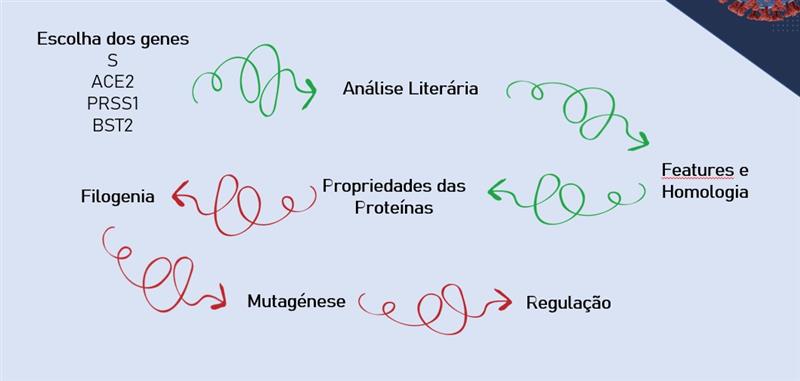
**Slide 1 – Apresentação**

Boa tarde, vamos apresentar o trabalho no âmbito da aplicação de ferramentas de bioinformática relacionado com o SARS-CoV-2.

**Slide 2 – Metodologia**

Para realizar este estudo escolhemos o vírus S e os genes …

****

**Slide 3 – Vírus S**

Os coronavírus (CoVs) são vírus de RNA (+) com envelope. Estes vírus entram nas células hospedeiras e, para iniciar a infeção, as suas membranas virais fundem-se com as membranas celulares do hospedeiro. A fusão da membrana é mediada pela glicoproteína transmembranar S do tipo I no invólucro viral e pelo recetor cognato na superfície das células hospedeiras. Para além de admitir a fusão da membrana, a superfície onde está a glicoproteína S é um alvo direto para as respostas imunes do hospedeiro, tornando-a o principal alvo dos anticorpos. Esta é capaz de regular negativamente o BST2 por degradação lisossomal, neutralizando a sua atividade antiviral.

A glicoproteína divide-se em duas subunidades: a proteína S1 e a proteína S2. A proteína S1 é responsável por incorporar o vírus na membrana celular quando este interage com o hospedeiro, ligando-se ao recetor ACE2, onde se inicia a infeção. A proteína S2 medeia a fusão do vírus com as membranas celulares do hospedeiro, agindo como uma proteína de fusão viral.

Esta proteína é o alvo principal para projetos de vacinas, como também para a terapêutica antiviral, devido aos seus papéis na infeção viral e na eliciação de respostas imunes humorais (transmitidas pelo sangue e pelo soro), mediadas por células nos hospedeiros durante a infeção.

**Slide 4 – ACE2**

O gene ACE2 é um gene codificador de proteína. A expressão específica desse gene desempenha um papel na regulação cardiovascular e renal do organismo e na fertilidade do mesmo. A proteína tem uma função crítica no volume sanguíneo, na resistência vascular sistémica e na homeostase cardiovascular. É ainda responsável por converter a Angiotensina 1 em Angiotensina 1-9 e a Angiotensina 2 em Angiotensina 1-7. Além disso, a proteína codificada por este gene é um recetor funcional da Proteína S1 do coronavírus humano e da Síndrome Respiratória Aguda Grave Humana Coronavírus, também conhecida como SARS-CoV e SARS-CoV-2. O ACE2 é, ainda, alvo no estudo de drogas para a COVID-19.

A expressão deste gene é específica de cada indivíduo. Pode expressar-se em células endoteliais de pequenas e grandes artérias, em células do músculo liso arterial (ao nível da proteína), em enterócitos do intestino delgado, do túbulo proximal renal e do coração (ao nível da proteína). No pulmão, expressa-se em baixos níveis nalgumas células alveolares Tipo2, como é o caso do TMPRSS2.

**Slide 5 – PRSS1**

O PRSS1 (Serine Protease 1) é um gene codificador de proteína. Este é responsável pela codificação de um tripsinogénio, que é um membro da família das serinas protéases da tripsina. Esta enzima é excretada pelo pâncreas e clivada, na sua forma ativa, no intestino delgado, sendo esta importante no processo digestivo. É ativo em ligações peptídicas envolvendo o grupo carboxilo da lisina ou da arginina. As mutações neste gene estão associadas à pancreatite hereditário. A proteína codificada por este gene tem atividade contra substratos sintéticos e é necessária para a clivagem no local S1 e S2 da proteína S permitindo assim a sua entrada nas células epiteliais dos brônquios.

**Slide 6 – BST2**

O BST2 (Bone Marrow Stromal Cell Antigen 2) é um gene de codificação de proteína. Tem como função o fator de restrição antiviral do hospedeiro induzido pelos IFN (Interferões), que bloqueiam eficientemente a libertação de diversos vírus com invólucro dos mamíferos, ao agarrar diretamente os viriões às membranas das células infetadas. Atua como uma espécie de “amarra” física direta, mantendo os viriões na membrana celular, ligando-se uns aos outros. Contém as isoformas 1 e a 2, em que a isoforma 1 atua como um ativador do NF-kappa-B e a isoforma 2 inibe esta atividade. Ambas as isoformas são fatores de restrição viral eficazes, no entanto têm antivirais e sinalizações diferentes, sendo que a isoforma 2 é resistente à degradação pelo vírus da imunodeficiência humana tipo 1 (HIV1) mediada pelo Vpu e restringe o budding na presença do Vpu.

A proteína BST2 pode ser regulada negativamente pela proteína S do SARS-CoV através da via da degradação lisossomal.

**Slide 7 – Análise das sequências e das features presentes no NCBI**

Após a análise literária, procedemos, então, à análise das features e dos qualifiers dos nossos genes de interesse. O código apresentado recebeu um ficheiro GenBank do gene de interesse da qual foram extraídos os records e selecionadas as informações acerca dos genes de interesse, bem como, informação referente às features e qualifiers.

**Slide 8 – Tabela**

Neste slide está representada alguma informação obtida do código anteriormente referido.

Gene S – Após a análise das features, obtivemos informação acerca do ID, da descrição (é um SARS coronavírus de genoma completo), do tamanho da sequência e da sua taxonomia. Através da análise dos qualifiers, obtivemos o sinónimo do gene (E2).

ACE2 – Após a análise dos features, conseguimos tirar informação sobre o ID, o tamanho da proteína e da sequência, entre outros tipos de informações características da proteína. Nos qualifiers, obtemos informação de interesse, como a sig\_peptide, os mat\_peptides e as misc\_features. Com os sig\_peptides e mat\_peptides conseguimos ver que o gene ACE2, quando é traduzido, vai ser constituído por um peptídeo de sinalização situado na posição 1-17 da nossa proteína. Obtivemos, também, 2 mat\_peptides, que são as chains que podem derivar de diferentes tipos de modificações pós-tradução, sendo um o ACE2 (com os aminoácidos na posição 18-805), e um Processed ACE2 (com os aminoácidos na posição 18-708). Este é obtido numa modificação pós-tradução da nossa proteína, e é sintetizada pela kinase, realizada pelo ADAM17, ou por serine proteases, como a TMPRSS2, TMPRSS11D e HPN/TMPRSS1. Obtivemos, ainda, várias misc\_features, que são features com regiões de interesse biológico que não podem ser descritas por nenhum outro recurso chave (novo ou raro). Nestas misc\_features obtivemos alguns artigos de interesse, onde foram retirados vários pontos essenciais sobre os genes em estudo. Numa destas misc\_features observamos que, se efetuassem a alteração da lisina na posição 32 e da tirosina na posição 41, podíamos detetar que se reduzia significantemente a associação com a proteína S1 da COVID-19. Se também se alterasse a lisina na posição 353, o ácido aspártico na posição 355 e a arginina na posição 357, era possível detetar uma redução significativa da associação do Sars-Cov-2, o que indica que a proteína S1 codificada pelo gene S do Vírus se pode ligar nestas posições do ACE2. Sabemos ainda que a infecciosidade não é alterada na presença de um inibidor que altera dramaticamente a forma do ACE2 e que a presença de TMPRSS2 aumenta a infecciosidade do Sars-Cov-2 a partir da kinase realizada por esta no ACE2.

PRSS1 – Nomeadamente aos features do nosso PRSS1 conseguimos tirar informação sobre o ID, o tamanho da proteína, o tamanho da sequência, e todos os tipos de informações de caracterização da nossa proteína. Na secção dos qualifiers verificamos algumas informações adicionais de relevo, como a existência de um péptido de sinalização e 3 chains (mat\_peptide) que são nomeadamente a Tripsina-1 e duas cadeias alfa. Também dão informações acerca das modificações pós-tradução, sendo que uma destas poderá ser a formação de uma cadeia apenas (tripsina-1) ou a formação de duas cadeias(alfa) produzidas após clivagem proteolítica e que a sulfatação em Tyr-154 aumenta a seletividade facilitando assim a digestão de proteínas dietéticas.

BST2 – Das features sobre o BST2 obtivemos o ID, a descrição, a taxonomia e o tamanho da sequência. Na sua descrição vemos que este gene se encontra no cromossoma 19 do Homo sapiens, sendo predominantemente expresso no fígado, pulmão, coração, placenta, nas células T, monócitos, células dendríticas, células NK e durante o desenvolvimento das células B e tendo baixos níveis no pâncreas, no rim, músculo esquelético e cérebro. Através dos qualifiers …

**Slide 9 – Blast**

Posteriormente à análise das features e qualifiers, que nos permitiram adquirir informação adicional, procedemos à execução do Blast para os genes e proteínas de interesse, de forma a encontrar genes e/ou proteínas homólogas. O código referente ao Blast dos genes de interesse recebeu um ficheiro GenBank destes genes e, após efetuado, foi criado um ficheiro com o resultado deste Blast. Este ficheiro é, posteriormente, filtrado de forma a listar os melhores resultados. O código referente ao Blast das proteínas, recebeu apenas o ID proveniente da UniProt da proteína de interesse, e foi efetuado o Blast, criando um ficheiro com o resultado do mesmo e, posteriormente, foi filtrado à semelhança do Blast anterior.

**Slide 10 – Tabela**

Neste slide estão representados alguns exemplos de resultados dos nossos Blasts. No caso das proteínas, o seu ID, nome, organismo, função molecular e estado, e no caso dos genes, o seu ID e organismo a que pertence.

S – Ao analisar o Blast do gene S, verificamos que os seus genes homólogos eram todos do organismo SARS coronavírus Tor2. No Blast das proteínas, verificamos que estas são todas homólogas do organismo SARS-CoV-2.

ACE2 – Ao analisar o Blast, detetamos que os genes homólogos do ACE2 eram todos do ser humano ou de primatas como o *Pan panisgus* e o *Gorilla Gorilla Gorilla*. Verificamos ainda que existem proteínas homólogas noutros mamíferos que tenham a mesma função que o ACE2, entre as quais algumas funções moleculares como metal ion binding, metallopeptidadse activity e peptidyl-dipeptidase activity.

PRSS1 – Após a analise do Blast para o gene do PRSS1 verificamos que os genes homólogos estão presentes em alguns primatas como por exemplo a *macaca mulata*,e no Blast para as proteínas verificamos que existem proteínas homólogas noutros mamíferos apresentando a mesma função de atividade de endopeptidase do tipo serina.

BST2 – Ao realizar o Blast do BST2, detetamos que os seus genes Homólogos também eram todos do ser humano ou de primatas do velho e novo mundo como *Pongo pygmaeus* ou *Gorilla Gorilla*. Ao fazer o Blast das proteínas Homólogas vimos que a semelhança entre eles é a função de resposta antiviral.

**Slide 11 – Ferramentas para a análise das propriedades das proteínas – Vírus S**

Para a análise das propriedades das proteínas utilizamos ferramentas como PDB, Phobius, NetNGlyc, LocTree, NetPhos e Virus-mploc.

A partir do Phobios, é possível verificar que a proteína S contém apenas um domínio transmembranar.

Utilizando o software Virus-mploc, prevemos que a localização do vírus S será no retículo endoplasmático do hospedeiro.

**Slide 12 e 13 – Ferramentas para a análise das propriedades das proteínas – ACE2**

A partir do Phobius podemos ver que esta proteína só contém 1 domínio transmembranar.

Utilizando o software LocTree3, que prevê a localização subcelular de proteínas, inferiu-se que a proteína é secretada, e que se encontrada agregado aos termos GO: espaço extracelular e membrana.

Utilizando o software NetNGlyc, podemos ver o número de N-glicosilações, verificou-se a existência de 6 N glicosilações nas posições 53, 90, 103, 322, 432, 690, sendo que na posição 90 uma glicosilação que diminui a infecciosidade da Proteína S.

Utilizando o software NetPhos, podemos detetar o número de Fosforilações que podes ocorrer na nossa proteína, e podemos ver pelo nosso gráfico um número elevado de fosforilações que podem ocorrer na nossa proteína ACE2.

**Slide 14 e 15 – Ferramentas para a análise das propriedades das proteínas – PRSS1**

Na análise às proteínas verificámos, através do Phobius, que esta não possui domínios transmembranares.

Através do software LocTree3 verificamos a localização da proteína em estudo, neste caso a PRSS1 é secretada no espaço extracelular. Nomeadamente às glicosilações verificamos que estas são inexistentes no PRSS1.

No gráfico seguinte obtido através do NetPhos podemos verificar um elevado número de fosforilações.

**Slide 16 e 17 – Ferramentas para a análise das propriedades das proteínas – BST2**

A partir do Phobius podemos ver que o gráfico gerado nos mostra que a proteína tem apenas 1 domínio transmembranar.

Utilizando o software LocTree3, observamos que a proteína BST2 se localiza na membrana plasmática.

Utilizando o software NetNGlyc, verificou-se a existência de 2 N-glicosilações nas posições 65 e 92.

Utilizando o software NetPhos, podemos observar através do gráfio que nos mostra o número elevado de fosforilações que pode ocorrer na proteína BST2. No entanto, podemos observar que não demonstra tantas fosforilações como a proteína ACE2 e apresenta mais fosforilações que a PRSS1.

**Slide 18 – Previsões futuras**

Futuramente e de forma a completar a Timeline apresentada anteriormente, iremos estudar o alinhamento múltiplo e a elaborar árvores filogenéticas, de modo a conseguirmos analisar zonas de conservação e comparar organismos diferentes para os genes estudados. Pretendemos ainda identificar interações regulatórias e de sinalização para compreender de que forma estes genes atuam e como são regulados por estas proteínas. Por fim, comparar as variantes mutagénicas e identificar o seu impacto biológico.