

# Mini Project: Comprehensive Sequence Analysis of the Human TNF Gene:

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

- I visited the NCBI homepage and searched for the human TNF gene.
- After finding the correct sequence labeled "Homo sapiens TNF," I downloaded it in FASTA format.
- I chose the version ([NCBI Reference Sequence: NC\\_000006.12](#)).
- Finally, I opened the sequence in BioEdit for further analysis.

## Output:

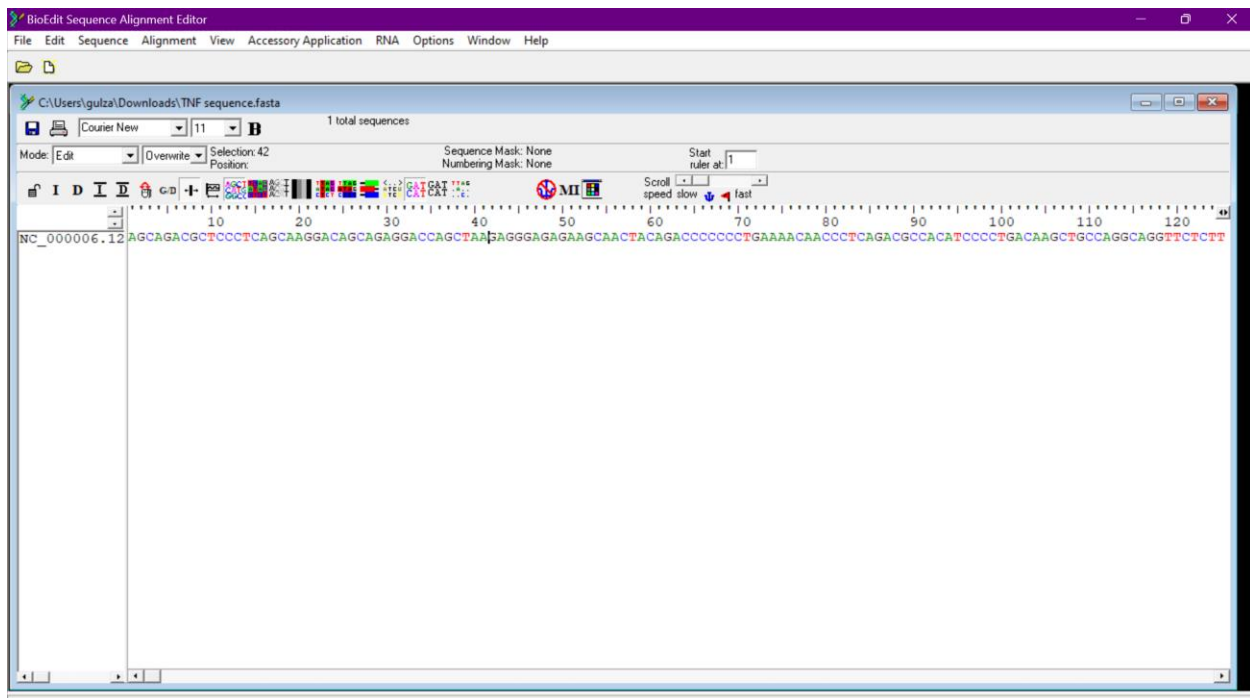


Figure 1: TNF Gene Sequence in BioEdit.

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

- I opened the gene sequence in BioEdit.
- Then, I clicked on the sequence.

- I navigated to the Sequence menu.
- I selected the "Translate" option.
- This generated the amino acid sequence.

## Output

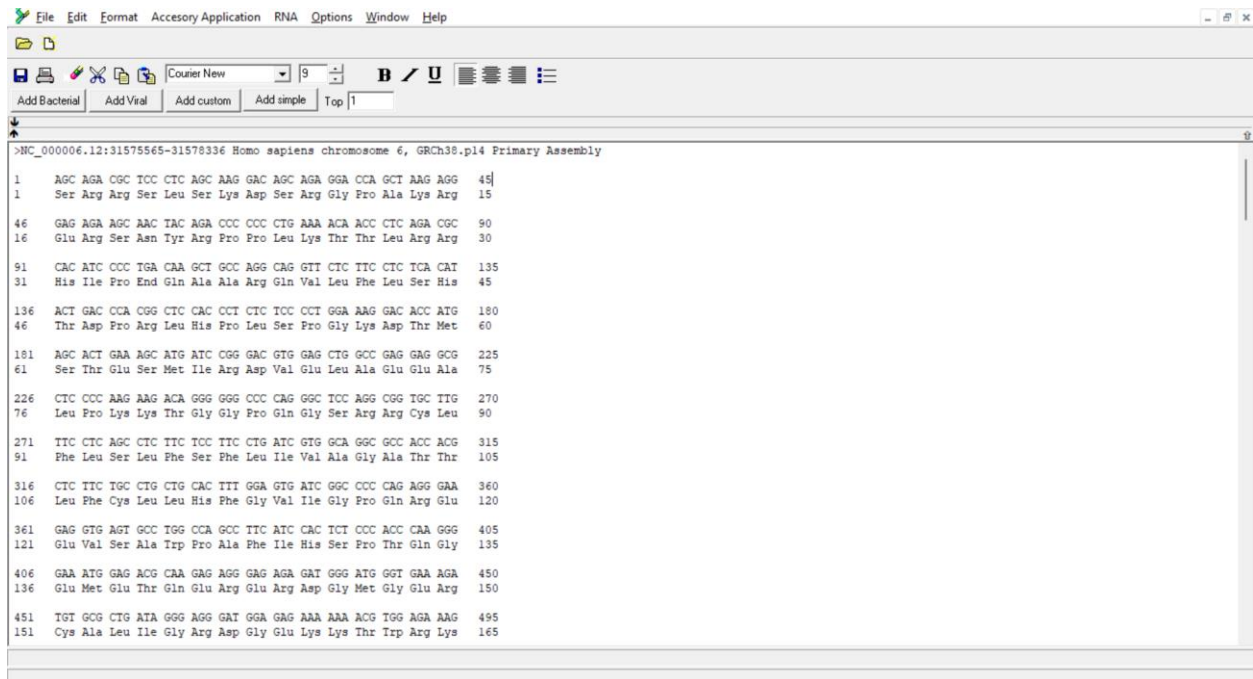


Figure 2: Translated Gene Sequence in BioEdit.

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

- I opened the gene sequence in BioEdit.
- Then, I clicked on the sequence.
- I went to the Sequence menu and selected "Nucleotide" and "Find Next ORF".
- Using this tool, I identified the ORF in the TNF sequence.
- Finally, I recorded the results.

## Output

```
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 178 to 606 (178 to 606): Frame 1 143 aa
MSTESMIRDLVLAEEALPKGTGGPGGSRRLFLSLFSLIVAGATLLPCLLHPGVIGPQREEVSANPAFIHSPITQEMETQERERDGMGERCALIGRDGEKKTWNTGMQKEMWQEMGRERERWRDRMSGTWKVLTKVNSE
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 450 to 587 (450 to 587): Frame 3 46 aa
MCADREGWREHNVKEDGDAERDVARDGEERERHMERQVWHHEGAS
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 470 to 610 (470 to 610): Frame 2 47 aa
MERHKGGERRCRKGCGHGWGREHKGDETCGLANGRCSLFVYGVME
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 659 to 781 (659 to 781): Frame 2 41 aa
MWGVRRREMEETSDMKDGETERAGNHTAKERDGGDKERRR
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 756 to 923 (756 to 923): Frame 3 56 aa
MGEIRREEDRVSGTQTLRERAVECLEGEYTDWREKTRHLRAKASAGQTGSQLFLL
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 850 to 1134 (850 to 1134): Frame 1 95 aa
MNGERKPDTSGLRAQARQAASCSFSGDSLDVNHSPSQPFPRDLSLISPLAQAVSKCLQTSFLILGLGLGVLPVWVWNGHFKVLVLGEDGWR
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1124 to 1288 (1124 to 1288): Frame 2 55 aa
MGGESRGVFSRKGSLSTFFLSLSSGSSSRIPSDKFAHVVGKSSSEVDVSNLEG
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1128 to 1235 (1128 to 1235): Frame 3 36 aa
MEVVGGVGLQSLRVSAFSLFSLQDLHLEFVIVSL
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1316 to 1519 (1316 to 1519): Frame 2 68 aa
MYGRTWRCCKEDSLSSREHWNHTGLSGILRTISWPGGMDRQRGQETQVGHAELEGQVSEPTWHE
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1434 to 1970 (1434 to 1970): Frame 3 179 aa
MTDREDNRNMGQSSRARMRVNRHGTDSPLFLSLFPANQAEQQLQWLNRRANALLANGVELRDNQLVTFSEGLYLIYSQVLFKQGCPSHVLHTHTISRAVSYQTKVNLLSAINKSPQRETPEGAENFWYEPYVLGGVFLQLEKQDLRAEINRPPDYLDFAESQGVYFGIILA
L
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1510 to 1584 (1510 to 1584): Frame 1 25 aa
MATILLLSLSPSLQQTLLRLGSSSG
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1717 to 2232 (1717 to 2232): Frame 1 172 aa
MCSPTSPSAASPPTFRSTSSLPAPARGRQGRGLRSPGMSPIWEGSSSWRVDSALRSIGFTISTPLSGRSTLGLSLPCEDEHPTFPNAPAPIPLLPSPSDTLNLFWLKKRIGGLGSEPKRLTSLNKTITSPGIGQECVACTVCKWQFLRIQTGASRTHWQLQL
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 2151 to 2336 (2151 to 2336): Frame 3 62 aa
MGLHSEVLATIKNSHWGLQNSLGFALIFDIWNHLETRPLVLARMQLQLRPHLEIDTSGF
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 2356 to 2472 (2356 to 2472): Frame 1 39 aa
MFPFLIETSRALMPEAPISIVTCVILFIYILFITYR
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 2420 to 2518 (2420 to 2518): Frame 2 33 aa
MFAIVITYILFIYILFITYLTDICVYLGDRGILGDM
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 2472 to 2624 (2472 to 2624): Frame 3 51 aa
MNVFIWETGVSWGTQRCSCLGSMFVSKTELNNRFLPCPLASVFSFYDF
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 2476 to 2550 (2476 to 2550): Frame 1 25 aa
MTTTRGSGGSGGSSSTTSGGSR
```

Figure 3: ORFs (Open Reading Frame) in TNF Sequence

Interpretation

- The ORF sequences of human TNF gene is located on chromosome 6 (GRCh38.p14 assembly). Each sequence is listed with its specific nucleotide positions and corresponding amino acid translations. These ORFs vary in length, from 25 to 179 amino acids, and are presented in different reading frames.
- The presence of repetitive and conserved motifs indicates areas that may be critical for the protein's function or its interaction with other molecules.

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

- I opened the sequence in BioEdit and clicked on it.
- I navigated to the Sequence menu and chose "Nucleotide Composition."
- Calculated the frequencies of nucleotide and GC content.

Output

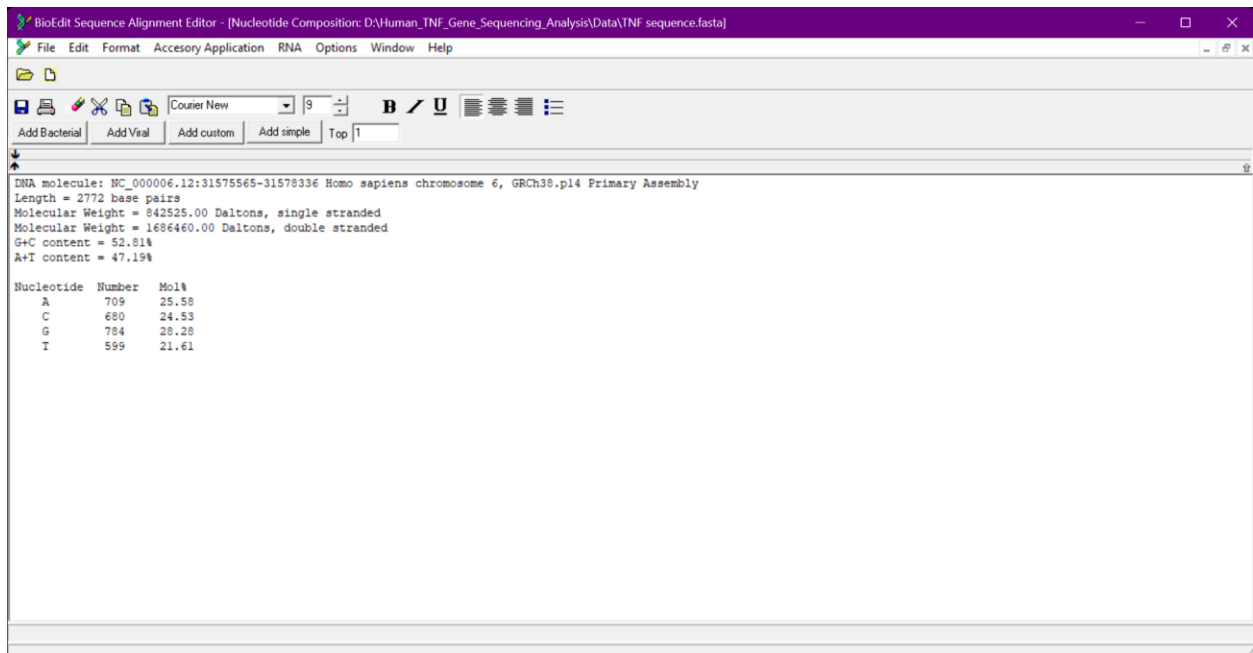


Figure 4(a): Nucleotide Composition

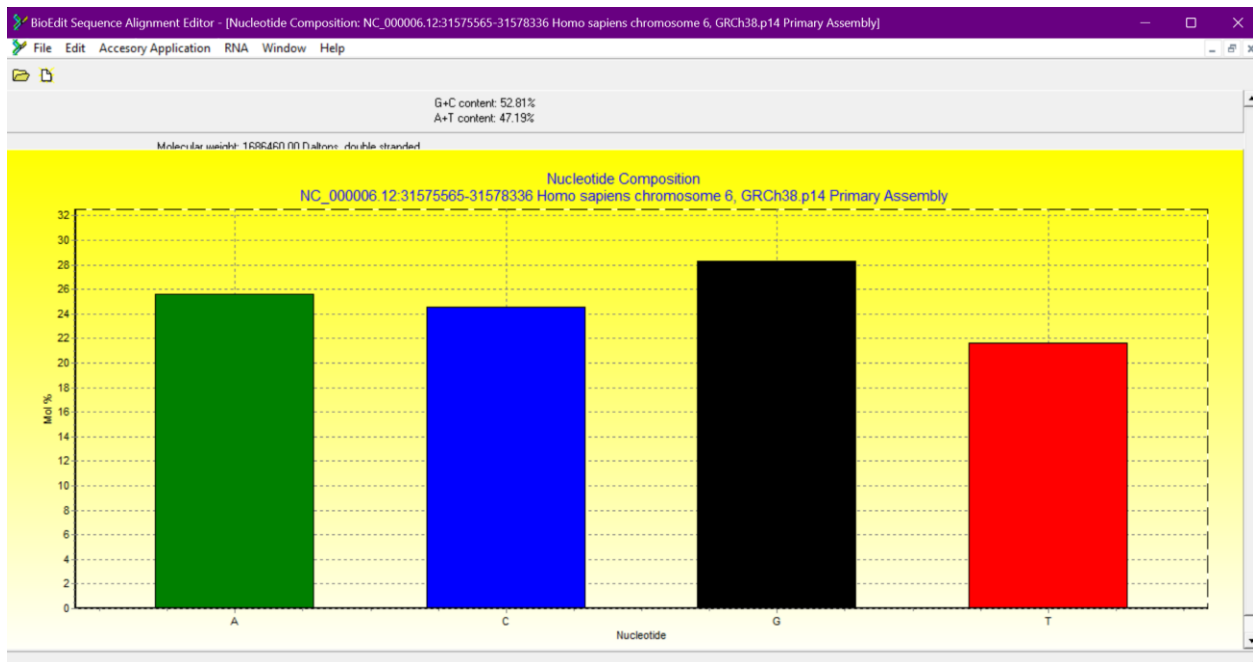


Figure 4(b): Graphical Illustration.

## Interpretation

- The DNA sequence from chromosome 6 is 2,772 base pairs long. Its GC content is 52.81%, indicating a slight GC bias, while the AT content is 47.19%. This information

suggests that the DNA sequence has a relatively balanced nucleotide composition, with a slight preference for guanine (G) and cytosine (C).

- A higher GC content often correlates with greater DNA stability due to the stronger hydrogen bonds between G and C. This might imply regions of the sequence with increased stability or potential regulatory importance.
- The molecular weight is 842,525 Daltons for the single-stranded form and 1,686,460 Daltons for the double-stranded form, useful for understanding its behavior during molecular processes.

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

- I went to the PROMO tool website to access it.
- The webpage had two steps to follow.
- In the first step, I selected "Human" as the species.
- Then, in the second step, I uploaded my sequence to identify transcription factor binding sites.
- Since I had a single sequence, I chose the first option, submitted my FASTA file, and analyzed the sequence.

## Output

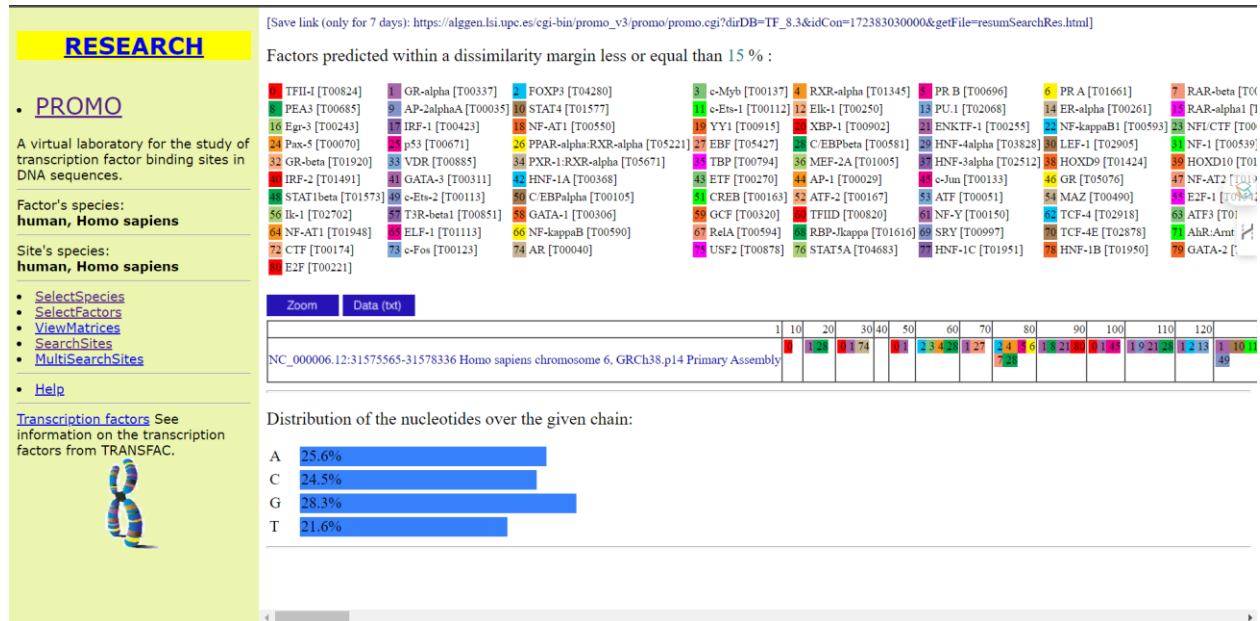


Figure 5: Transcription factor binding sites

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

- I accessed the MEME Suite tool through its website
- I uploaded the TNF gene sequence in FASTA format.
- I used the default settings and searched for motifs.

Output

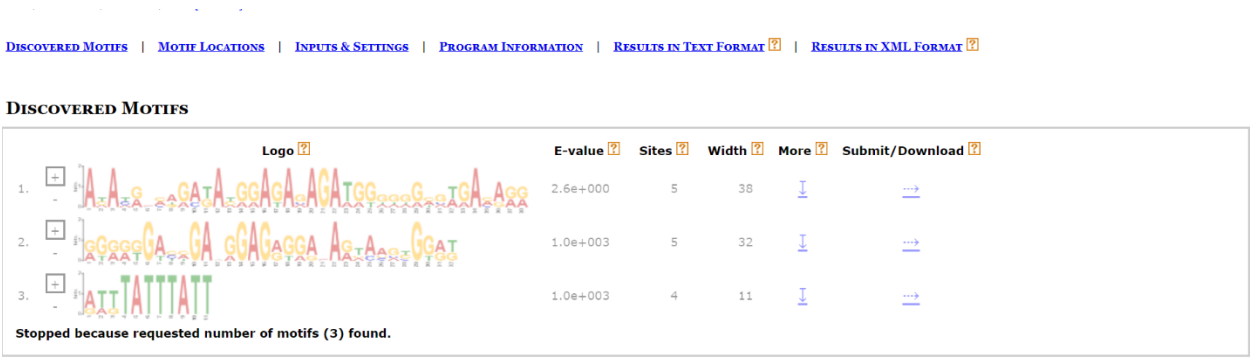


Figure 6(a): Functional motifs

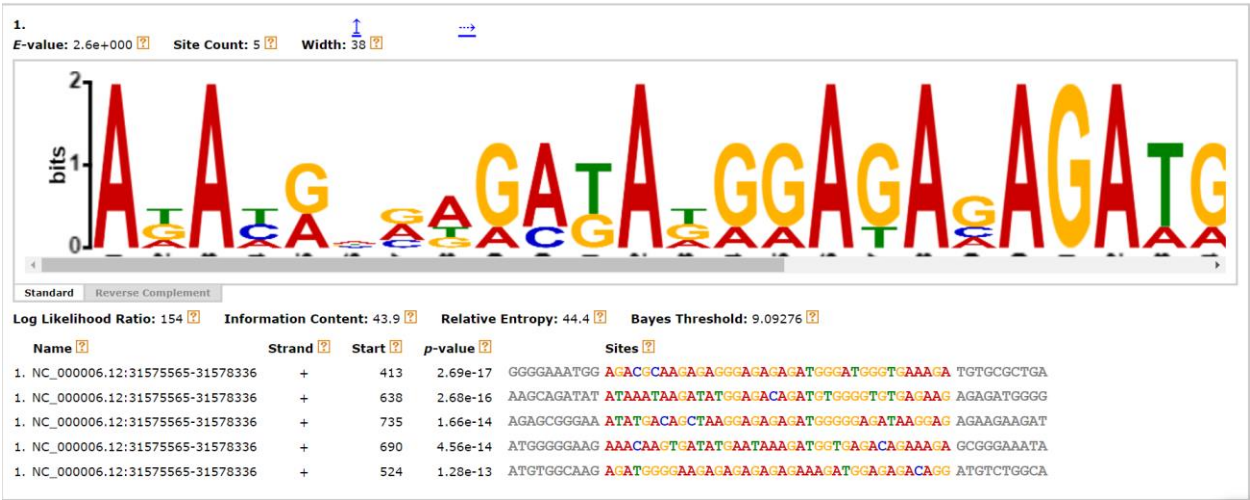


Figure 6(b): Functional motifs.



Figure 6(c): Location of motifs.

Interpretation

- The MEME Suite identified three motifs in the sequence. The motifs had low p-values indicating towards their importance. The clustering in some regions suggests that there was increased biological activity in that region.

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

- Accessed the GENSCAN tool.
- Uploaded the FASTA file of the TNF Sequence.
- Analyzed and predicted coding and non-coding region.

### Output

```
Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
-----
1.01 Init + 221 406 186 1 0 94 105 207 0.703 22.03
1.02 Intr + 1013 1058 46 1 1 106 89 4 0.929 0.77
1.03 Intr + 1246 1293 48 2 0 140 82 25 0.987 6.24
1.04 Term + 1595 2016 422 0 2 132 55 518 0.985 48.73
1.05 PlyA + 2792 2797 6 1.05
```

Figure: GENSCAN Output

### Interpretation

- GENSCAN output showed five exons on positive strand. The first exon is and initial exon. Then there are two internal exon. The fourth exon is Terminal exon. Finally the fifth exon is polyadenylation exon. The positions and length of the exons is also mentioned.



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

- Launched the TNF gene sequence in BioEdit.
- Employed the 'Save As...' option to change the file to PHYLIP format and then reviewed it in Text editor to verify.