REPORT

TNF (Tumor Necrosis Factor)

The following report, prepared within the scope of the Bioinformatics Laboratories curricular unit, was carried out by students Afonso Santos (up202406167), Carolina Jesus (up202406239) and Rafael Santos (up202405265), under the guidance of professors Pedro Ferreira and Miriam Santos.

Abstract

The TNF gene, located on chromosome 6, encodes the TNF-α, a pivotal cytokine in inflammation, immunity, and apoptosis, with significant roles in autoimmune diseases, cancer, and infectious diseases. This study employed bioinformatics tools to analyse TNF, focusing on homologous protein sequences, multiple sequence alignment, phylogenetic trees, and conserved motifs. The NCBI database provided essential resources and homologous sequences, while Clustal Omega produced alignments. WebLogo facilitated the identification of key functional motifs, iTOL generated phylogenetic trees, and the Genome Browser examined promoter regions and regulatory elements. The study revealed conserved and divergent regions of TNF, underscoring their functional and evolutionary implications. These findings are critical for understanding TNF's dual role in protective immunity and pathological inflammation, as well as for designing targeted therapies. The research highlights the power of integrated bioinformatics in advancing immunogenetics and precision medicine.

Introduction

The tumour necrosis factor (TNF) gene, located on chromosome 6, encodes a multifunctional proinflammatory cytokine, the TNF-α, which is a member of the TNF superfamily. Furthermore, it plays a crucial role in regulating a broad range of biological processes (including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation) and it has been associated with various diseases, such as autoimmune disorders, insulin resistance, psoriasis, rheumatoid arthritis, ankylosing spondylitis, tuberculosis, autosomal dominant polycystic kidney disease, and cancer.

The use of bioinformatics platforms and tools enables the analysis of various gene-related aspects, such as homologous protein sequences, multiple sequence alignments, phylogenetic trees, and motifs. This analysis facilitates the derivation of conclusions and enhances the understanding of mutations in the TNF gene and their implications.

It goes without saying that understanding mutations in the TNF gene is essential for uncovering disease mechanisms, identifying diagnostic biomarkers, and

developing targeted therapies. Its significant role in immune system functionality and inflammatory pathways underscores the importance of its study in addressing major health challenges.

Bioinformatics Tools | Description and utility

NCBI (National Center for Biotechnology Information)

NCBI is an organization that provides access to a variety of resources and databases on molecular biology and genetics.

In this project, the NCBI was useful for analysing the three-dimensional structure, information retrieval (via GenBank), the identification of 10 homologous protein sequences from species other than Homo Sapiens and extraction of relevant articles about the gene understudy (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.

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 analyse a phylogenetic tree from the multiple sequence alignment of homologous proteins, allowing the analysis of evolutionary relationships, the identification of conserved clades and other relevant conclusions about the TNF gene.

WebLogo

WebLogo is a bioinformatics tool designed to generate sequence logos, which visually represent conserved motifs in aligned DNA, RNA, or protein sequences. These logos illustrate the frequency of each nucleotide or amino acid at specific positions, highlighting patterns of conservation and variability.

Genome Browser

Genome Browser is a visualization tool that allows you to explore and analyse genomic sequences in great detail. Offers an interactive interface for viewing genetic information, including genetic variations and regulatory elements. In this context, the Genome Browser was useful to verify and analyze the regulatory elements associated with the TNF gene.

Results

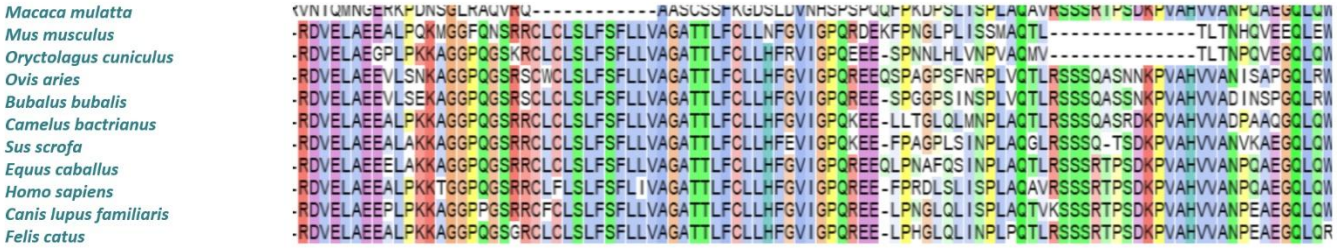

Figure 1 | Multiple Sequence Alignment for TNF (source: Clustal Omega)

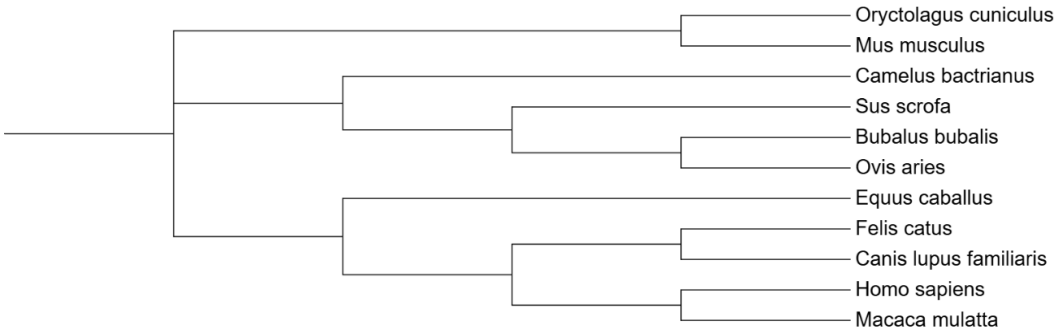

Figure 2 | Phylogenetic Tree with TNF homologues (source: iTOL)

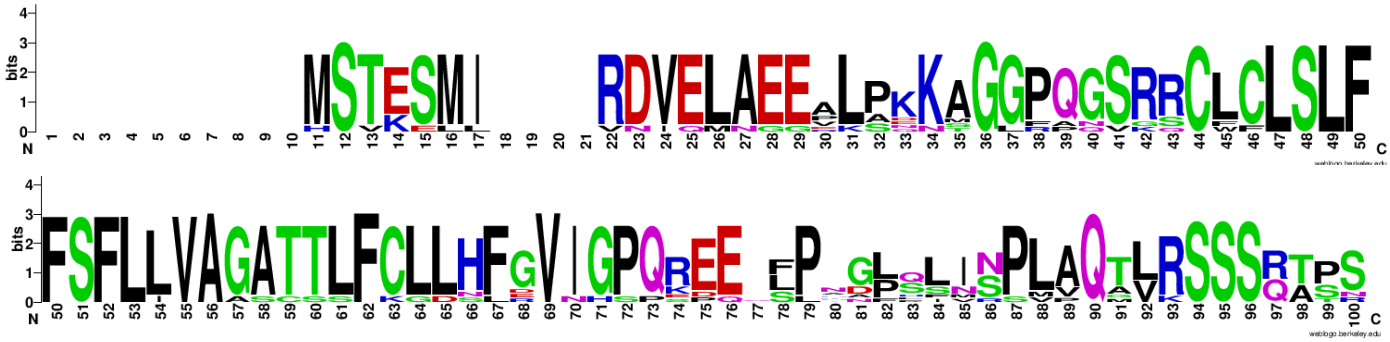

Figure 3 | TNF motifs generated by the 11 sequences (source: WebLogo)

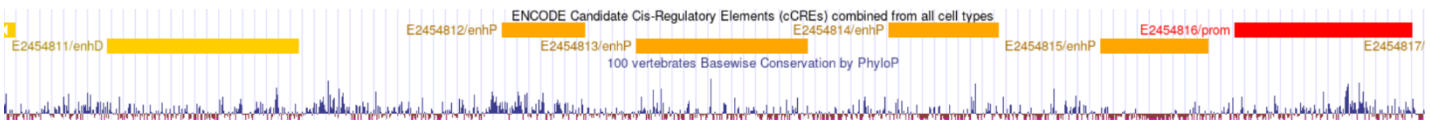

Figure 4 | Regulatory elements of the TNF gene (Source: Genome Browser)

Discussion of results

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

Domains

The TNF system involves transmembrane TNF (mTNF) cleavage by TACE—a metalloprotease with a catalytic TNF homology domain (THD)—releasing soluble TNF-α. This cytokine binds TNFR1, which triggers apoptosis and inflammation via its death domain (activating caspases and NF-κB), and TNFR2, promoting survival and immune modulation.

Lymphotoxin-alpha (LTα) diversifies signalling by forming heterotrimers (e.g., LTαB) that engage TNF receptors. TACE cleavage not only activates TNF-α but also leaves a remnant receptor fragment, potentially influencing cellular responses, highlighting TNF's dual roles in inflammation and cell fate regulation.

Multiple Sequence Alignments

Multiple sequence alignments are crucial for comparing and identifying conserved and variable regions among sequences.

On one hand, conserved regions, such as "CLCLSLFSFLLVAGATTLFCLLHFGV" 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 "LPQKMGGFONSRR" indicate portions of the sequence that can tolerate mutations without significant loss of function. Nevertheless, regions like "-LLTGLQLMN" show significant divergence (variability) across the homologous sequences.

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

Phylogenetic Tree

Bases on the analysis of the phylogenetic tree, several conclusions were drawn.

Firstly, *Homo sapiens* (humans) and *Macaca mulatta* (rhesus macaque) share a strong evolutionary link as primates, reflecting their common ancestry. Similarly, *Canis lupus familiaris* (dog) and *Felis catus* (cat) are closely related as

members of the order Carnivora, though they belong to different families (Canidae and Felidae, respectively).

Secondly, *Sus scrofa* (pig), *Bubalus bubalis* (water buffalo), and *Ovis aries* (sheep) cluster together as artiodactyls, while *Equus caballus* (horse) diverges slightly as a perissodactyl, yet remains within the broader ungulate grouping. *Camelus bactrianus* (Bactrian camel), though an artiodactyl, branches earlier, indicating a more distant relationship within this clade.

Lastly, *Oryctolagus cuniculus* (rabbit) and *Mus musculus* (mouse) show a notable divergence, as rabbits belong to the order Lagomorpha and mice to Rodentia, highlighting their distinct evolutionary paths despite both being small mammals.

Overall, the phylogenetic tree emphasizes both the close evolutionary ties within major mammalian groups (primates, carnivorans, and ungulates) and the clear divergence between more distantly related lineages. This underscores the complex branching patterns that define mammalian evolution.

Motifs

Through the MSA of 11 TNF homologous, 2 motifs were highlighted and consequently the respective structural and functional regions were identified.

Motif (1 – 50) | The alignment reveals a conserved charged region, RDVELAEE, dominated by glutamic acid (E) and arginine (R), likely critical for receptor binding or signalling due to its polarity. PQGSRR, a conserved region with proline (P) and serine (S), potentially enables conformational flexibility or phosphorylation-dependent regulation. Adjacent residues like LCLSLF show hydrophobic conservation (leucine, phenylalanine), suggesting structural stabilization or membrane interaction.

Motif (50–100) | A hydrophobic stretch (FSFLLVAGATTLF) with non-polar residues (leucine, phenylalanine, valine) indicates a potential transmembrane domain or core structural region essential for protein stability. Nearby, polar residues like glutamine (Q) and glutamic acid (E) (GPQREE) suggest roles in protein-protein interactions or signal transduction. The RSSS cluster imply phosphorylation sites (serine, threonine) for regulatory functions.

Regulatory Elements

Regulatory elements are specific DNA sequences (such as promoters, enhancers, and silencers) that control gene expression by interacting with transcription factors and other proteins to turn genes on or off in response to cellular signals.

The TNF gene is regulated by a complex array of cis-regulatory elements that ensure precise control of its expression in response to immune and inflammatory signals.

Among these, the promoter region, such as E2454816/prom, plays a central role in initiating transcription, containing multiple binding sites for key transcription factors like NF-κB, AP-1, and CREB, which are rapidly activated in response to inflammatory stimuli.

Adjacent to the promoter, proximal enhancer-like elements, for instance E2454815/enhP, enhance transcription by facilitating the recruitment of additional transcriptional machinery and stabilizing the interaction between enhancers and the core promoter.

Further upstream, distal enhancer-like elements, such as E2454811/enhD, contribute to long-range chromatin interactions that modulate TNF expression in a cell type and stimulus-specific manner. These elements often function through chromatin looping mechanisms, bringing enhancers into physical proximity with the promoter, thereby amplifying transcriptional responses under tightly regulated conditions.

Together, these regulatory domains form an integrated network that finely tunes TNF expression and prevents inappropriate or excessive inflammatory signaling.

Conclusion

The investigation of the TNF (Tumor Necrosis Factor) gene through bioinformatics approaches provided critical insights into its structure, function, and evolutionary conservation. Multiple sequence alignment of homologous TNF sequences across species, phylogenetic analysis, and motif identification revealed highly conserved regions, such as the TNF homology domain (THD), which is essential for receptor binding and trimerization. The presence of conserved motifs (e.g., RDVELAEE and LCLSLF) suggests their functional significance in immune signaling and apoptosis regulation.

These findings have important implications for understanding inflammatory and autoimmune diseases, including rheumatoid arthritis and Crohn's disease, where TNF plays a key role. The identification of conserved structural and functional domains could aid in the development of targeted therapies, such as TNF inhibitors (e.g., infliximab, adalimumab), and improve personalized treatment strategies for patients with dysregulated TNF activity.

The integration of bioinformatics tools, including sequence alignment algorithms, phylogenetic reconstruction, and motif analysis, provided a comprehensive understanding of TNF's molecular mechanisms. This study underscores the value of computational biology in biomedical research, enabling the exploration of disease-associated genetic variations and potential therapeutic targets.

https://github.com/afonsogsdantos/TNF-GENE