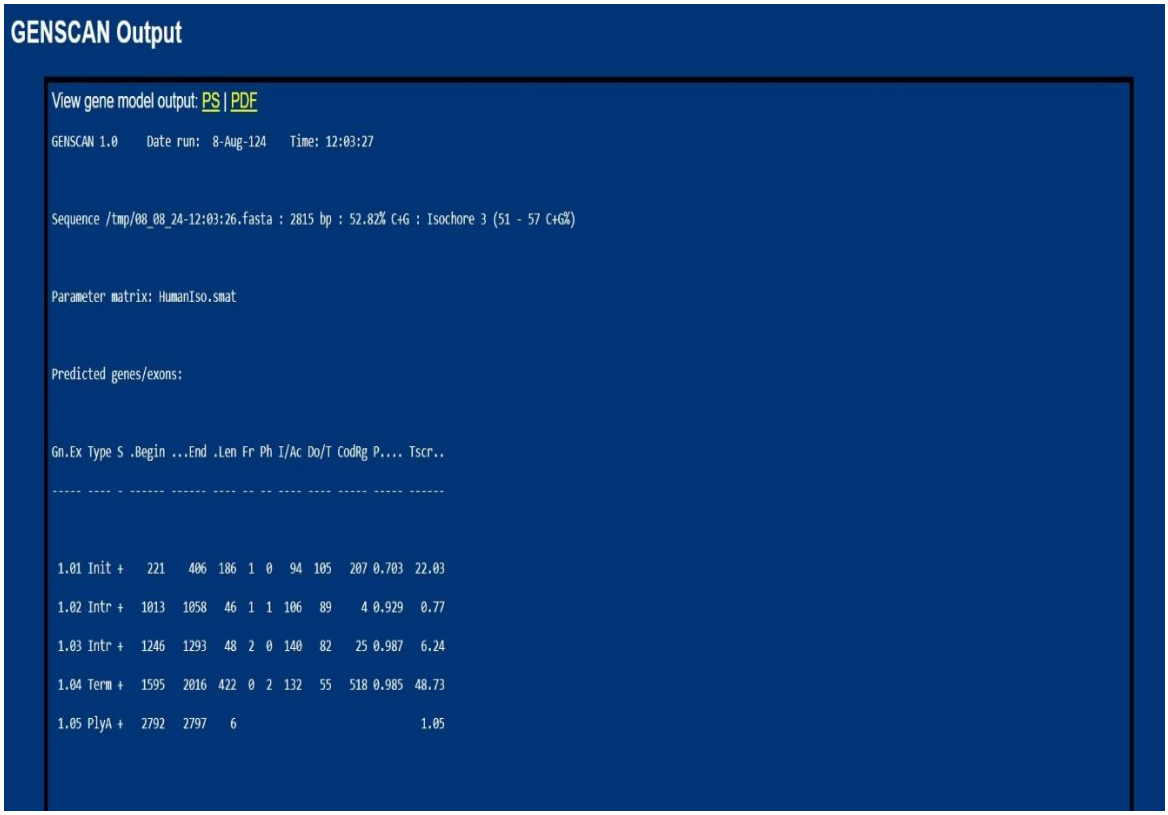

Figure 8: Identification of coding and non-coding regions within the TNF gene sequence using GENSCAN

Result interpretation:

This output provides information about each predicted gene/ exon, including (but not limited to) the length of the sequence, frame, GC content, type, S, and specific scores such as I/AC, Do/T, CodRg, P, and Tscr.

There are 5 predicted gene/ exon fragments, which include 2 internal (Intr), 1 initial (Init), 1 terminal (Term) and 1 polyA tail.

All of these are located on positive strands (as shown in the 3rd column), and most of them have high probabilities.

It's important to note that the predicted exons with higher probabilities are generally more accurate than those with lower probabilities.

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

Answer: Attached