Relatório

Work done by:

- Lara Ribeiro,
- Maria Benedita Marques,
- Mariana Pinho

Índice

Abstract

The APOE gene on chromosome 19 encodes apolipoprotein E, crucial in lipid metabolism and neurodegenerative diseases, particularly Alzheimer's. This study used bioinformatics tools to analyze APOE, focusing on homologous protein sequences, multiple sequence alignment, phylogenetic trees, and motifs. The NCBI database provided essential resources and homologous sequences, while Clustal Omega facilitated alignments. MEME identified key motifs, iTOL generated phylogenetic trees, and the Genome Browser examined regulatory elements. The study highlighted conserved and variable regions of APOE, emphasizing their functional and evolutionary importance. These insights are crucial for understanding genetic predisposition to neurodegenerative diseases and developing preventive and therapeutic strategies. The research demonstrates the value of integrated bioinformatics in genetic and biomedical advancements.

Introduction

The APOE gene, located on chromosome 19, encodes the apolipoprotein E, which is a critical player in lipid metabolism since it is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. However, the importance of the APOE gene goes beyond lipid metabolism. This gene is a key genetic factor in the pathogenesis of various neurodegenerative diseases, particularly notable for Alzheimer's disease, a devastating condition that affects millions of people worldwide.

Through the mastery of bioinformatics platforms and tools, it is possible to analyze various aspects inherent to the gene, such as homologous protein sequences, multiple sequence alignment, phylogenetic tree, and motifs. From these data, it is possible to draw conclusions and better understand the implications of mutations in the APOE gene.

The study of this gene can help understand the underlying mechanisms of genetic predisposition to Alzheimer's or other neurodegenerative diseases and assist in the development of preventive and therapeutic strategies. Furthermore, understanding the genetic variations in APOE can help identify individuals at risk and develop personalized interventions.

Bioinformatics Tools and Their Utility

NCBI(National Center for Biotechnology Information)
NCBI is an organization that provides access to a wide range of resources and databases on molecular biology and genetics.
In this project, the NCBI was useful for analyzing the three-dimensional structure, information retrieval (via GenBank), the identification of 10 homologous protein sequences from species other than *Homo Sapiens* (via BLASTp) and extraction of relevant articles about the gene under study (via PubMed).

Clustal Omega
Clustal Omega is a tool used for performing multiple sequence alignments (MSA), allowing the comparison of multiple sequences simultaneously.
This tool was used to obtain the alignment of the 11 homologous sequences that were subsequently crucial for the identification of conserved/non-conserved regions and evolutionary analysis.

MEME (Multiple Em for Motif Elicitation)

MEME is used to identify and analyze motifs in DNA, RNA, or protein sequences. In this case, MEME was used to obtain motifs from the protein sequences of the various species, which were fundamental for the identification of functional regions of the APOE gene and comparative analysis between species.

iTOL (Interactive Tree Of Life)

iTOL is an online tool that allows for the visualization, manipulation and annotation of phylogenetic trees. This tool was used to generate and analyze a phylogenetic tree from the multiple alignment sequences of homologous proteins, allowing the analysis of evolutionary relationships, the identification of conserved clades and other relevant conclusions about the APOE gene.

Genome Browser

Genome Browser is a visualization tool that allows you to explore and analyze genomic sequences in great detail. Offers an interactive interface for viewing genetic information including genes, genetic variations, regulatory elements and more.

In this context, the Genome Browser was useful to verify and analyze the regulatory elements associated with the APOE gene.

Results

Figure 1- APOE motifs generated by the 11 sequences (source: MEME)

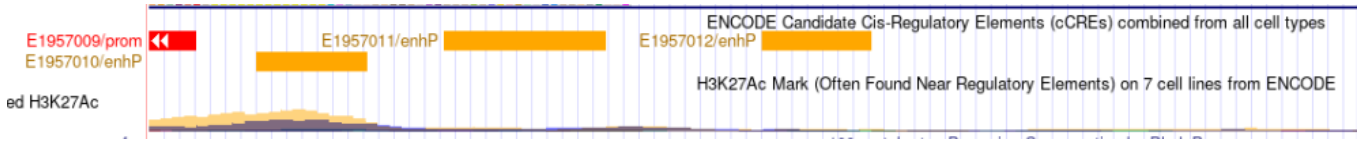

Figure 2- Regulatory elements of the APOE gene (Source: Genome Browser)

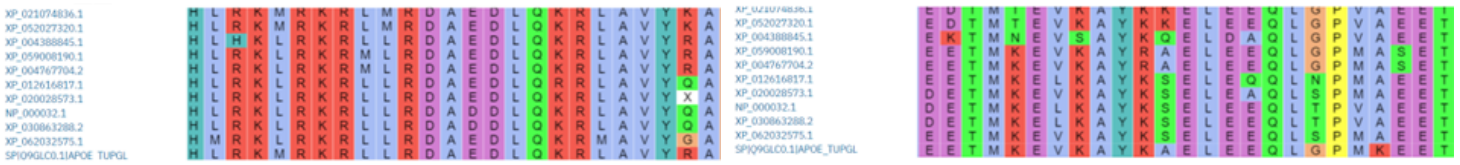

Figure 3, 4- Multiple Sequence Alignment for APOE (source: Clustal Omega)

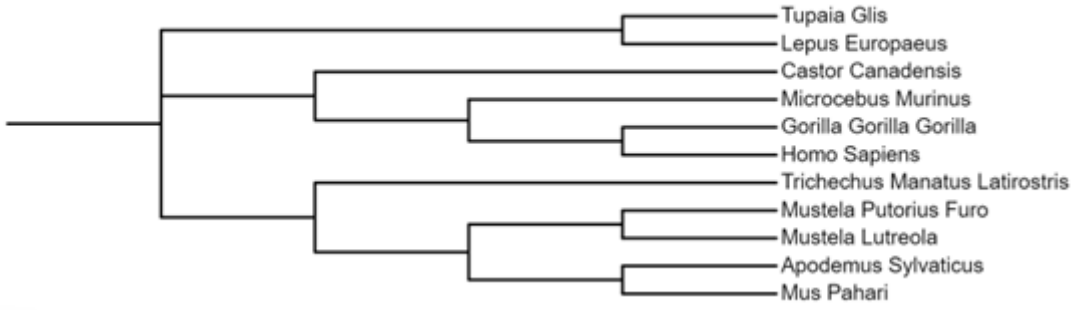

Figure 5- Phylogenetic Tree with APOE homologues in *Homo Sapiens* (source: iTOL)

Discussion of results

In the study carried out on the APOE gene, several important aspects were analyzed, including motifs, domains, alignments, phylogenetic tree and regulatory elements. The results reveal interesting and relevant characteristics about the functionality and importance of the gene.

Motifs

Motifs are recurring patterns in biological sequences, such as DNA, RNA or protein, that have a specific biological function or meaning.

Firstly, the motif discovery approach was used, 11 sequences of homologous proteins were used. In fact, three different motifs were generated, which present combinations of hydrophobic, hydrophilic, and charged residues, suggesting the formation of alpha-helices and beta-sheets. The first motif indicates involvement in protein-protein interactions and receptor binding. The second motif suggests electrostatic interactions with nucleic acids or other proteins, binding to DNA or RNA, and regulating gene expression. The third motif is associated with protein structure stabilization, protein-protein interactions, membrane anchoring, and lipid interaction. The discovery of these motifs in 11 different species highlights their strong evolutionary conservation, underscoring their biological importance in the APOE gene.
Secondly, the motif search approach was used where we know the motif in advance and aim to analyze the homologous protein sequence of the APOE gene for the presence of the motif. The analysis of the first motif suggests the protein's capacity to form ionic and hydrophobic interactions. The second motif shows a predominance of hydrophobic amino acids such as P, V, L, and some polar and charged ones like S, D, Q, suggesting that this region is important for the protein's structural stability. The third motif combines hydrophobic amino acids, indicating that the protein has the potential for hydrophobic and ionic interactions, with structural flexibility.

In conclusion, the analysis of motifs using two complementary approaches - motif search and motif discovery - provides a more comprehensive understanding of the conserved and functional regions of the APOE protein and its homologous proteins. This integrated analysis offers a more robust view of the functionality of the conserved regions of APOE, highlighting the importance of combined approaches for a complete understanding of complex proteins.

Domains

The APOE gene features two distinct domains: PCSK9_C-CRD and NR_LBD.

The PCSK9_C-CRD domain, located in the C-terminal region, is crucial for regulating hepatic receptors for low-density lipoproteins, playing a fundamental role in lipid metabolism. Moreover, it underscores its importance in maintaining lipid homeostasis and influencing cardiovascular

health. On the other hand, the NR_LBD domain, situated in the N-terminal region, includes members of the steroid hormone receptor family, thyroid hormone receptors, retinoid receptors, cholesterol byproduct receptors, and lipid receptors. This domain is essential for the regulation of interactions with these molecules, highlighting its critical role in managing hormonal and lipid signaling pathways.

Together, these domains contribute to the multifaceted functions of the APOE gene, influencing both lipid metabolism and hormone regulation. The dual functionality emphasizes the gene's significant impact on various physiological processes and its potential implications in metabolic and cardiovascular diseases.

Multiple Sequence Alignments

Multiple sequence alignments are crucial for comparing and identifying conserved and variable regions among sequences, revealing insights into the evolution and function of genes.

On one hand, conserved regions, such as "RDAEDLQ," which remain similar across different sequences, suggest functional or structural importance, as they are preserved by natural selection. These regions are likely critical for maintaining the essential functions of the protein.

On the other hand, variable regions, such as 'KAYKKELEEQLGPVAEETRARLAKEVQAAQAR,' indicate portions of the sequence that can tolerate mutations without significant loss of function. This reflects adaptability and functional diversity, allowing the protein and acquire new functions or adapt to different environmental conditions.

By analyzing 11 homologous sequences from species with considerable variation, it was concluded that there is a greater number of non-conserved regions compared to conserved regions. This observation aligns with the expectation that, while some parts of the protein must remain unchanged to maintain critical functions, other parts can evolve more freely, contributing to the overall adaptability and diversity of the protein's functions.

Phylogenetic Tree

Based on the analysis of the phylogenetic tree, several conclusions were drawn:

Firstly, *Homo sapiens* and *Gorilla gorilla gorilla* are closely related, reflecting the expected proximity between humans and gorillas. Interestingly, *Microcebus murinus* shows a surprising closeness to humans and gorillas, suggesting a close relationship within primates.

Secondly, *Tupaia glis* and *Lepus europaeus* are closely related, indicating a close phylogenetic relationship between these non-primate mammals. *Mustela putorius* furo and *Mustela lutreola*

are presented as sister groups, similarly, *Apodemus sylvaticus* and *Mus pahari* also form a group, indicating their close evolutionary relationship.

Castor canadensis and *Trichechus manatus latirostris* diverge earlier from other species, suggesting a greater evolutionary distance between them and the other listed groups.

Overall, the phylogenetic analysis underscores the evolutionary relationships among these species, highlighting both close relationships within groups and the distinct evolutionary paths of species.

Regulatory Elements

The expression of APOE is regulated by several regulatory elements that control when, where, and in what quantity the gene is expressed. In fact, the APOE gene presents 4 regulatory elements: EH38E1957009, EH38E1957010, EH38E1957011, and EH38E1957012.
The first element, located in the promoter region of the APOE gene, is essential for the initiation of transcription. Moreover, the second element enhances the expression of the APOE gene, especially in response to specific factors present in certain tissues such as the liver and the brain. The third element acts by increasing APOE expression under specific conditions, facilitating the binding of transcription co-activators. Finally, the fourth element works together with the other enhancers to promote adequate levels of APOE expression in response to metabolic and cellular signals.

Conclusion

The study of the APOE gene using bioinformatics tools revealed crucial information about its structure, function, and evolution. Analyses of homologous protein sequences, multiple sequence alignment, phylogenetic tree, and motifs allowed for the identification of conserved and variable regions, clarifying the functional importance of different parts of the APOE gene. These findings have significant implications for understanding the genetic predisposition to neurodegenerative diseases such as Alzheimer's and can contribute to preventive and therapeutic strategies, as well as the identification of individuals at risk for personalized interventions. The integrated bioinformatics approaches provided a robust view of the functional regions of APOE, highlighting the importance of these tools in genetic and biomedical research.

Link to GitHub: https://github.com/lararib0/Gene_APOE_Research