Relatório para o trabalho de Laboratórios de Bioinformática

A pesquisa sobre o gene HBB no âmbito da cadeira de laboratórios de bioinformática (Licenciatura de Bioinformática na Faculdade de Ciências da Universidade do Porto) foi realizada pelas alunas Cristiana Silva (up202305464), Filipa Ferreira (up202305895) e Filipa Marinha (up202304935). O presente relatório resume o trabalho efetuado, isto é, a pesquisa teórica sobre o gene que envolveu o uso de diversas ferramentas bioinformáticas e a criação e desenvolvimento de um site sobre o respetivo gene em estudo.

O gene HBB, também conhecido como gene da subunidade beta da hemoglobina, desempenha um papel crucial na síntese da hemoglobina, componente responsável pelo transporte de oxigénio presente nos glóbulos vermelhos. Localizado no cromossoma 11, este gene codifica a proteína beta-globina, que se combina com a alfa-globina para formar a hemoglobina A, a forma predominante de hemoglobina em adultos. Mutações no gene HBB são responsáveis por uma variedade de distúrbios, sendo os mais notáveis a doença falciforme e a beta-talassemia. Estes distúrbios genéticos levam a manifestações clínicas significativas, uma vez que o transporte de oxigénio é prejudicado e há destruição das hemácias.

Compreender a estrutura, função e mutações do gene HBB é essencial para entender a fisiopatologia dos distúrbios sanguíneos relacionados e para desenvolver estratégias terapêuticas direcionadas. Este relatório investiga a biologia molecular do gene HBB, os mecanismos genéticos subjacentes às suas mutações e as condições clínicas resultantes. Posto isto, a realização desta pesquisa e posterior desenvolvimento do site contribuíram para desenvolver conhecimento e experiência no uso dos diversos softwares bioinformáticos, identificar diversos aspetos sobre o gene em estudo e realizar alinhamentos e diversas árvores filogenéticas.

Relativamente às diversas ferramentas bioinformáticas utilizadas ao longo da pesquisa, ressalta-se a seguinte breve descrição das mesmas:

- O NCBI é uma das bases de dados de bioinformática mais utilizadas mundialmente. Explorámos este para obter as informações mais importantes sobre o nosso gene como por exemplo o ID do gene humano, o número de exões, a localização cromossómica, outras espécies que têm esse gene e a sua sequência. Também desfrutámos deste para obter as sequências das proteínas homólogas, que identificámos previamente no BLAST.
- O UniProt, tal como o NCBI é uma base de dados. Contudo, esta foca-se maioritariamente em proteínas. Aqui encontrámos diversas informações sobre a proteína como o tamanho, a função molecular, a sequência e a sua estrutura.
- O EMBL-EBI disponibiliza vários serviços e ferramentas bioinformáticas, porém utilizámos mais especificamente o Clustal Omega para fazer o alinhamento múltiplo de sequências e também para analisar a árvore filogenética.

- O iTOL foi usado para realizar a árvore filogenética, visto que esta é uma ferramenta específica para a manipulação e visualização de árvores filogenéticas.

Após a realização de toda a pesquisa sobre o gene HBB e a manipulação das diversas ferramentas bioinformáticas descritas anteriormente, foram retirados os resultados apresentados e discutidos abaixo.

Realizámos um alinhamento múltiplo de sequências (MSA) através do site EMBL-EBI, utilizando especificamente a ferramenta “Clustal Omega”. Foram selecionados os resultados obtidos do alinhamento a cores (figura acima), visto que é mais fácil distinguir as diferenças entre as sequências deste modo.

Todas as sequências utilizadas neste alinhamento possuem 147 aminoácidos, ou seja, possuem o mesmo comprimento, o que indica uma elevada conservação de informação ao longo da cadeia evolutiva dos espécimes selecionados. Apesar disso, as proteínas não são todas 100% iguais entre si.

Uma leitura mais aprofundada desta figura revela a existência de um total de 12 mutações que terão ocorrido a nível dos nucleótidos dos aminoácidos. A maior parte destas sucedeu antes do aminoácido 60, verificando-se apenas 4 após essa posição. Estas mudanças encontram-se representadas por diferentes símbolos: o “:” indica que todos os aminoácidos possuem o mesmo tamanho de resíduo e hidropatia, o “.” que possuem o mesmo tamanho de resíduo ou hidropatia e o “ “ que não têm nenhum dos aspetos em comum.

Destes termos, a hidropatia mostra-se especialmente importante, visto que categoriza um resíduo em hidrofílico ou hidrofóbico. A mudança desta propriedade num aminoácido pode revelar-se impactante no modo como a proteína atua, pois pode reagir de modo diferente em certas interações químicas.

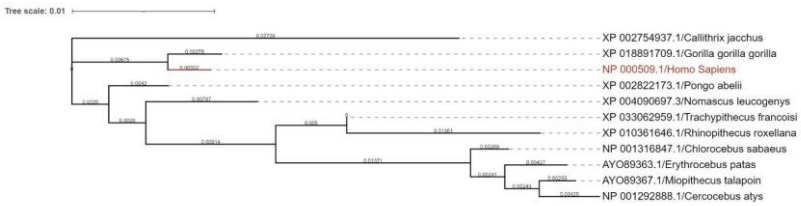
Para além destes símbolos, existe mais um, o "*". Ao contrário dos outros, este não serve para sinalizar uma disparidade, mas sim uma elevada semelhança de aminoácidos nessa posição. Como a grande maioria do MSA encontra-se legendado por este símbolo, pode-se concluir que os genes selecionados possuem uma elevado grau de semelhança entre si, visto que possuem alta compatibilidade tanto em termos de tamanho de resíduo como em hidropatia, tendo propriedades físico químicas idênticas.

Esta similaridade normalmente indica uma função semelhante e uma origem evolutiva comum, o que corresponde aos resultados que seriam de esperar tendo em conta a proximidade entre as espécies. Confirma-se então que estas sequências têm relações funcionais, estruturais e evolutivas idênticas entre si.

Para além disto, também é possível concluir que a maioria da sequência possui um papel relevante no funcionamento da proteína, visto que se verificaram poucas mutações significativas.

Por fim, ainda foi criado um logo utilizando o site "WebLogo" para se observar a análise realizada de um modo mais intuitivo. Este logo tem como base a frequência com que cada nucleótido ocorre em cada posição. Letras de maiores dimensões indicam a predominância de um certo aminoácido nessa região.

A grande maioria das posições possuem apenas uma letra, mas algumas, tal como a 10, demonstram o quão suscetíveis certos locais podem ser a mutações.

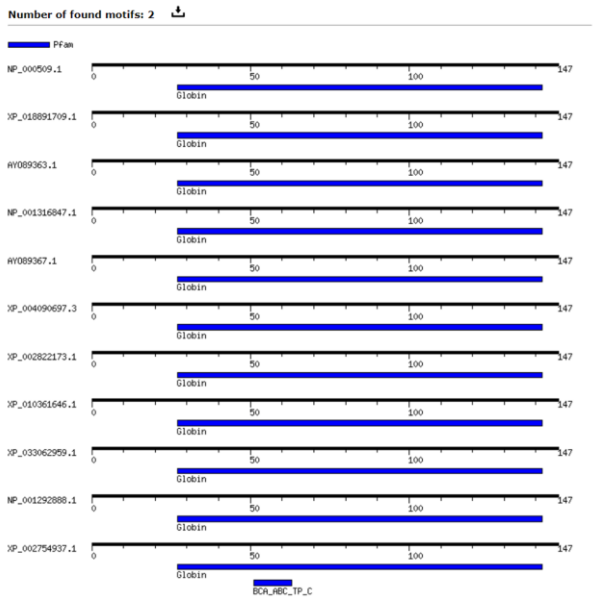


Utilizando o MSA, foi criada uma árvore filogenética através o site “ITOL”. Todas as sequências encontram-se legendadas com o nome da respetiva espécie a que pertencem de modo a permitir uma visualização facilitada do gráfico.

O ramo do Homo Sapiens, que se encontra destacado a vermelho, é o segundo mais próximo da raíz, indicando uma diferença mínima relativamente ao ancestral comum de todas as sequências analisadas. Relativamente a este ramo, ainda se pode concluir que a espécie mais próxima a esta é a do Gorilla Gorilla Gorilla, apresentando uma diferença mínima em termos de distância evolucionária. Por outro lado, a espécie do Cercocebus Atys é a que apresenta a maior disparidade.

Devido à elevada semelhança entre o ser humano e a espécie do gorila, foi realizado um alinhamento entre as suas sequências para contemplar a desigualdade que existe. Verificou-se que existe a diferença de apenas um aminoácido e, como ocorreu a mudança de arginina para lisina, existe a possibilidade de esta ter sido provocada pela transformação de um nucleótido.

Ainda é possível contemplar que a espécie Callithrix jacchus é a mais divergente delas todas, encontrando-se completamente isolada no seu ramo.



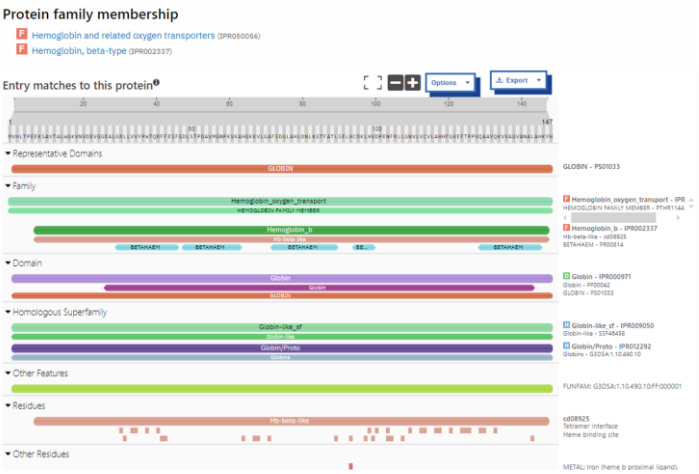
Pfam (2 motifs)

Query	Pfam	Position(Independent E-value)		Description
NP_000509.1	Globin	27..142(4.3e-33)	Detail	PF00042, Globin
XP_018891709.1	Globin	27..142(2e-33)	Detail	PF00042, Globin
AYO89363.1	Globin	27..142(7.2e-34)	Detail	PF00042, Globin
NP_001316847.1	Globin	27..142(5.6e-34)	Detail	PF00042, Globin
AYO89367.1	Globin	27..142(7.2e-34)	Detail	PF00042, Globin
XP_004090697.3	Globin	27..142(1.2e-33)	Detail	PF00042, Globin
XP_002822173.1	Globin	27..142(1.1e-33)	Detail	PF00042, Globin
XP_010361646.1	Globin	27..142(2.3e-33)	Detail	PF00042, Globin
XP_033062959.1	Globin	27..142(1.2e-33)	Detail	PF00042, Globin
NP_001292888.1	Globin	27..142(3.9e-34)	Detail	PF00042, Globin
XP_002754937.1	Globin	27..142(5.2e-33)	Detail	PF00042, Globin
XP_002754937.1	BCA_ABC_TP_C	51..63(0.12)	Detail	PF12399, Branched-chain amino acid ATP-binding cassette transporter

Através do site “GenomeNet”, foram gerados os motivos de todas as sequências envolvidas no estudo. Nestas figuras, é possível observar que todas elas possuem um motivo em comum, caracterizado por um elevado comprimento, o que seria de

esperar, tendo em conta o baixo número de mutações anteriormente verificados, solidificando a ideia que a maioria da sequência se encontra ligada ao funcionamento do gene.

Apesar disso, uma das sequências destaca-se das outras: o gene XP_002754937.1, da espécie *Callithrix jacchus*, possui um motivo que não se encontra presente nos outros. Tendo em conta que este é o gene mais díspar dos outros, a presença deste motivo coincide com a sua alta divergência evolutiva.



A imagem acima apresenta uma análise realizada pelo site “InterPro”, demonstrando diversas categorias dentro do gene, sendo uma delas a “Domain”. Nesta, é possível verificar que o domínio principal é o da globina, que corresponde ao transporte do sangue realizado pelas hemácias. Este domínio é observável em todas as sequências envolvidas, o que comprova que a função deste gene se encontra preservada nos espécimes selecionados.

A nível de dificuldades sentidas, destaca-se a análise dos motivos e dos domínios, onde foram percorridos múltiplos sites até ser encontrado um que fornecesse resultados fáceis e claros de serem interpretados. A grande quantidade de informação disponibilizada pelas ferramentas mostrou-se uma adversidade, sendo mais difícil filtrar a informação crucial e a secundária.

Em suma, este relatório apresenta uma análise detalhada do gene HBB, responsável pela codificação da subunidade beta da hemoglobina, uma proteína crucial para o transporte de oxigénio no corpo humano. Utilizando ferramentas e bases de dados bioinformáticas, explorámos várias informações sobre o gene HBB, incluindo a sua localização cromossómica, número de exões, sequência nucleotídica e outras espécies que possuem esse gene. Além disso, realizámos comparações de sequências entre o gene HBB humano e proteínas homólogas em diferentes espécies, utilizando métodos de alinhamento e análise filogenética. Os resultados destacam a conservação evolutiva do gene HBB e fornecem insights sobre a sua função molecular e importância biológica. Este estudo exemplifica a aplicação eficaz de ferramentas bioinformáticas na compreensão e análise de genes essenciais para a saúde humana.