**Astrid Liliana Vargas Sánchez**

**BioinformHER Mini Project: Analysis of the Human TNF Gene**

This repository contains the development of the BioinformHER mini project, in which the human **TNF** gene is analyzed. The proposed tasks and their results are described below:

**Task 1. Downloading a Biological Sequence from NCBI and View/ Edit it**

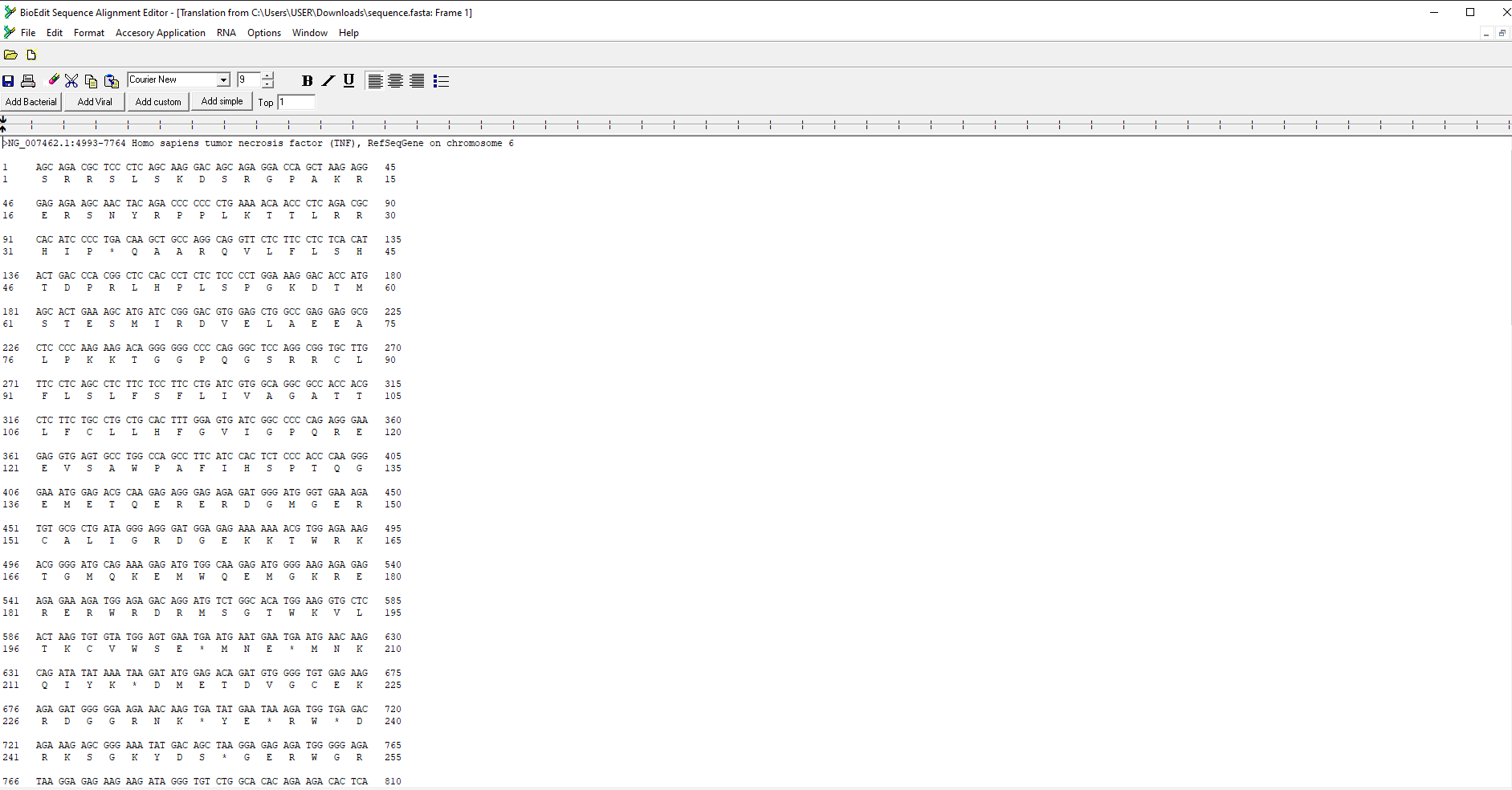
The TNF gene sequence was downloaded and visualized in BioEdit. The sequence can be seen in the following image:



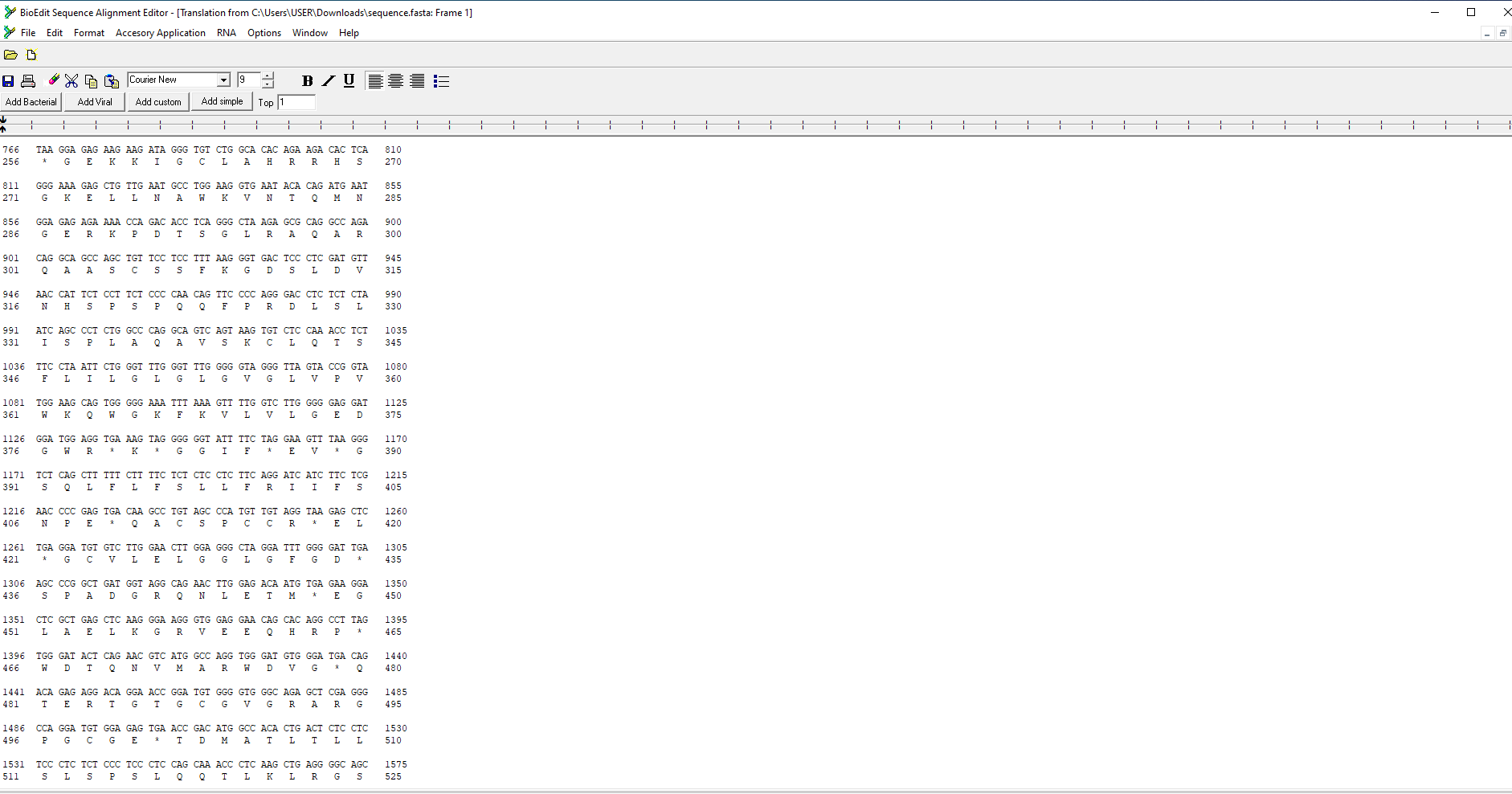
<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%201.png>

**Task 2. Generate a Translation of a DNA or RNA Sequence intro Amino Acids**

The obtained amino acid sequence is presented in the following images:



<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%202-1.png>



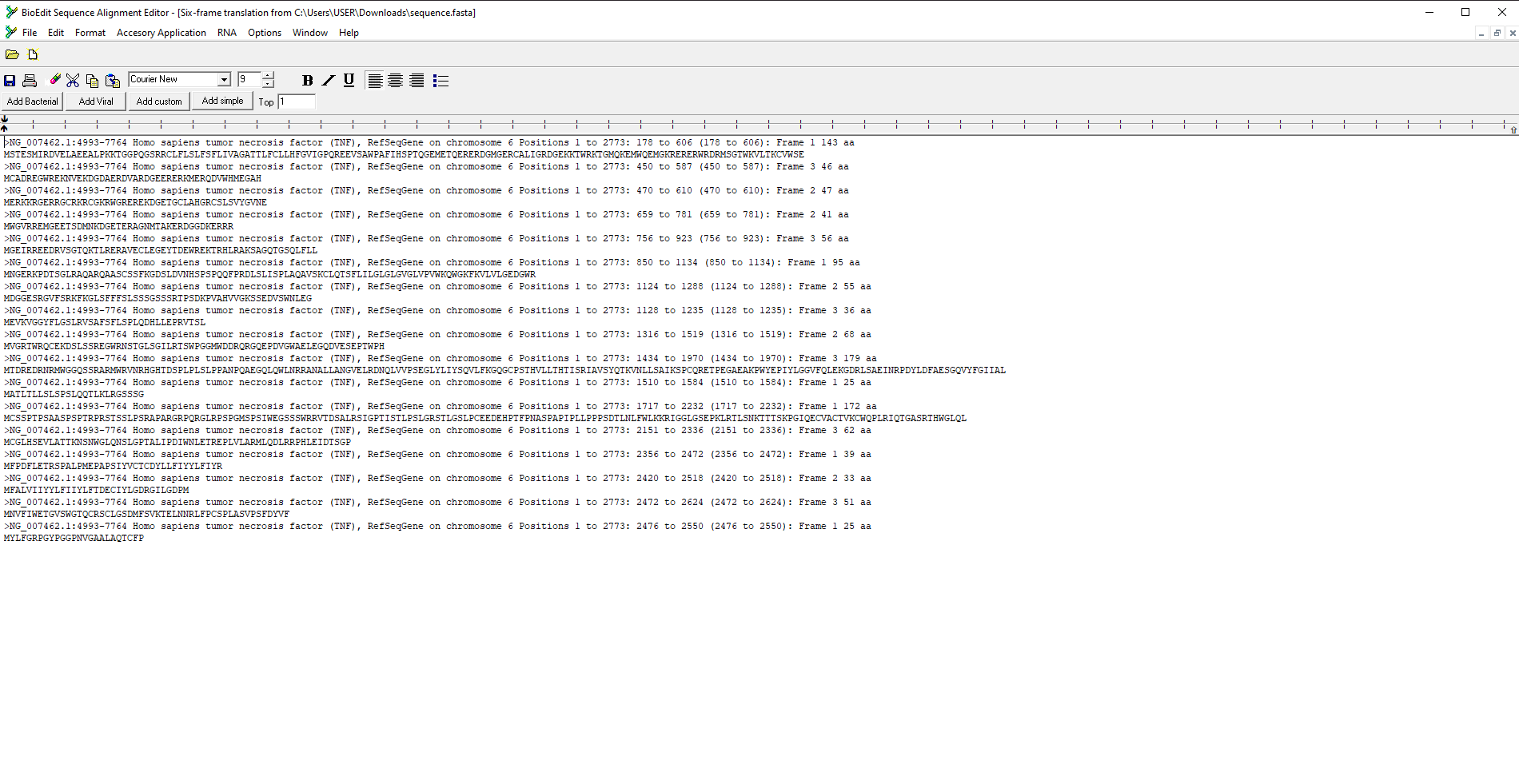
<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%202-2.png>



<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%202-3.png>

**Task 3. Find ORFs (Open Trading Frames) in a DNA or RNA Sequence**

ORFs are regions that begin with a start codon (usually ATG) and extend to the first stop codon. BioEdit identified 17 possible ORFs at different positions and with various lengths. For example, the first ORF is located between positions 178 and 606, and has a length of 143 amino acids.

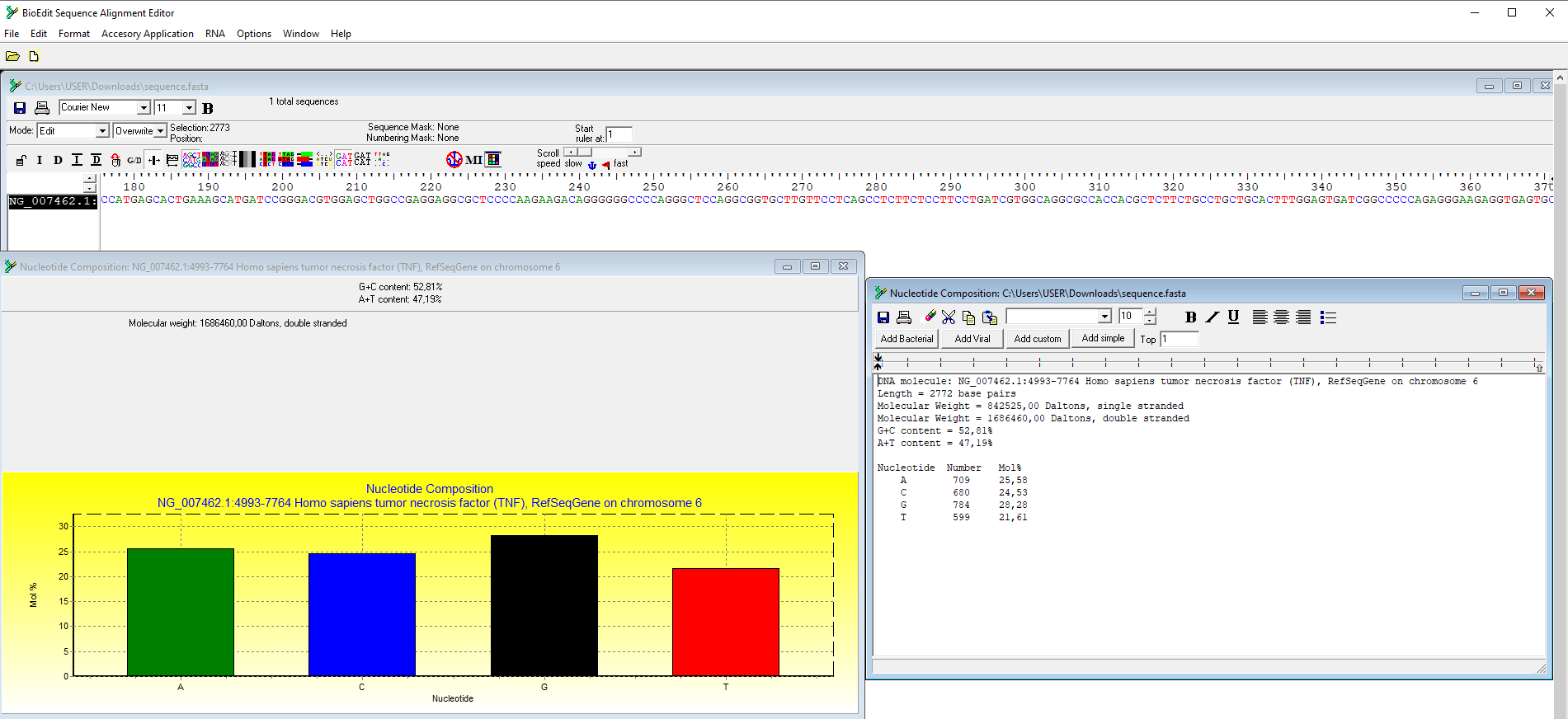


<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%203.png>

<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%203.ORFs>

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

The nucleotide composition of the sequence was analyzed. The **Homo sapiens TNF** gene is located on chromosome 6 and has a length of 2772 base pairs. The sequence has a G+C content of 52.81% and an A+T content of 47.19%. The most prevalent nucleotide is Adenine (25.58%), while the least prevalent is Thymine (21.61%). The composition details can be seen in the following image:

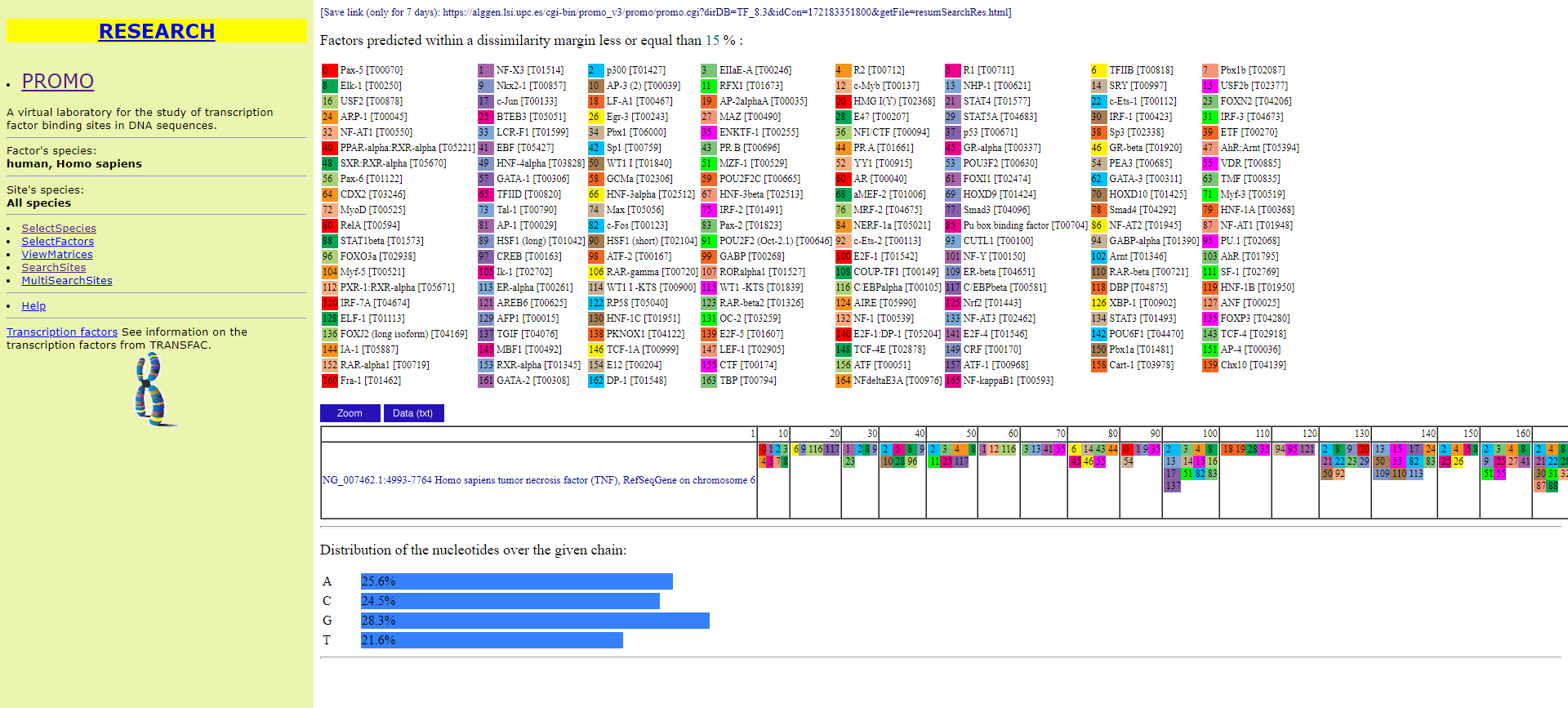


<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%204.Nucleotide_composition>

<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%204.png>

**Task 5. Identify Transciption Factor Binding Sites Using the PROMO Tool**

The image shows a list of transcription factors that can bind to the sequence.

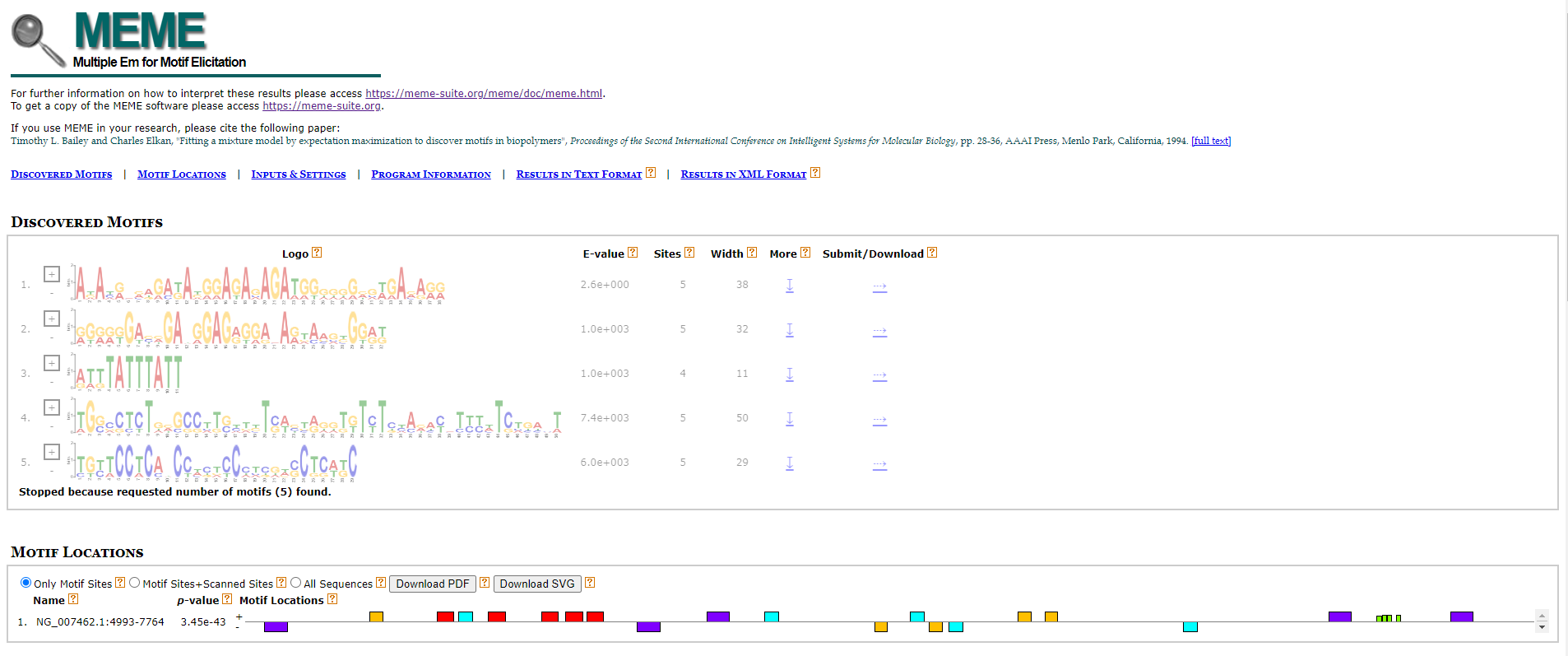


<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%205.png>

For each identified transcription factor, details such as the position of the binding site on the sequence and the name of the transcription factor are provided. For example, the first factor on the list is Pax-5 [T00070], with a binding site at positions 10 and 90. A summary of the nucleotide distribution in the sequence is also included.

**Task 6. Search for Funcional Motifs in a Genome or Transcriptome Using MEME**

The MEME tool identified several functional motifs in the sequence. Functional motifs are specific patterns of nucleotides or amino acids with a biological function. Five functional motifs were identified, with larger nucleotides being more predominant. The associated E-value is significant, indicating that the motifs are unlikely to have been found by chance. The positions and sizes of the motifs are detailed in the following images:



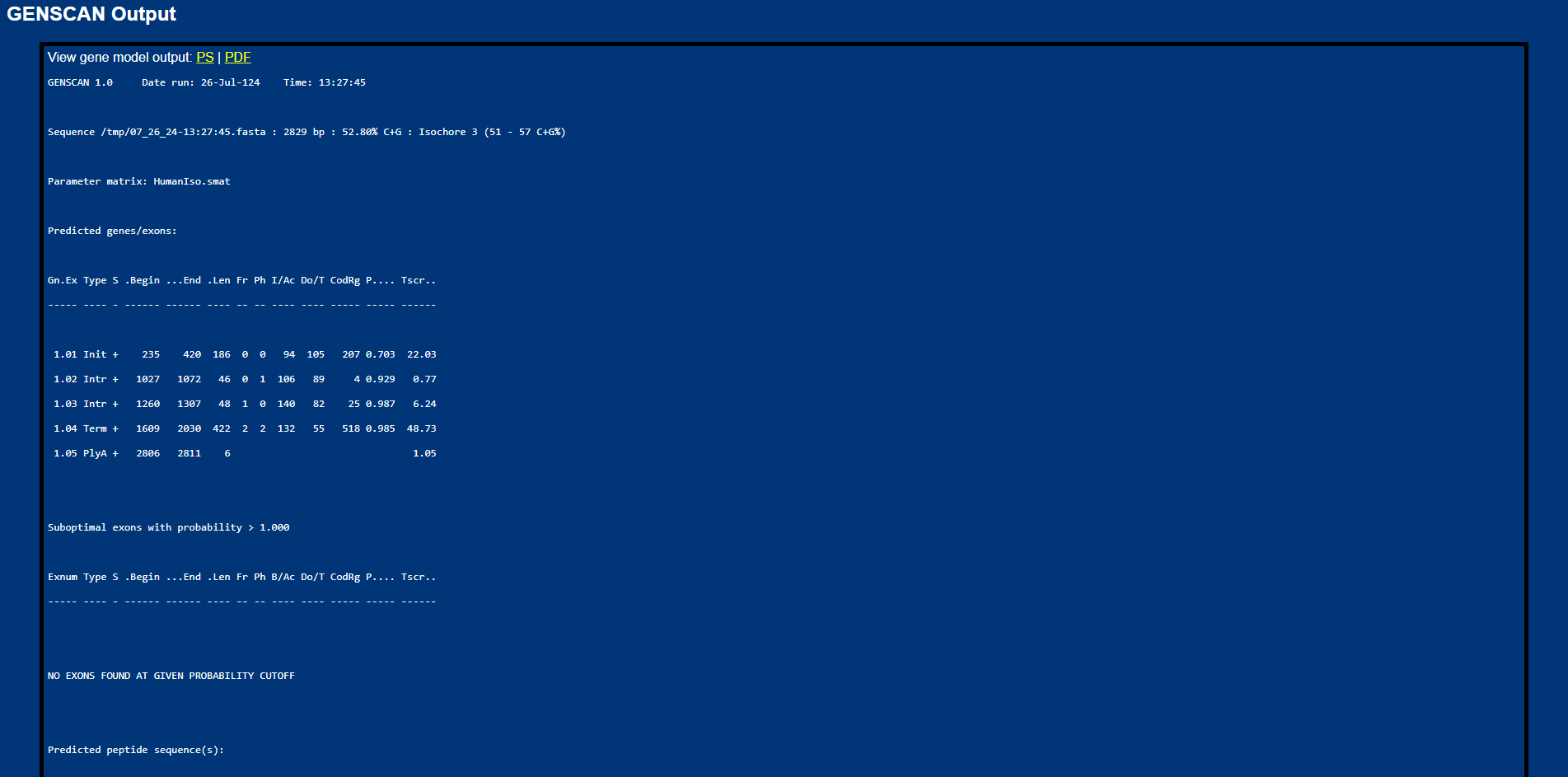
<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%206-1.png>



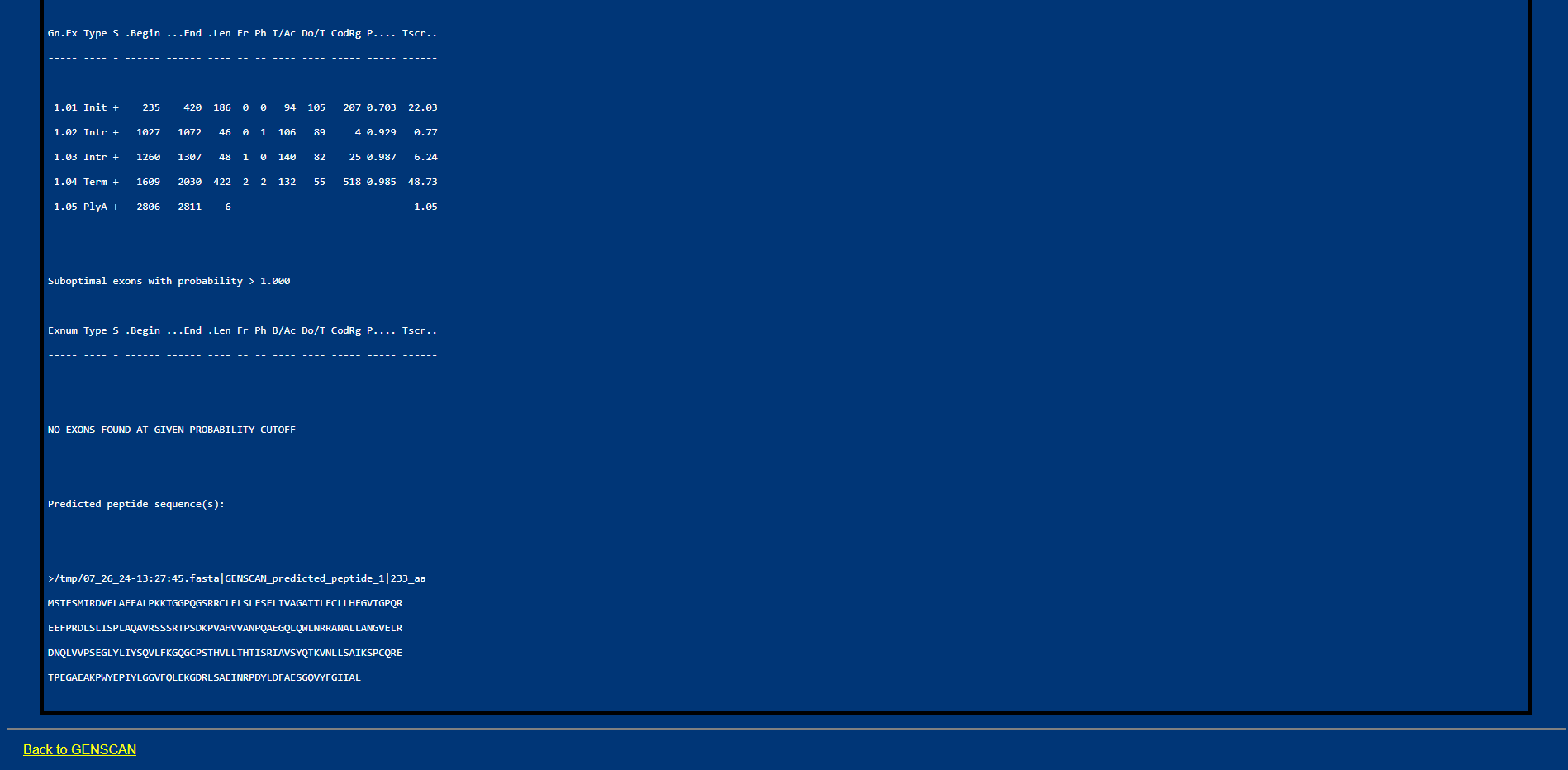
<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%206-2.png>

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

GENSCAN is a bioinformatics tool used to predict the structure and location of genes in DNA sequences, identifying both coding regions (exons) and non-coding regions (introns, intergenic regions).



<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%207-1.png>



<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%207-2.png>

The predictions of exons and the encoded amino acid sequence can be seen in the attached images. These predictions are useful for identifying new genes or studying regulatory elements and other DNA sequences that do not code for proteins but play important roles in gene regulation.

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

The file converted to PHYLIP format can be viewed at the following link:



<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/Task%208.png>

<https://github.com/Liliana223/-BioinformHER_Module-1-Mini-Project/blob/main/Module-1-Mini-Project/Files/sequence_TNF_phylip.phy>