**Guião Bioinformática**

Step 1: General and Theoretical Research

The project began with detailed general and theoretical research on the chosen gene, APOE. This preliminary phase was crucial for establishing a solid knowledge base about the gene in question. Platforms such as NCBI (National Center for Biotechnology Information) and PUBMED were used to access a wide range of information. The research included data on the structure and function of the APOE gene, the different alleles associated with it, and the diseases related to its mutations.

Step 2: Obtaining the 3D Image of the Protein

Through the Protein Data Bank (PDB) on the NCBI platform, we obtained the three-dimensional image of the protein associated with the APOE gene. This allowed us to visualize specific details, such as how different regions of the protein are organized in space.

Step 3: Selection of Homologous Sequences

Using the NCBI platform, we not only obtained the protein sequence of the APOE gene but also selected 10 homologous protein sequences from various species. This step was fundamental for performing comparative analyses (such as the phylogenetic tree, which will be discussed later). The sequences were collected in FASTA format and grouped into a single file, forming a set of 11 homologous sequences, which served as the basis for subsequent analyses.

Step 4: Multiple Sequence Alignment (MSA)

The file containing the 11 homologous sequences was uploaded to the Clustal Omega platform, which processed the sequences and generated a multiple sequence alignment. This was essential for identifying conserved regions among the different species. Additionally, the platform produced an initial phylogenetic tree and its corresponding text format.

Step 5: Phylogenetic Tree

The phylogenetic tree generated in Clustal Omega was subsequently uploaded to the iTOL (Interactive Tree Of Life) platform. This tool allowed us to visualize the phylogenetic tree in different formats: rectangular, circular, and unrooted. Moreover, we used iTOL's features to edit and enhance the representations of the phylogenetic trees, highlighting important characteristics and improving the clarity of the data presentation.

Step 6: Identification of Motifs

Using the MEME platform, motifs of the homologous sequences were obtained. We started by uploading the file with the 11 sequences, and the platform generated 3 motifs. The following presets were used in this step:

* Discovery mode: classic (the platform discovers enriched motifs in the set of sequences, with enrichment measured against a random model based on the frequency of the provided sequence letters)
* Sequence alphabet: DNA, RNA, or protein (in this specific case, homologous protein sequences were used)
* Distribution: zero or one occurrence per sequence (the platform assumes each sequence can contain at most one occurrence of each motif)
* Number of motifs sought: 3

Step 7: Identification of Regulatory Elements

Using the Genome Browser site, we searched for regulatory elements of the APOE gene. This analysis was fundamental for identifying DNA regions that control gene expression. The Genome Browser allowed us to visualize genomic data in a broader context, providing information on promoters, enhancers, and other regulatory elements that can influence the activity of the APOE gene.