

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS BIOLÓGICAS  
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS

Ohanez Mamigonian

Seleção de neoepítomos como base para desenvolvimento de uma vacina contra câncer de  
mama triplo-negativo

Florianópolis, Santa Catarina

2023

Ohanez Mamigonian

**Seleção de neopítopos como base para desenvolvimento de uma vacina contra câncer de  
mama triplo-negativo**

Trabalho de Conclusão de Curso de bacharelado em  
Ciências Biológicas pela Universidade Federal de Santa  
Catarina

Orientador: Prof. Dr. Edroaldo Lummertz da Rocha

Florianópolis, Santa Catarina

2023

Mamigonian, Ohanez

Seleção de neoepítomos como base para desenvolvimento de uma vacina contra câncer de mama triplo-negativo / Ohanez Mamigonian ; orientador, Edroaldo Lummertz da Rocha, 2023.

36 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Graduação em Ciências Biológicas, Florianópolis, 2023.

Inclui referências.

1. Ciências Biológicas. 2. Câncer. 3. Imunoterapia. 4. Medicina de Precisão. 5. Bioinformática. 6. Neoantígenos. I. Rocha, Edroaldo Lummertz da. II. Universidade Federal de Santa Catarina. Graduação em Ciências Biológicas. III. Título.

Ohanez Mamigonian

**Seleção de neoepítomos como base para desenvolvimento de uma vacina contra câncer de mama triplo-negativo**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel e **APROVADO** em sua forma final pelo Curso de Ciências Biológicas

Florianópolis, 7 de junho de 2023

---

Coordenação do Curso

**Banca Examinadora:**

---

Prof. Edroaldo Lummertz da Rocha, Dr.

Orientador

UFSC/CCB/MIP

---

Prof. Alfeu Zanotto Filho, Dr.

UFSC/CCB/MIP

---

Prof. Daniel Santos Mansur, Dr.

UFSC/CCB/FMC

Florianópolis, 2023

Dedico este trabalho à minha avó,  
Obrigado por todo carinho, e por ainda visitar-me nos sonhos

## AGRADECIMENTOS

Gostaria de agradecer ao meu orientador, Prof. Dr. Edroaldo Lummertz da Rocha, pela atenção nas inúmeras vezes que passei em sua sala para discutir sobre o andamento do projeto, e outras ideias loucas da minha cabeça. Fui sempre bem recebido, e sinto que estes momentos todos agregaram não só ao meu trabalho, mas a mim como pessoa. Fico muito feliz pelas experiências que tive, e anseio pelo que o futuro resguarda.

Nada do que fiz teria sido possível sem o apoio da minha família. Não tenho como agradecer o suficiente pelo apoio oferecido pela minha mãe e minha tia, pessoas muito carinhosas e especiais, que tenho o privilégio de ter na minha vida. Foram imprescindíveis também na minha escolha por trilhar o caminho da ciência. Duas mulheres muito dedicadas, trabalhadoras e amáveis, sei que orgulharam minha avó.

Todos buscamos encontrar alguém com quem sabemos que podemos entregar nossos corações. Fico muito feliz de ter passado estes últimos anos ao lado de alguém tão especial quanto minha *люблюмая, золотая*, Helena Caio. Obrigado por me aguentar estes anos todos, nos mais sensíveis momentos. És uma das pessoas que mais admiro por toda tua dedicação e paixão, obrigado por me ajudar a aproveitar mais a vida, e adocicá-la.

Agradeço ao meu pai pelos bons momentos que passamos juntos. Tive o imensurável privilégio de sempre ter ao meu lado um professor que me instigasse a conhecer coisas novas desde minha tenra idade. Mesmo nas adversidades e atritos, sei que guardarei algumas memórias com carinho.

Seria uma tarefa inexecutável nominar todos aqueles que foram importantes para mim nestes anos. Sou grato a todos os meus amigos pelo companheirismo e apoio, e aguardo com carinho pelos momentos que compartilharemos juntos. Agradeço aos meus colegas de laboratório por todas as conversas que tivemos, por mais breves que fossem, sempre alegraram meu dia. Agradeço a todos os professores e professoras que compartilharam seus saberes e participaram da minha formação.

Por fim, agradeço aos professores que compõem a presente banca, pelo tempo e conhecimentos. Sei que seus apontamentos contribuirão para o aperfeiçoamento deste e elaboração de futuros trabalhos.

## RESUMO

Estima-se que, para o triênio de 2023 a 2025, aproximadamente 73 mil brasileiras sejam diagnosticadas anualmente com câncer de mama, correspondendo a aproximadamente 10% de todos os novos casos de neoplasias. Embora ainda clinicamente subutilizado no Brasil, o sequenciamento de tumores malignos viabiliza a escolha de tratamentos que melhor guiem pacientes à recuperação. Imunoterapias, embora atualmente limitadas a alguns tipos de cânceres, possuem seus horizontes ampliados com o advento de novas técnicas. Mutações e alterações cromossômicas afetam a funcionalidade de proteínas, promovendo oncogenicidade, gerando, em contrapartida, neoantígenos, que contém o potencial de eliciar resposta imune anti-tumoral. Ainda é difícil inferir quais mutações seriam de fato imunogênicas, incitando resposta potente e específica contra células tumorais. Por consequência, desenvolver novas estratégias que potencializem a predição de neoantígenos a partir de dados genômicos mostra-se como peça chave para revolucionar as práticas clínicas. A partir de um montante superior a 15 mil candidatos, selecionamos, por meio de análises *in silico*, 16 peptídeos, oriundos do sequenciamento de tumores da linhagem 4T1 de camundongos BALB/c, com alto potencial de agir como neoepítomos. Duas metodologias diferentes foram implementadas, a primeira visando a agretopicidade entre os peptídeos mutados e pares próprios, e a segunda buscando peptídeos longos com curtos embutidos, visando apresentação dinâmica entre células T CD4<sup>+</sup> e CD8<sup>+</sup>. A incorporação de neoepítomos na elaboração de estratégias terapêuticas inovadoras, como na forma de vacinas, é um passo rumo à medicina de precisão e revolução das práticas clínicas contemporâneas.

Palavras-chave: Câncer; Imunoterapia; Medicina de precisão; Bioinformática; Neoantígenos;

## ABSTRACT

It is estimated that, for the three-year period from 2023 to 2025, approximately 73,000 Brazilian women are to be diagnosed annually with breast cancer, accounting for approximately 10% of all new cases of neoplasms. Although still clinically underused in Brazil, the sequencing of malignant tumors enables the selection of treatments that will best guide patients to recovery. Immunotherapies, although currently limited to some types of cancers, have their horizons broadened with the emergence of new techniques. Mutations and chromosomal alterations alter the functionality of proteins, promoting oncogenicity, while generating in turn neoantigens, which have the potential to elicit an anti-tumor immune response. It is still difficult to determine which mutations would actually be immunogenic, triggering a potent and specific response against tumor cells. Therefore, developing new strategies that enhance the prediction of neoantigens from genomic data is a key to revolutionize current clinical practice. Through *in silico* analysis of over 15,000 candidates, derived from sequencing of the 4T1 tumor cell-line model, we selected 16 peptides with strong neopeptide potential. Two different methodologies were implemented, the first targeting agretopicity between mutated peptides and pairwise self-ones, while the second sought after long peptides with embedded short ones, aiming for dynamic presentation between CD4<sup>+</sup> and CD8<sup>+</sup> T cells. The integration of neopeptides into the design of novel therapeutic approaches, such as vaccines, represents a step towards precision medicine and the revolution of current clinical practices.

Keywords: Cancer; Immunotherapy; Precision Medicine; Bioinformatics; Neoantigens;

## Sumário

<b>INTRODUÇÃO.....</b>	<b>8</b>
<b>OBJETIVOS.....</b>	<b>11</b>
Objetivo geral.....	11
Objetivos específicos.....	11
<b>MATERIAL E MÉTODOS.....</b>	<b>11</b>
1. Origem de dados.....	11
2. Softwares usados.....	12
3. Especificações técnicas das máquinas utilizadas.....	12
<b>RESULTADOS.....</b>	<b>12</b>
<b>1. BUSCA E CAPTAÇÃO DE DADOS PARA PREDIÇÃO DE NEOEPÍTOPOS....</b>	<b>12</b>
<b>2. SELEÇÃO DE SOFTWARES.....</b>	<b>13</b>
<b>3. LIMPEZA DE DADOS.....</b>	<b>14</b>
<b>4. PREPARAÇÃO DE INPUTS AOS SOFTWARES DE PREDIÇÃO.....</b>	<b>15</b>
<b>5. SELEÇÃO DE PEPTÍDEOS.....</b>	<b>19</b>
<b>5.1. SELEÇÃO DE NEOEPÍTOPOS.....</b>	<b>19</b>
<b>5.1.1. SELEÇÃO POR AGRETOPICIDADE.....</b>	<b>19</b>
<b>5.1.2. SELEÇÃO POR INTEGRAÇÃO DE SEQUÊNCIAS.....</b>	<b>24</b>
<b>5.2 SELEÇÃO DE CONTROLES POSITIVOS E TESTES DE PERFORMANCE             DOS SOFTWARES DE PREDIÇÃO.....</b>	<b>25</b>
<b>DISCUSSÃO.....</b>	<b>27</b>
<b>CONCLUSÃO.....</b>	<b>31</b>
<b>REFERÊNCIAS.....</b>	<b>31</b>

## INTRODUÇÃO

Cânceres são doenças de natureza celular, em que as células do corpo passam a se replicar de forma incontrolada, e se espalham para outras regiões do corpo.<sup>[1]</sup> Estima-se que para o ano de 2020, no Brasil, entre 225 e 250 mil pessoas perderam a vida em função da doença,<sup>[2,3]</sup> sendo o número de novos casos para o período quase três vezes maior.<sup>[4]</sup> Estimativas mostram que, para os anos de 2023 a 2025, teremos algo próximo a 73 mil brasileiras sendo diagnosticadas anualmente com câncer de mama, o que corresponde a pouco mais de 10% de todos os novos casos de neoplasias previstas para o período.<sup>[4]</sup> Embora taxas de sobrevivência de 5 anos para cânceres de mama sejam relativamente altas, próximas de 90% em países como Estados Unidos<sup>[5]</sup> e Inglaterra,<sup>[6]</sup> a mesma taxa com foco em tumores triplo-negativos cai para 77,6%.<sup>[5]</sup> Cânceres de mama do subtipo triplo-negativo correspondem a aproximadamente 10-20% dos casos.<sup>[7-9]</sup> Embora não tenha sido encontrada uma estimativa concreta, uma amostragem do Rio Grande do Sul relatou uma prevalência de 14,6% deste subtipo.<sup>[9]</sup> Tratamentos hormonais, que são comumente empregados para cânceres de mama,<sup>[10]</sup> não funcionam para o subtipo triplo-negativo, devido à ausência dos receptores de estrogênio, progesterona e *Human Epidermal growth factor Receptor 2* (HER2),<sup>[11]</sup> culminando na queda dos índices de sobrevivência relatados.

A maioria dos cânceres são essencialmente células do corpo que acumularam mutações, assim adquirindo alta capacidade proliferativa, imortalidade replicativa, evasão imune e, eventualmente, invasibilidade, migração e colonização.<sup>[12-14]</sup> Estas capacidades são promovidas por anomalias genéticas e epigenéticas.<sup>[14-16]</sup> Quando mutações ocorrem sobre genes codificantes, geram proteínas defeituosas, proporcionando, por exemplo, ganho de função oncogênica ou perda de supressão tumoral. O tratamento das neoplasias é classicamente alicerçado em três formas de intervenção: cirurgia, radioterapia e quimioterapia.<sup>[10,17]</sup> Embora prometida com ares romantizados há quase 20 anos atrás,<sup>[18,19]</sup> as imunoterapias, a passos vagarosos, vêm evoluindo e se demonstrando como abordagens terapêuticas promissoras.<sup>[20-31]</sup>

A imunoterapia é um tipo de tratamento fundamentado no emprego do sistema imunológico ao combate de várias infecções e doenças, como o câncer.<sup>[20,32-35]</sup> Diferentes estratégias constituem esse tipo de intervenção, englobando desde terapias celulares,<sup>[27,28,33]</sup> uso de anticorpos inibidores de *checkpoints* imunológicos,<sup>[21,29-31,36,37]</sup> dentre outros.<sup>[20,38]</sup> Uma barreira que dificulta o uso destas terapias guiadas na oncologia é a ausência de marcadores claros que diferenciem células cancerígenas de saudáveis. Embora tenham sido registrados

avanços notáveis, principalmente no caso de leucemias,<sup>[39,40]</sup> este tipo de tratamento ainda avança a passos mais lentos em tumores sólidos.<sup>[41,42]</sup>

Mutações e alterações cromossômicas em cânceres levam à formação de proteínas defeituosas, estas são capazes de funcionar como neoantígenos.<sup>[43-45]</sup> Por meio de vias de processamento e apresentação de peptídeos, fragmentos de antígenos dão origem a epítomos próprios, que são apresentados na superfície celular para células do sistema imune. No caso de células cancerígenas, os neoantígenos, por meio das mesmas vias, produzirão neoepítomos. Estes novos antígenos e epítomos tumorais, por não estarem presentes durante o desenvolvimento dos linfócitos nos órgãos geradores, podem ser capazes de induzir uma resposta imune antitumoral.<sup>[46,47]</sup> A identificação destes, porém, pode ser um tanto desafiadora. Não se sabe ao certo quais mutações seriam imunogênicas, ou seja, quais realmente induzirão uma resposta imune potente e específica o suficiente para combater as células tumorais.<sup>[46-49]</sup> Por consequência, o desenvolvimento de métodos para predição de neoepítomos é peça estratégica na viabilização de formas inovadoras de tratamento oncológico.

A demanda por métodos capazes de selecionar neoepítomos já resulta na corrente oferta de diversos fluxos de trabalho que tentam suprir o anseio por explorar esse potencial nicho.<sup>[50,51,52]</sup> Estes *pipelines* usam ferramentas construídas para fazer predições acerca de alguma lacuna do processo de apresentação de antígenos, como processamento,<sup>[53]</sup> transporte<sup>[54]</sup> e, talvez mais importante, associação com moléculas MHC (*Major histocompatibility complex*). Há alguns *softwares* treinados para definir quais peptídeos são mais prováveis de se ligarem às proteínas MHC,<sup>[55,56]</sup> com bons resultados já sendo apresentados para MHC-I,<sup>[55,56]</sup> mas as predições voltadas aos alelos de MHC-II ainda não têm conseguido prover o mesmo desempenho.<sup>[55]</sup> Mesmo com as limitações atuais, estudos pioneiros já começam a demonstrar a competência de vacinas personalizadas em explorar o potencial dos neoepítomos, utilizando-se de predição *in silico* de neoantígenos, em ensaios clínicos.<sup>[22-26]</sup>

É neste cenário que o presente projeto se concentra, buscando identificar, priorizar e selecionar um conjunto de potenciais neoepítomos a partir de dados de sequenciamento do exoma da linhagem celular tumoral 4T1,<sup>[57]</sup> modelo murino de câncer de mama triplo-negativo. Visando incorporação em ensaios vacinais usando o modelo de camundongos BALB/c, foram selecionados neoepítomos como base para o desenvolvimento de uma plataforma vacinal peptídica, sendo empregados *softwares* já bem estabelecidos na literatura.<sup>[55]</sup> Fomentar-se-á, assim, posteriormente a confecção de novas estratégias vacinais para direcionar o sistema imune contra células constituintes do câncer.

## **OBJETIVOS**

### **Objetivo geral**

Selecionar possíveis neoepítomos para incorporação em vacinas antitumorais baseadas em peptídeos para o modelo de câncer de mama triplo-negativo

### **Objetivos específicos**

1. Proceder a procura e captação de dados genômicos;
2. Realizar o levantamento e seleção de *softwares* para a predição de neoantígenos;
3. Efetuar a limpeza de dados por meio do controle de qualidade das amostras;
4. Promover a geração de peptídeos para *input* nos *softwares* de predição;
5. Realizar a seleção de peptídeos candidatos;

## **MATERIAL E MÉTODOS**

Para cumprir com os objetivos do trabalho, foram utilizados dados públicos de sequenciamento de exoma do modelo murino BALB/c, assim como dados provenientes de células 4T1 implantadas em BALB/c. Foram utilizados alguns *softwares* para manipular e fazer predições sobre os dados. Também estão aqui descritas as especificações técnicas do equipamento computacional utilizado. Detalhes acerca de como cada etapa foi executada encontram-se na porção dos resultados, estratégia empregada visando um entendimento mais natural, trazendo a noção de como cada etapa dialogou uma com a outra.

### **1. Origem de dados**

Foram utilizados dados de chamada de variantes e uma tabela de peptídeos testados imunologicamente, disponibilizados pelo artigo de Schrörs e colaboradores.<sup>[57]</sup> Os dados tiveram de passar por algumas etapas de limpeza e padronização, processos estes descritos em maiores detalhes na porção de resultados.

Para a montagem de cada proteína utilizou-se também o arquivo FASTA anotado de BALB/c contendo todas as versões de proteínas do modelo murino, que pode ser encontrado via domínio do banco Ensembl.<sup>[58]</sup>

## 2. Softwares usados

Para a realização do trabalho foram gerados *scripts* em linguagem Python (v. 3.10.6)<sup>[59]</sup> e BASH (5.1.16(1)-release)<sup>[60]</sup>. Para as análises de ranqueamento de pares peptídeo com alelo MHC foram empregados os softwares NetMHCpan (v. 4.1)<sup>[55]</sup> e NetMHCIIpan (v. 4.0),<sup>[55]</sup> para predições com os alelos de camundongos para as classes de *Major Histocompatibility Complex* 1, e 2, respectivamente.

Houve um esforço para minimizar a utilização de ferramentas extras, que necessitariam instalação à parte, em favor das já disponíveis em instalações básicas do Sistema Operacional Linux. Usaram-se os softwares em linha de comando BASH: AWK (mawk 1.3.4 20200120), GREP (GNU grep 3.7), SED (GNU sed 4.8), CP (GNU coreutils 8.32), RM (GNU coreutils 8.32), CUT (GNU coreutils 8.32), GZIP (v. 1.10), TAIL (GNU coreutils 8.32), LIBREOFFICE (LibreOffice 7.3.7.2 30(Build:2)).

Nos *scripts* de Python, foram utilizadas as bibliotecas: pandas (v. 1.5.2), numpy (v. 1.24.1), biopython (1.80).

## 3. Especificações técnicas das máquinas utilizadas

Nesta última sessão, para fins de reprodutibilidade computacional, são descritos alguns detalhes técnicos do computador que foi usado para a elaboração do presente trabalho. Vale ressaltar que o ranqueamento dos peptídeos foi feito nas versões online de seus respectivos softwares,<sup>[55]</sup> portanto, não sendo possível trazer requerimentos ideais para utilização dessas ferramentas *in loco*.

Especificações de *hardware*: processador 12<sup>a</sup> Geração Intel® Core™ i7-12700K, 20 núcleos de processamento; 16 Gb de memória RAM; 2 Tb de armazenamento, arquivos usados totalizando em torno de 160 Mb.

Especificações de *software*: Linux Ubuntu 22.04.02 LTS como Sistema Operacional.

## RESULTADOS

### 1. BUSCA E CAPTAÇÃO DE DADOS PARA PREDIÇÃO DE NEOEPÍTOPOS

Foi feito um levantamento inicial de trabalhos publicados no banco de sequências de transcriptomas Gene Expression Omnibus (GEO),<sup>[61]</sup> procurando pela linhagem de camundongos que seria usada no projeto, “BALB/c”, e pela linhagem celular tumoral, “4T1”.

Os dados selecionados para o desenvolvimento deste projeto vieram do artigo “*Multi-Omics Characterization of the 4T1 Murine Mammary Gland Tumor Model*”, de Barbara Schrörs e colaboradores.<sup>[57]</sup> Este artigo foi selecionado pela natureza exploratória e complementar dos dados, sendo feito sequenciamento de DNA tanto do modelo murino BALB/c, quanto das células tumorais 4T1, que carregam as mutações somáticas, características da linhagem celular. Nele, além dos dados genômicos, foram também gerados dados de sequenciamento de RNA de 4T1, bem como análises imunológicas feitas com alguns peptídeos selecionados pelos autores. Estas últimas seriam especialmente interessantes, pois também foi empregada uma estratégia vacinal, fundamentando o racional do presente projeto. Os dados provindos destes experimentos prévios também deram base para avaliações de performance dos *softwares* usados, e desenho de peptídeos controle positivo.

Os dados de chamada de variantes (.VCF) seguiram por outras etapas de processamento, descritas nas porções 3-5 dos resultados, culminando, por fim, no resultado final do trabalho, os peptídeos selecionados. Já os dados dos peptídeos de imunogenicidade testada serviram para “desenhar” um limiar ideal ao qual os neoepítomos selecionados deveriam chegar, bem como a geração de peptídeos controle positivo, dos quais, imunogenicidade específica já havia sido averiguada pelos autores do artigo supracitado.

Além destes dados previamente citados, é relevante informar que para a montagem dos peptídeos a partir das mutações, foi necessário parear com as proteínas descritas do modelo BALB/c. Para tal, foi usado o arquivo de referência de proteínas anotadas específico desta linhagem. Este arquivo pode ser encontrado no domínio do banco de referências Ensembl.<sup>[58]</sup>

## 2. SELEÇÃO DE SOFTWARES

A primeira peça de literatura lida, para obter uma contemplação inicial do tema de neoantígenos, foi a publicação do artigo de opinião “*The problem with neoantigen prediction*”, na revista *Nature Biotechnology*.<sup>[48]</sup> Como relatado por ela, e outros trabalhos mais recentes reiteram,<sup>[62]</sup> muitos foram os esforços na elaboração de *softwares* que abordassem a problemática da predição de neoantígenos por diferentes flancos. Dentre eles: afinidade de peptídeos com MHC,<sup>[55,56]</sup> clonalidade,<sup>[63]</sup> processamento de epítomos,<sup>[53,54]</sup> dentre outros.<sup>[62]</sup>

Após avaliar o cenário, optou-se pela utilização do software NetMHCpan,<sup>[55]</sup> por um conjunto de fatores: seu impacto no meio científico, com até o momento, 40 patentes citando a versão 4.0 da rede treinada para MHC-I,<sup>[64]</sup> e aproximadamente 961 citações em outros

artigos até o presente momento,<sup>[65]</sup> as melhorias apresentadas nas versões 4.0 quanto à predição no contexto de MHC-I,<sup>[64]</sup> o fato do software NetMHCpan possuir um equivalente, mesmo que ainda com performance a desejar, capaz de predição para MHC-II, NetMHCIIpan,<sup>[55]</sup> Sendo relevante enfatizar o fato de que tanto a versão voltada a MHC-I quanto a II são capazes de predição em alelos de camundongos (*Mus musculus*), com os haplótipos presentes em BALB/c dentre os agraciados pelos modelos de predição, conforme tabela.<sup>[66]</sup>

### 3. LIMPEZA DE DADOS

A limpeza de dados está sendo citada aqui como uma etapa, contudo, deve-se esclarecer que a limpeza de dados não foi feita uma única vez, tendo papel crucial primeiramente após a captação dos dados (Resultados 1.), e também uma vez obtidas as predições oriundas dos *softwares* viabilizando a manipulação dos *outputs* rumo à seleção final (Resultados 5.).

Acerca da limpeza dos dados de chamada de variantes, captados do artigo,<sup>[57]</sup> foram necessárias algumas medidas de pré-processamento dos dados, garantindo sua coerência em contexto biológico. Além de filtrar pelas informações desejadas, como “gene”, “tipo de substituição de aminoácidos”, “transcrito”, etc, como era esperado, a primeira problemática se desenvolveu. Como queremos reconstruir um fragmento de peptídeo correspondendo a cada posição mutada, é de suma relevância que o aminoácido original, que deveria ser encontrado na posição do resíduo que está mutado, seja o mesmo encontrado na versão da proteína correspondente ao transcrito anotado. Caso contrário, pode-se montar erroneamente os fragmentos de peptídeo. Por exemplo, baseando-se em uma mutação “Q385L”, estariam inapropriadas proteínas oriundas de transcritos do gene que não possuam uma Glutamina (Q) no resíduo 385. Quaisquer fragmentos montados sobre essas proteínas, colocando uma Leucina (L) na posição 385, estariam fundamentalmente errados. Ou seja, fragmentos originados de pareamento inapropriado falhariam em representar o cenário biológico real, comprometendo toda a análise. Eis que o seguinte problema se desenvolve: a tabela VCF publicada concatenou todas as mutações encontradas, bem como seus respectivos transcritos, em apenas uma linha. Ou seja, a informação que parecia uma mutação residual com seu respectivo transcrito estava perdida. Mutações e transcritos estavam embaralhados, individualizados apenas pelo gene. A forma de anotação dos transcritos encontrava-se na padronização UCSC, dificultando o processo de encontrar proteínas correspondentes para

reconstruir os neoantígenos, visto que a formatação presente no banco de referências de sequências, Ensembl,<sup>[58]</sup> usa outro formato de identificação.

Para solucionar o problema supracitado, a abordagem utilizada foi: a partir da informação de cada mutação e gene, como o caso da Glutamina (Q385L), foram interrogados todos os transcritos descritos aos respectivos genes, em busca de quais proteínas seriam apropriadas para conter a mutação, de acordo com critérios a seguir. Para tal, primeiramente foram consideradas todas as proteínas oriundas de transcritos anotadas de um gene, para cada mutação, seguindo à seleção de apenas aqueles que corresponderam, na posição do resíduo mutado, ao aminoácido de referência. Por exemplo, todos os produtos de transcritos de referência que possuíam o aminoácido Glutamina na posição 385 entraram para a análise da mutação da Glutamina para Leucina neste resíduo, conforme anotação “Q385L”. Todos os que possuísem outros aminoácidos nessa posição seriam inválidos, e portanto, foram descartados, conforme pode ser observado na Figura 1, na segunda porção, “Limpeza de Dados”. Repetindo esse processo para cada mutação documentada, verificando a posição mutada contra todas as proteínas de transcritos descritas daquele gene, conseguiu-se aproveitar 253 das 429 mutações anotadas, uma proporção de ~59%. As demais foram perdidas, possivelmente pelo fato de que as mutações foram anotadas com base no genoma de referência de camundongo (*Mus musculus*), e não na versão específica relativa à linhagem BALB/c.<sup>[58]</sup> Com essa porção efetivamente mais útil dos dados em mãos, começou-se a prepará-los para lançamento aos softwares de predição.

#### **4. PREPARAÇÃO DE INPUTS AOS SOFTWARES DE PREDIÇÃO**

Feito o pareamento das 253 mutações com seus respectivos transcritos viáveis, chegamos então a aproximadamente 366 pares de código de mutações associados a proteínas oriundas de transcritos viáveis. Os softwares de predição de epítomos geram uma associação quantitativa relativa de um peptídeo a um alelo de molécula MHC. Basicamente o que precisamos fornecer aos softwares são as sequências de aminoácidos que devem ser interrogadas (na forma de um arquivo FASTA), e selecionar quais alelos MHC devem ser testados. Pode parecer que nesta etapa de geração seriam simplesmente montadas proteínas com as mutações presentes no arquivo VCF, e pronto. Esse olhar, no entanto, se encontraria longe da realidade. Sendo computacionalmente muito custoso, e logisticamente difícil trabalhar com os resultados diretos de uma entrada tão grande de dados. Foram sim montadas, para cada par proteína-mutação, versões de proteína que possuísem o resíduo mutado em sua

sequência. Porém, diversas medidas foram tomadas conjuntamente de forma a tornar o processamento mais eficiente, e as análises seguintes mais fáceis.

O software selecionado, NetMHCpan,<sup>[55]</sup> opera de tal forma que, dar como *input* as proteínas inteiras para cada mutação relatada, resultaria em um processamento muito lento, inviabilizando a conclusão da tarefa. Após alguns testes, foi verificado que o *software* funcionava fragmentando cada proteína em janelas (“*frames*”) de leitura, e fazendo a predição individualmente e sequencialmente, seguindo um padrão de janelas deslizantes (*sliding frames*). Desta forma, uma proteína sendo interrogada por epítomos de 8 aminoácidos de comprimento (“*8mers*”), será interrogada dos resíduos 1-8, em seguida de 2-9, seguindo este padrão até seu último resíduo. Essa observação nos proporciona um *insight* importante: dentro de uma proteína inteira haverá muitas janelas, sendo que apenas uma proporção pequena delas realmente nos interessarão, por incluir o resíduo mutado. Tendo isso em mente, após alguns novos testes, foi comprovado que as predições feitas não levam em conta o “contexto” completo da proteína, apenas a janela que o *software* “lê”. Ou seja, pressupõem-se que todos os fragmentos de uma proteína têm iguais chances de passar pelo processo de degradação de proteínas, e dar início no processo de apresentação de antígenos, independente de encontrar-se no começo, meio ou final de uma proteína de 300 ou mais de 2000 aminoácidos. Este aspecto permite que ao invés de usar a proteína mutada inteira, poderia ser usado apenas um fragmento pequeno que contenha o resíduo mutado. Assim, foram feitas várias funções em python que permitiram o desenho de fragmentos para cada mutação, em que todas as janelas de leitura incluíssem o resíduo mutado. Desta forma, a problemática do processamento conseguiu ser evitada, tornando o processo consideravelmente mais rápido. Agora as 366 “proteínas-mutação” seriam melhor lidas como 366 “fragmentos-mutação”.

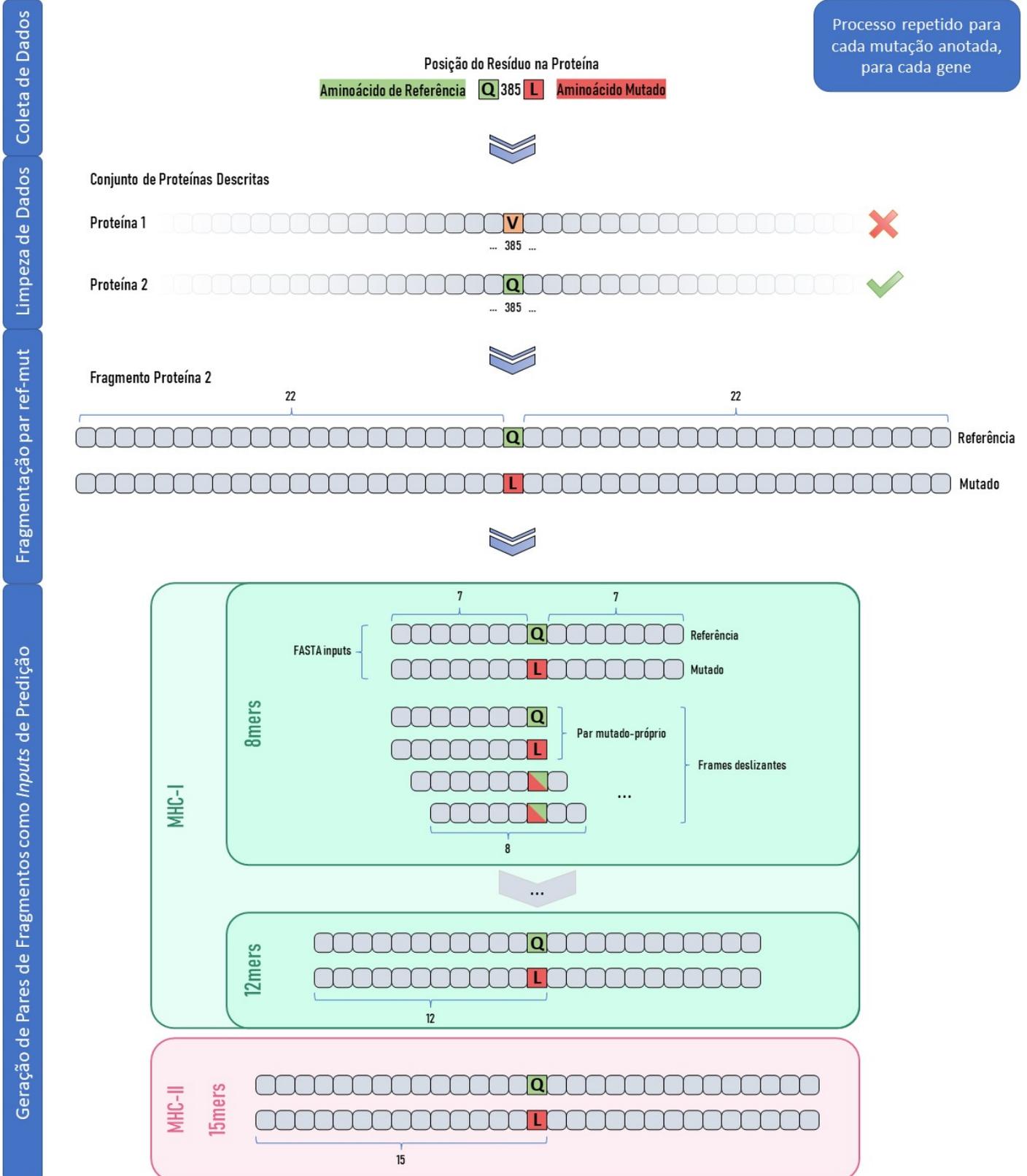
Quando falamos de neoepítomos, é importante termos em mente que: os peptídeos contendo a mutação só serão efetivamente imunogênicos se conseguirem superar a tolerância imunológica central. Um conceito que nos permite mensurar esse tópico de forma mais palpável é o de agretopacidade diferencial (*differential agretopicity*)<sup>[62]</sup>. Um peptídeo ao se ligar a uma molécula MHC possui um certo nível de afinidade, essa afinidade não será a mesma para outros alelos MHC, nem a outros peptídeos. Ou seja, podemos, ao interrogar a afinidade de um peptídeo mutado a um alelo de MHC, também realizar a predição para seu “par próprio”, que carrega o aminoácido de referência, e comparar os resultados para cada par de janelas. Isso permitirá, durante a próxima fase, eliminar os candidatos mutados que possuam seus pares próprios com níveis de ranqueamento também elevados aos alelos MHC.

Assim, os 366 fragmentos-mutação se multiplicaram para 732, cada peptídeo mutado com seu “par endógeno”, não mutado.

Tendo os 732 fragmentos, estes foram então adaptados aos comprimentos de sequência de interesse para gerar as predições. O programa NetMHCpan<sup>[55]</sup> aceita peptídeos de 8-14 aminoácidos de comprimento. Foi optado por gerar arquivos FASTA de 8-12 aminoácidos. Já o programa NetMHCIIpan<sup>[55]</sup> parece que funciona melhor com peptídeos de 15 aminoácidos, então foi gerado apenas um arquivo contendo os fragmentos neste comprimento. A geração dessas subsequências teve como objetivo que o resíduo sempre estivesse presente na janela de leitura do software de predição, seja de 8 em 8, ou 15 em 15 aminoácidos. Desta forma, todos os peptídeos mutados gerarão apenas janelas ou “frames” mutados, tornando a seleção dos candidatos mais direta e eficiente. Qualquer *frame* oriundo de fragmentos mutados não contendo a mutação, claramente falharia com seu propósito. Como nesta etapa multiplicou-se consideravelmente o número de sequências, efetivamente 6 vezes, foi importante “limpar” os dados novamente. Optou-se, então, por procurar e eliminar sequências que se repetissem dentro de cada arquivo FASTA, o que foi feito usando o *software* AWK via terminal. Essas repetições são resquícios do “re-pareamento” das mutações com cada uma das proteínas compatíveis de carregarem a mutação, processo descrito na etapa de limpeza dos dados. Isso se deu porque, por mais que várias proteínas descritas possam abrigar a mutação, os fragmentos delas que serão usados na análise nem sempre se diferenciam, visto que, às vezes, a diferença nas proteínas se dá em exons distantes, longe de onde o resíduo da mutação se encontra. Após essa “poda” dos fragmentos duplicados, obtivemos 2764 fragmentos, somando todos os arquivos FASTA. Uma subtração de aproximadamente 37% do volume de fragmentos total anterior à poda.

São apresentados também alguns dos detalhes a respeito das escolhas quanto aos parâmetros usados na execução dos *softwares*. Foram geradas predições para os atributos EL (*Eluted Ligant*) e BA (*Binding Affinity*). Estes *scores* mensuram a associação de peptídeos às moléculas de MHC por diferentes “vias”, conforme origem dos dados usados para treinar suas redes neurais. Os alelos selecionados para as análises foram: H-2-Dd, H-2-Kd e H-2-Ld para MHC-I; H-2-IAd e H-2-IEd para MHC-II. Estes alelos foram escolhidos conforme a haplotipagem de linhagens de camundongos, por serem os presentes em BALB/c, conforme tabela<sup>[66]</sup>. Um resumo do processo de coleta e tratamentos descritos até então encontra-se esquematicamente exposto na Figura 1, apresentada a seguir.

Figura 1 - Fluxo de análises e manipulação de dados até preparação de *inputs* aos *softwares*. Demonstração esquemática do fluxo de trabalho de coleta de dados até geração de peptídeos de múltiplos comprimentos para *input* nos *softwares* de predição. O processo foi repetido a cada mutação presente na anotação de variantes. Foram selecionados todos os variantes de proteínas viáveis de carregar a mutação, conforme ilustrado em “Limpeza de Dados”. Foram fragmentadas as proteínas em porções menores por uma questão computacional, conforme ilustrado em “Fragmentação por mut-ref”. Por fim, foram gerados os diferentes arquivos de *input*, de forma a direcionar o processamento do *software* às janelas deslizantes (*sliding frames*) de leitura que contenham sempre o resíduo mutado ou o aminoácido de referência. A inclusão dos pares próprios viabiliza mensurar a agretopicidade dos candidatos (peptídeos mutados), de forma a melhor garantir que estes últimos sejam vistos como não próprios (*non-self*).



Processo repetido para cada mutação anotada, para cada gene

Fonte: Imagem gerada pelo autor.

## 5. SELEÇÃO DE PEPTÍDEOS

O objetivo desta etapa foi selecionar os peptídeos candidatos a ingressar nos testes imunológicos *in vivo*. Para cumprir com este objetivo, duas linhas foram traçadas: na primeira delas buscou-se efetivamente pelos neoepítomos teoricamente mais capazes de eliciar resposta imune; já na outra linha, a prioridade era construir formas de validação das predições. Por exemplo, pela delimitação de *scores* atingidos por peptídeos validados como imunogênicos, e seleção de peptídeos que servirão para os testes *in vivo* como controles positivos.

Nesta etapa, após executar as predições, foram novamente necessárias algumas práticas de limpeza dos dados. O input dos dados nas plataformas online dos *softwares*<sup>[55]</sup> apresentaram problemas não antecipáveis, como a desconfiguração de *headers* dos arquivos FASTA, além de que a arquitetura dos dados de output estava também organizada de tal forma que dificultava qualquer manipulação e análise dos mesmos. Embora tenham configurado contratempos, todos os problemas foram solucionados.

### 5.1. SELEÇÃO DE NEOEPÍTOPOS

Após obter os valores de predição oriundos dos softwares NetMHCpan e NetMHCIIpan, os dados seguiram por uma série de análises diferentes. Duas foram as estratégias adotadas na procura pelos neoepítomos: uma baseando-se na agretopicidade, ou seja, na diferença entre peptídeos mutados e seus pares próprios; e a segunda, embasada na hipótese de que peptídeos imunogênicos longos, que se encaixam nos MHC-II, quando possuírem a sequência de curtos “embebidos” na sua, acabam viabilizando a apresentação desses neoepítomos de forma mais dinâmica para células T CD4+ e CD8+.

#### 5.1.1. SELEÇÃO POR AGRETOPICIDADE

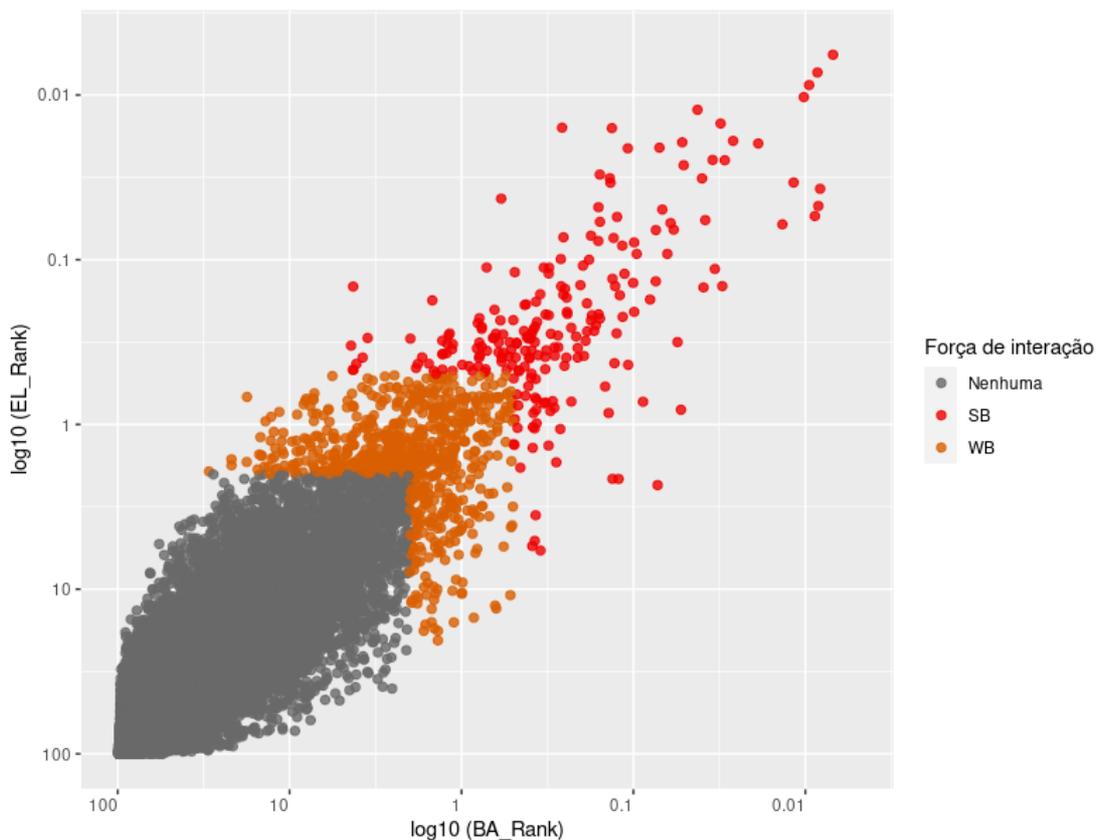
A partir da busca fundamentada na agretopicidade, obtiveram-se 8 peptídeos candidatos para MHC-I, e 4 para MHC-II. A saída direta dos softwares de predição resultou em mais de 32 mil candidatos, todos formados por sequências de aminoácidos contendo um resíduo mutado sendo associadas a algum dos 5 alelos de MHC testados, 3 alelos de MHC-I e 2 de MHC-II, seguindo os planos da etapa anterior. Para selecionar quais seriam bons candidatos foram inicialmente usados os limiares indicados pela própria equipe desenvolvedora dos *softwares* preditores. Segundo os quais, deveriam ser considerados: SB (*Strong Binding*), ou “candidatos fortes”, aqueles peptídeos cujo ranqueamento de predição

estivesse abaixo de 0,5 para o NetMHCpan, e abaixo de 1 para NetMHCIIpan; WB (*Weak Binding*), ou “candidatos fracos”, aqueles cujo ranqueamento estivesse entre 2 e 0,5 para o NetMHCpan, e entre 5 e 1 para NetMHCIIpan. Esses limiares incidiriam sobre os valores de “EL\_Rank” e “BA\_Rank” de cada par peptídeo-alelo. Aqui passam a se diferenciar as abordagens para os peptídeos associados a MHC-I e MHC-II, tanto por possuírem limiares diferentes, quanto por questões de precisão que serão discutidos posteriormente.

Tratando dos candidatos para MHC-I, 223 possuíam a característica de “SB” pelo atributo “EL\_Rank”, “BA\_Rank” ou por ambos, conforme Figura 2. Para filtrar por peptídeos que satisfazem ambos os atributos, foi criado um novo atributo, derivado da média entre ambos, chamado “ELBA\_Rank”. Sobre este, então, foi aplicado um limiar no valor 0,5. Assim, dos 223 peptídeos “bons candidatos” iniciais, 148 passaram pelo novo limiar, como pode ser observado na Figura 3.

Figura 2 - Candidatos mutados iniciais

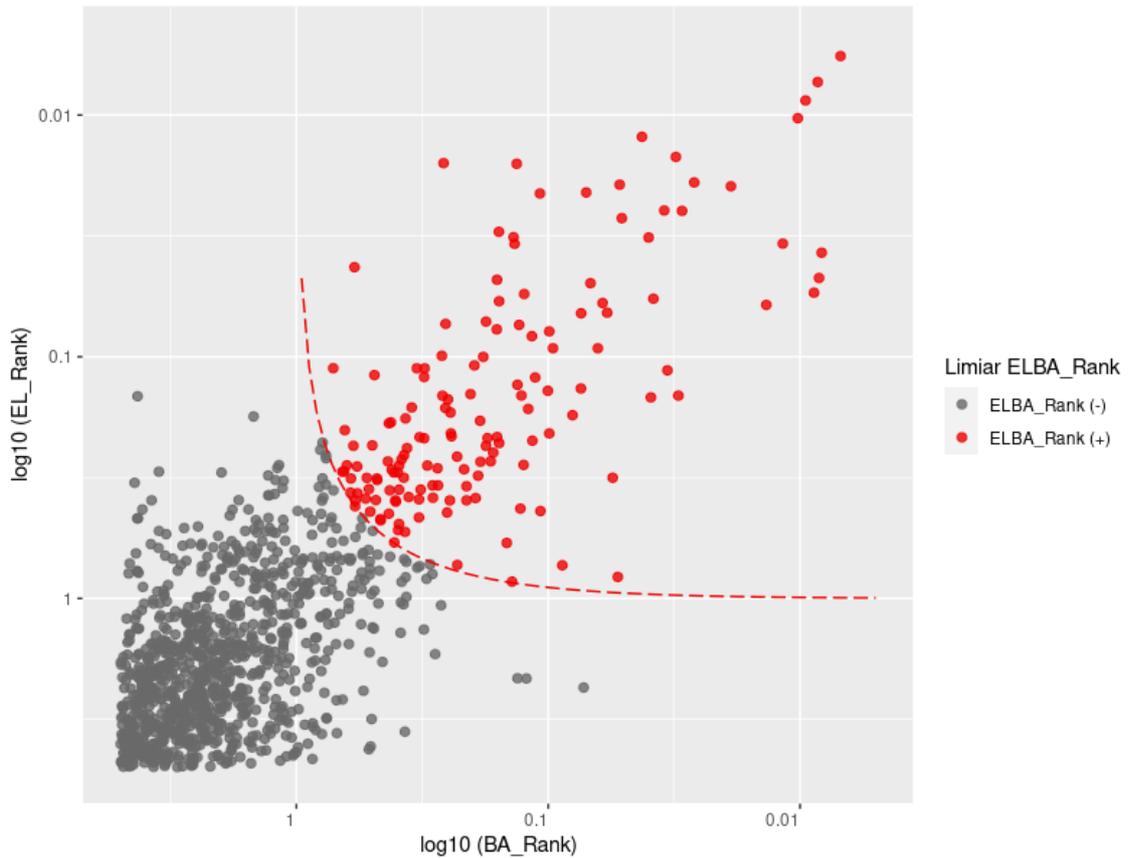
Gráfico de dispersão dos candidatos (peptídeos mutados) preditos para interação com MHC-I. Cada ponto representa um peptídeo em associação com um alelo MHC-I. Ranqueamento (100-0) do atributo BA (*Binding Affinity*) (eixo x), e ranqueamento (100-0) do atributo EL (*Eluted Ligand*) (eixo y). A escala é logarítmica, e invertida. Os candidatos são classificados em 3 categorias, com relação ao quão melhor se mostraram no ranqueamento: SB (*Strong Binding*), “ligação forte”, com ranqueamento entre os 0,5% melhores peptídeos associados com o dado alelo; WB (*Weak Binding*), “ligação fraca”, com ranqueamento entre os 2% melhores, mas inferior aos 0,5%; ou nenhuma, com ranqueamento inferior aos 2% melhores para cada alelo.



Fonte: Figura gerada pelo autor

Figura 3 - Limiar ELBA\_Rank

Demonstração do limiar imposto ao parâmetro “ELBA\_Rank”, obtido da média dos dois parâmetros preditos pelo *software* NetMHCpan. Com o dado limiar, candidatos que possuíam a média de seus ranqueamentos de EL e BA dentro os 0,5% melhores peptídeos, encontram-se em vermelho, e foram selecionados para as etapas seguintes de análise por agretopacidade, agindo como um filtro positivo.

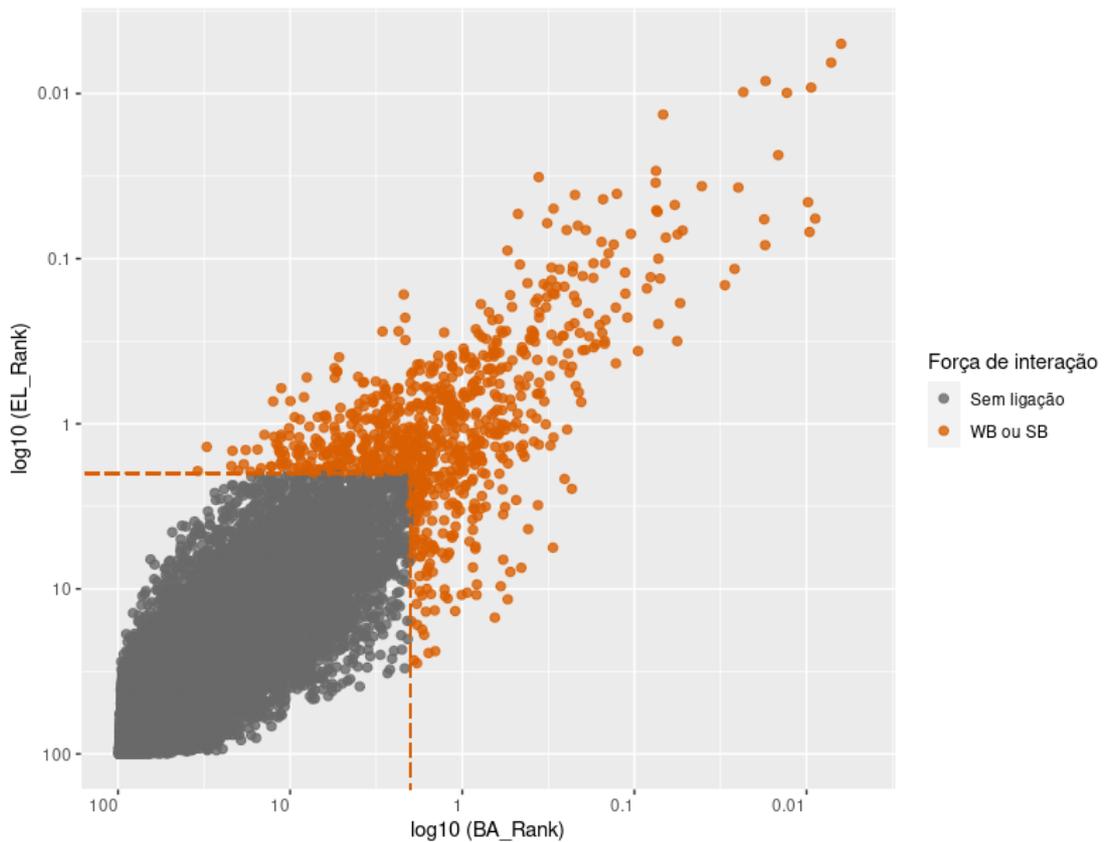


Fonte: Figura gerada pelo autor

O último passo foi fazer a seleção por agretopacidade. Para tal, usaram-se os peptídeos próprios, que também passaram pelas análises de predição às moléculas de MHC. Para avaliar a agretopacidade foi necessário acessar o ranqueamento dos pares próprios de cada um dos 148 candidatos que passaram o limiar ELBA\_Rank. Foi então verificado que os pares próprios destes candidatos não estivessem em limiares consideráveis plausíveis de interação com os alelos MHC. Isto é, não poderiam se encontrar nem no considerado “WB”, que denotaria uma ligação fraca, mas ainda uma ligação. Desta forma, foram eliminados candidatos mutados cujos pares próprios estivessem dentre os 853 que possuem alguma ligação, estes últimos visíveis em laranja, conforme a Figura 4. Por fim, selecionaram-se 8 dos 148 candidatos, nos quais a agretopacidade era alta, conforme a Figura 5.

Figura 4 - Ranqueamento dos pares próprios

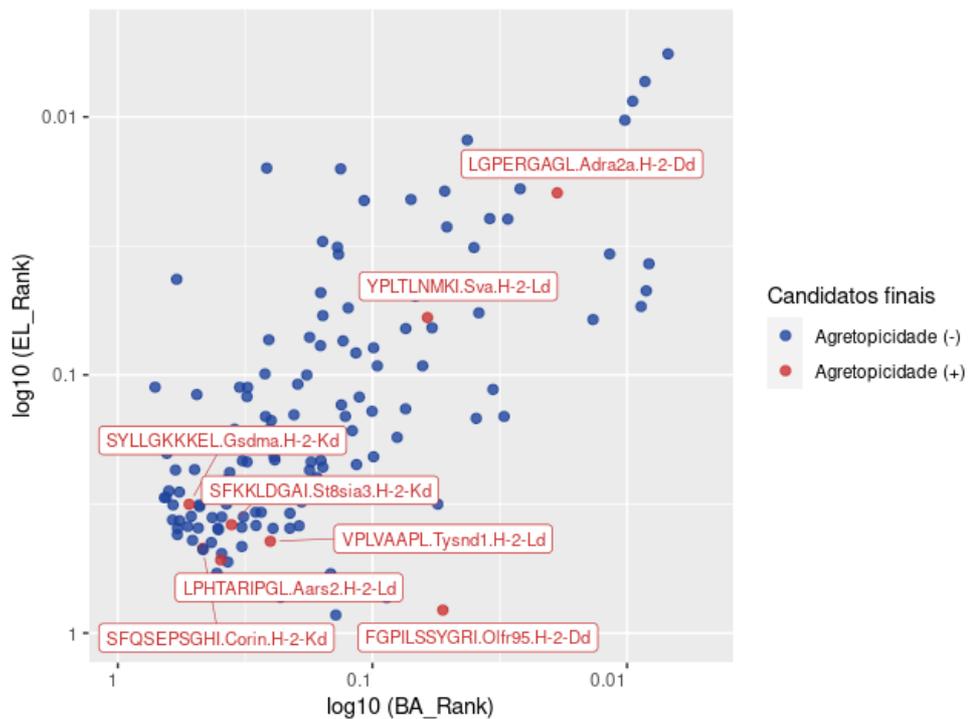
Demonstração dos pares de peptídeos próprios considerados dentro de limiares preocupantes de ranqueamentos, dentre os 2% melhores candidatos para algum dos dois atributos, EL ou BA. Os candidatos (peptídeos mutados) não poderiam apresentar seus pares próprios dentre os aqui representados em laranja, agindo como um filtro negativo.



Fonte: figura gerada pelo autor

Figura 5 - Neoepítópos candidatos por agretopacidade

Demonstração dos 8 candidatos finais selecionados pelo método de agretopacidade, referente aos alelos de MHC-I. Dos 148 candidatos inicialmente considerados pelo filtro de inclusão (Fig. 3), após aplicação do filtro negativo (Fig. 4), selecionaram-se ~5% deles, considerados de grau de agretopacidade alto.



Fonte: figura gerada pelo autor

Semelhante abordagem foi empregada para os peptídeos associados aos alelos de MHC-II, porém, com estes foi necessário deixar o filtro de agretopicidade um pouco menos criterioso. Isto se deu devido ao fato de que nenhum par mutado-próprio, cujos mutados estivessem acima do ranque 1 (limiar “SB” para MHC-II), possuía os seus pares de peptídeos próprios fora do limiar de ligação fraca. Deste modo, foram selecionados os candidatos de ranqueamento “SB”, que possuísem seus pares próprios fora do limiar de “SB”, tolerando, portanto, os próprios que caíssem na classificação “WB”. Outra peculiaridade da seleção para MHC-II é que optou-se por usar os resultados de EL\_Rank, e não o ranqueamento misto ELBA\_Rank, como havia sido feito no cenário de MHC-I. Essa escolha foi baseada na baixa performance de BA\_Rank do *software* NetMHCIIpan,<sup>[55]</sup> descrito em etapa subsequente, que tratará a respeito de testes feitos sobre uma amostragem de peptídeos cuja imunogenicidade fora previamente aferida. Detalhes acerca dos candidatos selecionados podem ser melhor observados nas Tabela 1, correspondendo aos candidatos por predição a MHC-I, e Tabela 2, para MHC-II.

Tabela 1 - Candidatos Curtos por Agretopicidade

Listagem dos 8 candidatos finais selecionados pelo método de agretopicidade, referente aos alelos de MHC-I. Trazendo informações acerca de seus atributos e características biológicas.

Gene	Aminoácido	Peptídeo	Alelo	EL_Rank	BA_Rank	ELBA_Rank	Agretopicidade
Tysnd1	A278P	VPLVAAPL	H-2-Ld	0,4410	0,2518	0,3464	0,0572
Adra2a	P276L	LGPERGAGL	H-2-Dd	0,0197	0,0188	0,0193	0,0063
St8sia3	L235F	SFKKLDGAI	H-2-Kd	0,3804	0,3580	0,3692	0,0864
Sva	T41P	YPLTLNMKI	H-2-Ld	0,0599	0,0608	0,0604	0,0251
Corin	L820F	SFQSEPSGHI	H-2-Kd	0,4697	0,4638	0,4668	0,1069
Gsdma	H289Y	SYLLGKKKEL	H-2-Kd	0,3170	0,5247	0,4209	0,1459
Aars2	A697P	LPHTARIPGL	H-2-Ld	0,5213	0,3942	0,4578	0,0539
Olf95	L212P	FGPILSSYGRI	H-2-Dd	0,8148	0,0529	0,4339	0,0330

Tabela 2 - Candidatos Longos por Agretopicidade

Listagem dos 4 candidatos finais selecionados pelo método de agretopicidade, referente aos alelos de MHC-II. Trazendo informações acerca de seus atributos e características biológicas.

Gene	Aminoác.	Peptídeo	Alelo	EL_Rank	BA_Rank	ELBA_Rank	Agretop.
Pram1	Q572L	DPKLLKLIRKAEKAE	H-2-IEd	1,6500	0,3500	1,0000	0,5089
Pram1	Q572L	PKLLKLIRKAEKAEER	H-2-IEd	0,9700	0,2300	0,6000	0,4478
Wdr33	H13Y	GSPPRFFYMPRFQHQ	H-2-IEd	1,2400	0,5000	0,8700	0,4780
1110038F14Rik	G133R	PTPKQKEQAVRAIRT	H-2-IAAd	1,0100	0,4700	0,7400	0,2236

### 5.1.2. SELEÇÃO POR INTEGRAÇÃO DE SEQUÊNCIAS

A segunda abordagem na busca pelos neoepítomos consistiu na procura por peptídeos candidatos longos, que contivessem internamente a sequência de peptídeos curtos também carregando a mutação. Estes primeiros teriam 15 aminoácidos de comprimento, destinados a encaixar-se em moléculas de MHC de classe II, e os segundos, menores, de 8 a 12, conforme foram coletados os dados após a predição. Ambos os peptídeos aqui precisariam apresentar bons valores de predição. Foram aplicadas diferentes combinações de limiares, entre selecionar pelo menos um deles como SB, e outro pelo menos com WB. Mas no final optou-se por filtrar os candidatos apenas com nível de predição forte (SB). Assim, obtiveram-se 4 peptídeos longos, os quais possuíam internamente um total de 5 curtos. Alguns dos peptídeos curtos estavam contidos num mesmo peptídeo longo, e alguns se repetiram em mais de um longo, como pode ser notado no caso do gene Wdr33, com 2 peptídeos longos e 3 curtos. Alguns desses detalhes podem ser melhor observados na Tabela 3 e Figura 6.

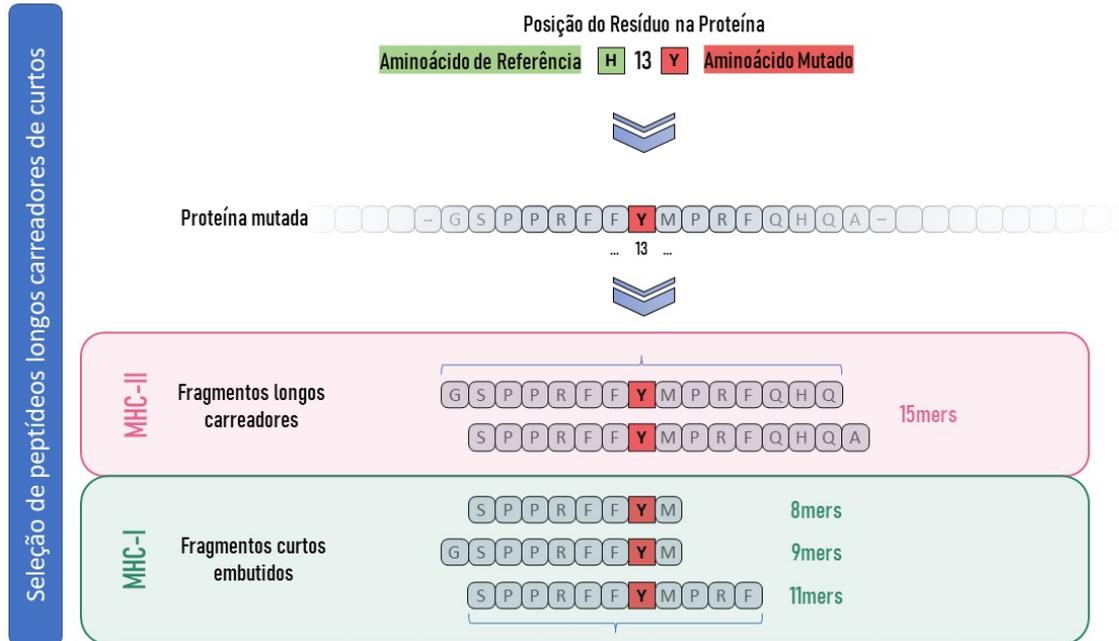
Tabela 3 - Candidatos por Integração de Sequências

Listagem dos 5 candidatos finais longos selecionados pelo método de integração de sequências com peptídeos curtos. Trazendo informações acerca de seus atributos e características biológicas.

Gene	Peptídeo MHC-I	Peptídeo MHC-II	Alelo MHC-I	Alelo MHC-II	ELBA_ Rank_I	ELBA_ Rank_II
Wdr33	SPPRFFYM	GSPRFFYMPRFQHQ	H-2-Dd	H-2-IEd	0,1362	0,8700
Wdr33	GSPRFFYM	GSPRFFYMPRFQHQ	H-2-Dd	H-2-IEd	0,1025	0,8700
Wdr33	SPPRFFYMPRF	GSPRFFYMPRFQHQ	H-2-Ld	H-2-IEd	0,4004	0,8700
Wdr33	SPPRFFYM	SPPRFFYMPRFQHQA	H-2-Dd	H-2-IEd	0,1362	0,3000
Wdr33	SPPRFFYMPRF	SPPRFFYMPRFQHQA	H-2-Ld	H-2-IEd	0,4004	0,3000
Stard13	IYHYVLNSM	IYHYVLNSMVPHP SR	H-2-Kd	H-2-IAd	0,2000	0,7750
Ros1	VPLTFVCHRSL	VPLTFVCHRSLKSHK	H-2-Ld	H-2-IEd	0,1422	0,2550

Figura 6 - Peptídeos longos carreadores de curtos

Demonstração ilustrativa de peptídeos longos carreadores de curtos embebidos em sua sequência de aminoácidos (aa). Foram selecionados candidatos (peptídeos mutados) apenas aqueles que tivessem sido ranqueados dentro limiares considerados fortes (SB) para cada um dos *softwares* de predição, 0,5% e 1%, para MHC-I e MHC-II, respectivamente. Desta forma, tanto peptídeos entre 8-12 aa de comprimento deveriam estar entre os 0,5% melhores, para EL ou BA, quanto os longos deveriam estar entre os 1% melhores para mesmos atributos.



Fonte: figura gerada pelo autor

## 5.2 SELEÇÃO DE CONTROLES POSITIVOS E TESTES DE PERFORMANCE DOS SOFTWARES DE PREDIÇÃO

Com os resultados oriundos dos *softwares* de predição em mãos, tornou-se interessante verificar a performance destes sobre a parcela de peptídeos testados imunologicamente, estes, apresentados pelo artigo que originou os dados, de Schrörs e colaboradores<sup>[57]</sup>. Por mais que a proporção dos testados seja pequena frente ao total dos possíveis neoepítotos gerados pela análise *in silico* aqui descrita – aproximadamente na ordem de 1:650 – esse pequeno espaço amostral, mesmo assim, tem potencial de fornecer informações úteis, como: sensibilidade na avaliação de positivos; limiares considerados toleráveis, por abranger peptídeos comprovadamente imunogênicos; e redesenho de peptídeos controle positivo, capazes de auxiliar as análises que venham a se desdobrar do trabalho aqui descrito.

Acerca da sensibilidade dos *softwares* em identificar peptídeos como positivos, obteve-se resultados interessantes. O preditor NetMHCpan, destinado a predição de neoepítotos curtos, teve boa performance. Como podemos observar na Tabela 4, acerca dos ranqueamentos para MHC-I: 2 dos 2 peptídeos testados foram confirmados como positivos pelo método de EL\_Rank, e 1 pelo método de BA\_Rank. Já a performance, infelizmente, não

se repetiu com a predição a MHC-II, visível na Tabela 5, o que era esperado. Sendo capturados 4 dos 7 que deveriam ser positivos via EL\_Rank, e 3 por BA\_Rank. O que destacou-se negativamente nesta segunda metodologia, por afinidade de ligação, foi a média e desvio padrão altos no ranqueamento, com valores de média 9,70 e desvio padrão 8,28, frente a 4,61 e 4,09, respectivamente, obtidos pela metodologia mais recente, EL\_Rank. Motivo pelo qual foi previamente declarado optar-se por usar apenas a nova forma de ranqueamento na seleção de candidatos longos para agretopicidade usando a presente versão do *software*.

Tabela 4 - Limiares de Ranqueamento NetMHCpan

Demonstração dos limiares atingidos pelos candidatos melhor ranqueados previamente verificados como imunogênicos pelos testes *in vivo* do artigo de Barbara Schrörs e colaboradores. A imunogenicidade destes peptídeos fora confirmada como específica para células CD8<sup>+</sup>.

Gene	Aminoácido	Peptídeo	EL_Rank	BA_Rank	ELBA_Rank	Alelo
Gprc5a	F119L	AHALNLIKL	0,3456	0,8131	0,5794	H-2-Kd
Pzp	G1199E	RPEDVQVKAL	0,6053	2,8286	1,7169	H-2-Ld

Tabela 5 - Limiares de Ranqueamento NetMHCIIpan

Demonstração dos limiares atingidos pelos candidatos melhor ranqueados previamente verificados como imunogênicos pelos testes *in vivo* do artigo de Barbara Schrörs e colaboradores. A imunogenicidade destes peptídeos fora confirmada como específica para células CD4<sup>+</sup>.

Gene	Aminoácido	Peptídeo_BA	Peptídeo_EL	EL_Rank	BA_Rank
Enho	L10I	AAISQGAIIVCNG	AAISQGAIIVCNG	13,12	8,03
Zfr	K411T	HQKVVTLHTKLGKPI	KHQKVVTLHTKLGKP	2,24	3,44
Ilkap	I127S	EREEMQDAHVSLNDI	GEREEMQDAHVSLND	1,02	24,99
Cenpf	D1327E	ESRTECITATSQMTA	ELNESRTECITATSQ	4,56	12,06
Cep120	H68N	RKVLHQNRLQRTPIK	DRKVLHQNRLQRTPI	2,04	1,55
Malt1	T523A	MKFLKDRILLEDKkia	EDKKIAVLLDEVAED	5,92	14,54
Polr2a	M1102I	ITLNTFHYAGVSAKN	ITLNTFHYAGVSAKN	3,40	3,35

Conjuntamente à análise de sensibilidade, a observação do ranqueamento de peptídeos imunogênicos com os valores atingidos pelos candidatos selecionados viabilizou uma “pré-validação” do potencial dos neoepítomos selecionados. É possível ser relatado satisfatoriamente aqui que: todos os candidatos selecionados encontram-se acima dos limiares traçados pelos peptídeos sabidamente imunogênicos. Ou seja, todos oferecem certa confiança no seu potencial em elicitar resposta imune antitumoral.

Peptídeos redesenhados, com melhor ranqueamento, oriundos de mutações validadas como imunogênicas anteriormente,<sup>[57]</sup> foram selecionados para servirem como controles

positivos para futuros testes *in vivo*. Oferecendo-se, assim, uma forma de aferir a integridade dos processos de estratégia vacinal e seleção *in silico* de candidatos. Estes podem ser vistos na Tabela 6.

Tabela 6 - Peptídeos Controle Positivo

Listagem dos candidatos melhor ranqueados, dentre os previamente verificados como imunogênicos pelos testes *in vivo* do artigo de Barbara Schrörs e colaboradores, visando incorporação nos testes vacinais como possíveis controles positivos da seleção peptídica. A imunogenicidade destes peptídeos fora confirmada como específica para células CD8<sup>+</sup>, no peptídeo curto, e CD4<sup>+</sup>, no caso dos dois peptídeos longos.

Gene	Aminoácido	Peptídeo	Alelo	EL_Rank	BA_Rank	ELBA_Rank
Gprc5a	F119L	AHALNLIKL	H-2-Kd	0,3456	0,8131	0,57935
Ilkap	I127S	GEREEMQDAHVSLND	H-2-IAd	1,02	28,94	14,98
Cep120	H68N	DRKVLHQNRLQRTPI	H-2-IEd	2,04	2,11	2,075

## DISCUSSÃO

No presente trabalho foram selecionados peptídeos oriundos da análise de variantes de células tumorais do modelo murino para câncer mamário triplo-negativo, 4T1, com o objetivo de, em trabalhos futuros, constituir uma vacina com atividade antitumoral. Novas formas de tratamento como esta, visando elicitar atividade imunológica antitumoral na forma de vacinas, estão progressivamente tomando mais o empenho de grupos de pesquisa<sup>[21-26]</sup>. Não existem regras padronizadas que governem a seleção destes alvos terapêuticos. Desta maneira, estratégias diferentes foram empregadas, cujos detalhes serão discorridos em seguida.

A agretopicidade já é um conceito proposto na construção teórica dos neoepítomos. A diferença dos novos epítomos, oriundos de mutações, de seus pares próprios, é teorizada como viabilizadora da atividade antitumoral de células T citotóxicas. No entanto, formas de interrogar essa diferenciação ainda são escassas na literatura. Tendo então a tarefa de selecionar quais mutações seriam mais promissoras na formulação de uma estratégia vacinal, o presente trabalho se viu na atribuição de elaborar formas práticas de abordar a problemática da diferenciação destes peptídeos. O principal método de garantir agretopicidade entre os peptídeos promissores foi selecionar neoepítomos com ranqueamento bom (ELBA\_Rank<0,5%), cuja mutação causou distanciamento considerável de onde foram ranqueados seus pares próprios. Ou seja, usou-se o próprio software de predição que avalia ligação às moléculas de MHC tanto para interrogar a associação destas com os candidatos, quanto para certificar-se da dessemelhança destes com peptídeos próprios. Além de garantir esse distanciamento no ranqueamento dos candidatos, mensurou-se a razão direta entre o

ranqueamento mutado-próprio dos melhores candidatos, na tentativa de quantificar a mudança ocasionada pela mutação.

Uma das explicações do porquê a agretopicidade levaria à constituição de neoepítomos se dá na narrativa que peptídeos próprios, antes incapazes de associação com moléculas de MHC, poderiam passar a ter maior afinidade após a mudança de um de seus resíduos. Assim, o novo resíduo poderia, por exemplo, induzir uma interação com sítios de ancoramento do complexo de apresentação de antígenos não existente na presença do resíduo original, próprio. Uma vantagem dessa abordagem é que, em teoria, como os peptídeos não seriam mais percebidos como iguais, não seria tão relevante uma expressão diferencial do gene mutado, que favorecesse a apresentação destes peptídeos no contexto tumoral, pois as células T associadas ao neoepítomo não interagiriam com as células do tecido que não apresentem a mutação. Seria ainda necessário, claro, um nível de expressão mínima desses genes mutados e integridade das vias de apresentação de antígenos.

Um dos problemas evidenciados durante a condução das análises foi o desempenho insatisfatório da predição associada a MHC-II, relatado na avaliação de performance com peptídeos cuja imunogenicidade fora previamente verificada. Esta questão já era esperada, pelos desafios conhecidamente ligados ao complexo de classe II. O fato de o sulco presente nessa classe de moléculas ser maior que o presente nas de classe I permite um encaixe mais promíscuo de peptídeos, permitindo que sequências não associadas diretamente com o ancoramento flanqueiem a sequência que efetivamente faria o ancoramento. Esse cenário mais instável da interação peptídeo-MHC torna a predição de epítomos para a classe II uma tarefa consideravelmente mais laboriosa. Por isso, já era esperada uma performance menor, e, por consequência, faz-se necessário lembrar de que os candidatos longos, selecionados para interação com células T CD4+, por dependerem de ligação a MHC-II, possuem nível de confiança inferior aos candidatos selecionados a MHC-I. Uma vez que os desafios promovidos pelo complexo MHC-II ainda não foram superados pelos *softwares* de predição.

Como forma de complementar a abordagem de agretopicidade, brevemente mencionada, buscou-se por neoepítomos longos, destinados a apresentação via MHC-II, cujas sequências possuíssem internamente as sequências de peptídeos curtos, que seriam apresentados via MHC-I. Esta estratégia de imunidade cruzada entre células T CD4+, dependentes de apresentação de peptídeos longos, e CD8+, dependentes de apresentação de peptídeos curtos, foi inspirada pela predileção destes feita pelo trabalho de Heitmann e colaboradores.<sup>[67]</sup> É admitida a hipótese que a apresentação tautócrona de epítomos, por células apresentadores de antígenos competentes, proporcione uma resposta imunológica mais

dinâmica e eficaz. É importante ressaltar que aqui não é priorizada a agretopicidade entre os neoepítomos e seus pares próprios, mas sim, a sobreposição entre os epítomos.

Por fim, deve-se trazer à tona o fato de que nenhum dos peptídeos usados nos testes imunológicos apresentados pelo trabalho de Schrörs e colaboradores,<sup>[57]</sup> e verificados como imunogênicos, coincidiram com os selecionados pelo presente trabalho, ainda que os dados sejam os mesmos. Esta questão poderia passar despercebida, ou taxada simplesmente como mero acaso, mas discuti-la pode ser interessante para compreender melhor sobre a metodologia aplicada pelo presente estudo. Duas indagações principais derivam desta observação: por que o grupo de mutações validadas não apareceram pelas análises aqui apresentadas? O fato de as mutações esperadas não aparecerem invalida o método?

O primeiro levantamento é de natureza mais técnica, e para compreendê-lo é bom termos em mente a arquitetura de uma matriz de confusão. Nesta matriz dividem-se valores preditos e reais em positivos e negativos, ambos possuindo atributos “verdadeiros” e “falsos”, ou seja, há “verdadeiros positivos”, “falsos positivos”, “verdadeiros negativos” e “falsos negativos”. Para conseguirmos mensurar a performance verdadeira de um método seria necessário verificar cada resultado individualmente, na totalidade de uma população, ou para mensurar uma aproximação, poderia-se fazer a interrogação de uma determinada amostragem, que carregará uma margem de confiança denotando o quão bem a amostra reflete a realidade expressa pela população total. O fato é: cada ferramenta usada para predizer algo trará resultados corretos e incorretos. O conjunto total de peptídeos imunogênicos para um determinado grupo de alelos MHC é desconhecido, mal estamos começando a compreender melhor aspectos gerais acerca da imunologia da apresentação de antígenos. Os níveis de complexidade abstraídos pelos modelos de predição são associados ao quão limitado nosso conhecimento real sobre estes tópicos são. Sendo assim, cada ferramenta encontra-se em um *trade-off*: priorizar a captura de mais resultados positivos, aumentando a sensibilidade, mas diminuindo a precisão, ou seja, permitindo-se mais falsos positivos passarem pelos filtros. Ou priorizar a precisão, inevitavelmente deixando alguns positivos incógnitos (falsos negativos), capturando menos resultados positivos para testar, mas com maior confiança destes. Compreendendo esta compensação de escolhas, deve-se dizer que alguns neoepítomos carregando mutações validadas passaram pelos limiares do software NetMHCpan inicialmente, com associações de nível WB a SB. Porém, como foram mais de 223 candidatos iniciais, com associação SB para algum dos atributos de ranqueamento (EL\_Rank ou BA\_Rank), conforme descrito na porção de seleção dos resultados, expôs-se a necessidade de uma apuração mais minuciosa. Minúcia cumprida pela aplicação da metodologia de

agretopicidade, elegendo, enfim, 8 candidatos para MHC-I. Assim, muitos possíveis neoepítomos foram descartados das análises futuras, mas esta é uma decisão inerente a qualquer método de seleção de alvos biológicos em potencial.

O segundo apontamento, talvez mais afrontoso, que questiona a veracidade dos resultados aqui apresentados, é naturalmente mais laborioso e delicado de se responder. Como foi explanado há pouco, escolhas sobre o quão criteriosos são os limiares impostos pelas técnicas irão, invariavelmente, limitar o montante de resultados final. Mesmo assim, o fato de que aqueles resultados validados não terem surgido como candidatos ao final da análise, mas funcionariam para o propósito, no final das contas, denotaria ineficiência da seleção empregada? Infelizmente, não há resposta clara capaz de sanar as dúvidas aqui. A questão da seleção de neoepítomos não só é de natureza altamente biológica, com várias camadas de complexidade, e, portanto, de difícil resolução, como também possui muitas lacunas ainda a se preencher. Por este motivo, tantos trabalhos prévios não foram capazes, até o momento, de nos levar muito longe<sup>[48,49]</sup>. Carecem-nos conhecimentos acerca das interações entre agentes imunes, detalhes moleculares que implicam no processamento, apresentação e reconhecimento de neoantígenos, etc. Enxergando apenas o começo e o destino final, o que nos resta é conjecturar sobre o que pode haver neste caminho, e como nossas escolhas podem afetar o objetivo final. Tendo isso em mente, um dos cenários cabíveis, caso fossem usados peptídeos mutados não diferenciáveis dos pares próprios para imunização, talvez indivíduos sujeitos a estes procedimentos apresentassem sintomas característicos de resposta autoimune em algum nível, como relato em outras imunoterapias.<sup>[31]</sup> Visto que foram, em tese, expandidas populações de leucócitos para combater as células defeituosas, e, em razão da não diferenciação entre os epítomos, agora estas células citotóxicas poderiam erroneamente reconhecer células saudáveis como alvo também. Ou os neoepítomos poderiam servir contrariamente ao propósito almejado, serem pró-tumorais, como já são levantadas discussões acerca de linfócitos com fenótipo pró-tumoral.<sup>[62]</sup>

O que absorvemos destes apontamentos é que mais atenção deve ser voltada ao preenchimento das lacunas que minam nossa compreensão sobre os vários elementos que envolvem os neoepítomos. Quão antes e melhor as questões que regem a identificação de neoantígenos forem elucidadas, mais nos aproximaremos de um cenário em que as plataformas de vacinação terapêutica promoverão desfechos clínicos relevantes aos pacientes.

## CONCLUSÃO

A seleção de neoepítos foi concluída, sendo selecionados 8 candidatos curtos, 4 longos, e 4 longos contendo curtos internalizados. O resultado das comparações entre estes candidatos e outros comprovadamente imunogênicos prenuncia o êxito do processo. No entanto, confirmações biológicas ainda serão necessárias para qualificar as predições aqui feitas. Esta carência será parcialmente abordada em experimentos *in vivo* futuros, permitindo, assim, acessar a eficácia dos candidatos incluídos no estudo.

Ademais das perspectivas biológicas ainda a se validar, pode-se dizer que um fruto ainda em maturação, desenvolvido concomitantemente à seleção dos neoepítos, é o próprio conjunto de *scripts* desenvolvidos para selecioná-los. Há expectativas de que estes venham a compor futuramente um *software* capaz de performar como selecionador de neoepítos, agregado a funções ainda a se desenvolver. Quando mais lapidado, capaz de executar chamada de variantes e algumas competências complementares, terá potencial de auxiliar no desenvolvimento de vacinas antitumorais personalizadas para uso clínico no futuro.

No presente trabalho foram apontadas lacunas e limitações que ainda obstruem a identificação de neoepítos, postergando a exploração do poder neles contidos. Apesar dos obstáculos, ensaios clínicos pequenos já expressam resultados iniciais em favor do uso de vacinas terapêuticas antitumorais. Foi aqui evidenciado o poder oferecido pelas análises *in silico* no fomento a estas novas estratégias terapêuticas, bem como suas presentes restrições. Pretendemos continuar trabalhando no desenvolvimento de ferramentas que provejam estas terapias de informações que melhor assegurem seu sucesso.

## REFERÊNCIAS

1. NATIONAL CANCER INSTITUTE (NCI). **What Is Cancer?**. Disponível em: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. Acesso em: 10 julho de 2022.
2. MINISTÉRIO DA SAÚDE. **Sistema de Informação sobre Mortalidade, 2021**. <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>. Acesso em: 10 julho 2022.
3. WORLD HEALTH ORGANIZATION (WHO). **Global Cancer Observatory**. <https://gco.iarc.fr/today/home>. Acesso em: 21 maio 2023.
4. INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Estimativa 2023**. Incidência de Câncer no Brasil. P.62. Rio de Janeiro: INCA, 2022. <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//estimativa-2023.pdf>.
5. SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS (SEER). **Cancer Stat Facts Female Breast Cancer Subtypes**. 5-Year Relative Survival Percent, Female Breast Cases by Cancer Subtype, SEER 22 (Excluding IL/MA) 2013–2019. <https://seer.cancer.gov/statfacts/html/breast-subtypes.html>. Acesso em: 21 maio 2023.

6. CANCER RESEARCH UK. Survival Statistics Data Table: England and Wales survival summary. Abril 2014.  
[https://www.cancerresearchuk.org/sites/default/files/cs\\_dt\\_survival.xlsx](https://www.cancerresearchuk.org/sites/default/files/cs_dt_survival.xlsx). Acesso em: 21 maio 2023.
7. SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS (SEER). Cancer Stat Facts Female Breast Cancer Subtypes. Percent of Female Breast Cases by Cancer Subtype, SEER 22 2016–2020. <https://seer.cancer.gov/statfacts/html/breast-subtypes.html>. Acesso em: 21 maio 2023.
8. MORRIS, G. J. et al. Differences in breast carcinoma characteristics in newly diagnosed African–American and Caucasian patients. **Cancer**, v. 110, n. 4, p. 876–884, 15 ago. 2007.  
<https://doi.org/10.1002/cncr.22836>.
9. CALEFFI, M. et al. Breast cancer survival in Brazil: How much health care access impact on cancer outcomes? **The Breast**, v. 54, p. 155–159, 1 dez. 2020.  
<https://doi.org/10.1016/j.breast.2020.10.001>.
10. CENTERS FOR DISEASE CONTROL AND PREVENTION. **Breast Cancer Treatment (Adult) (PDQ®)–Health Professional Version - National Cancer Institute**. Disponível em: <<https://www.cancer.gov/types/breast/hp/breast-treatment-pdq>>.
11. CENTERS FOR DISEASE CONTROL AND PREVENTION. **Triple-Negative Breast Cancer**. Disponível em: <<https://www.cdc.gov/cancer/breast/triple-negative.htm>>.
12. HANAHAN, Douglas; WEINBERG, Robert A.. The Hallmarks of Cancer. **Cell**, [S.L.], v. 100, n. 1, p. 57-70, jan. 2000. Elsevier BV. [http://dx.doi.org/10.1016/s0092-8674\(00\)81683-9](http://dx.doi.org/10.1016/s0092-8674(00)81683-9).
13. HANAHAN, Douglas; WEINBERG, Robert A.. Hallmarks of Cancer: the next generation. **Cell**, [S.L.], v. 144, n. 5, p. 646-674, mar. 2011. Elsevier BV.  
<http://dx.doi.org/10.1016/j.cell.2011.02.013>.
14. HANAHAN, Douglas. Hallmarks of Cancer: new dimensions. **Cancer Discovery**, [S.L.], v. 12, n. 1, p. 31-46, 1 jan. 2022. American Association for Cancer Research (AACR).  
<http://dx.doi.org/10.1158/2159-8290.cd-21-1059>.
15. PRUITT, K.. Molecular and Cellular Changes During Cancer Progression Resulting From Genetic and Epigenetic Alterations. **Progress In Molecular Biology And Translational Science**, [S.L.], p. 3-47, 2016. Elsevier. <http://dx.doi.org/10.1016/bs.pmbts.2016.09.001>.
16. BAYLIN, Stephen B.; JONES, Peter A.. Epigenetic Determinants of Cancer. **Cold Spring Harbor Perspectives In Biology**, [S.L.], v. 8, n. 9, p. 1-35, 18 maio 2016. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/cshperspect.a019505>.
17. BIDRAM, Elham; ESMAEILI, Yasaman; RANJI-BURACHALOO, Hadi; AL-ZAUBAI, Nuha; ZARRABI, Ali; STEWART, Alastair; DUNSTAN, Dave E.. A concise review on cancer treatment methods and delivery systems. **Journal Of Drug Delivery Science And Technology**, [S.L.], v. 54, p. 101350, dez. 2019. Elsevier BV.  
<http://dx.doi.org/10.1016/j.jddst.2019.101350>.
18. MALMBERG, Karl-Johan. Effective immunotherapy against cancer. **Cancer Immunology, Immunotherapy**, [S.L.], v. 53, n. 10, p. 879-892, 28 jul. 2004. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s00262-004-0577-x>.
19. SCHUSTER, Manfred; NECHANSKY, Andreas; KIRCHEIS, Ralf. Cancer immunotherapy. **Biotechnology Journal**, [S.L.], v. 1, n. 2, p. 138-147, fev. 2006. Wiley.  
<http://dx.doi.org/10.1002/biot.200500044>.
20. FARKONA, Sofia; DIAMANDIS, Eleftherios P.; BLASUTIG, Ivan M.. Cancer immunotherapy: the beginning of the end of cancer?. **Bmc Medicine**, [S.L.], v. 14, n. 1, art 73, 5 maio 2016. Springer Science and Business Media LLC.  
<http://dx.doi.org/10.1186/s12916-016-0623-5>.

21. OSÓRIO, J. Cancer immunotherapy research round-up. **Nature**, v. 552, n. 7685, p. S70–S71, dez. 2017. <https://doi.org/10.1038/d41586-017-08703-6>.
22. OTT, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. **Nature**, v. 547, n. 7662, p. 217–221, jul. 2017. <https://doi.org/10.1038/nature22991>.
23. BISWAS, N. et al. Designing neoantigen cancer vaccines, trials, and outcomes. **Frontiers in Immunology**, v. 14, 9 fev. 2023. <https://doi.org/10.3389/fimmu.2023.1105420>.
24. KESKIN, D. B. et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. **Nature**, v. 565, n. 7738, p. 234–239, 19 dez. 2018. <https://doi.org/10.1038/s41586-018-0792-9>.
25. SAHIN, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. **Nature**, v. 547, n. 7662, p. 222–226, jul. 2017. <https://doi.org/10.1038/nature23003>.
26. SHIBATA, H. et al. Personalized Cancer Vaccination in Head and Neck Cancer. **Cancer Science**, 27 dez. 2020. <https://doi.org/10.1111/cas.14784>.
27. TURTLE, C.; RIDDELL, S.; MALONEY, D. CD19-Targeted chimeric antigen receptor-modified T-cell immunotherapy for B-cell malignancies. **Clinical Pharmacology & Therapeutics**, v. 100, n. 3, p. 252–258, 17 jun. 2016. <https://doi.org/10.1002/cpt.392>.
28. FRY, Terry J.; MACKALL, Crystal L.. T-cell adoptive immunotherapy for acute lymphoblastic leukemia. **Hematology**, [S.L.], v. 2013, n. 1, p. 348-353, 6 dez. 2013. American Society of Hematology. <http://dx.doi.org/10.1182/asheducation-2013.1.348>.
29. VAN CUTSEM, E. et al. Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. **New England Journal of Medicine**, v. 360, n. 14, p. 1408–1417, 2 abr. 2009. <http://doi.org/10.1056/NEJMoa0805019>.
30. HUDIS, C. A. Trastuzumab — Mechanism of Action and Use in Clinical Practice. **New England Journal of Medicine**, v. 357, n. 1, p. 39–51, 5 jul. 2007. <http://doi.org/10.1056/NEJMra043186>.
31. HODI, F. S. et al. Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. **New England Journal of Medicine**, v. 363, n. 8, p. 711–723, 19 ago. 2010. <http://doi.org/10.1056/NEJMoa1003466>.
32. BOECKH, Michael; COREY, Lawrence. Adoptive Immunotherapy of Viral Infections: should infectious disease embrace cellular immunotherapy?. **The Journal Of Infectious Diseases**, [S.L.], v. 216, n. 8, p. 926-928, 23 set. 2017. Oxford University Press (OUP). <http://dx.doi.org/10.1093/infdis/jix360>.
33. SEIF, Michelle; EINSELE, Hermann; LÖFFLER, Jürgen. CAR T Cells Beyond Cancer: hope for immunomodulatory therapy of infectious diseases. **Frontiers In Immunology**, [S.L.], v. 10, n. 2711, p. 1-8, 21 nov. 2019. Frontiers Media SA. <http://dx.doi.org/10.3389/fimmu.2019.02711>.
34. A SABATOS-PEYTON, Catherine; VERHAGEN, Johan; WRAITH, David C. Antigen-specific immunotherapy of autoimmune and allergic diseases. **Current Opinion In Immunology**, [S.L.], v. 22, n. 5, p. 609-615, out. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.coi.2010.08.006>.
35. RAKSHIT, Sagar; MOLINA, Julian R.. Immunotherapy in patients with autoimmune disease. **Journal Of Thoracic Disease**, [S.L.], v. 12, n. 11, p. 7032-7038, nov. 2020. AME Publishing Company. <http://dx.doi.org/10.21037/jtd-2019-cptn-10>.
36. KIMIZ-GEBOLOGLU, Ilgin; GULCE-IZ, Sultan; BIRAY-AVCI, Cigir. Monoclonal antibodies in cancer immunotherapy. **Molecular Biology Reports**, [S.L.], v. 45, n. 6, p. 2935-2940, 11 out. 2018. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s11033-018-4427-x>.

37. HAFEEZ, Umbreen; GAN, Hui K; SCOTT, Andrew M. Monoclonal antibodies as immunomodulatory therapy against cancer and autoimmune diseases. **Current Opinion In Pharmacology**, [S.L.], v. 41, p. 114-121, ago. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.coph.2018.05.010>.
38. COVENTRY, Brendon J.. Therapeutic vaccination immunomodulation: forming the basis of all cancer immunotherapy. **Therapeutic Advances In Vaccines And Immunotherapy**, [S.L.], v. 7, p. 1-28, jan. 2019. SAGE Publications. <http://dx.doi.org/10.1177/2515135519862234>.
39. AURELI, A. et al. Immunotherapy as a Turning Point in the Treatment of Acute Myeloid Leukemia. v. 13, n. 24, p. 6246–6246, 13 dez. 2021.. <http://dx.doi.org/10.3390/cancers13246246>.
40. HUNGER, S. P. et al. Improved Survival for Children and Adolescents With Acute Lymphoblastic Leukemia Between 1990 and 2005: A Report From the Children's Oncology Group. **Journal of Clinical Oncology**, v. 30, n. 14, p. 1663–1669, 10 maio 2012. <http://dx.doi.org/10.1200/jco.2011.37.8018>.
41. PHAM, T. et al. An Update on Immunotherapy for Solid Tumors: A Review. **Annals of Surgical Oncology**, v. 25, n. 11, p. 3404–3412, 23 jul. 2018. <http://dx.doi.org/10.1245/s10434-018-6658-4>.
42. MENON, Smitha; SHIN, Sarah; DY, Grace. Advances in Cancer Immunotherapy in Solid Tumors. **Cancers**, [S.L.], v. 8, n. 12, p. 106, 24 nov. 2016. MDPI AG. <http://dx.doi.org/10.3390/cancers8120106>.
43. SCHUMACHER, Ton N.; SCHREIBER, Robert D.. Neoantigens in cancer immunotherapy. **Science**, [S.L.], v. 348, n. 6230, p. 69-74, 3 abr. 2015. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.aaa4971>.
44. TRAN, Eric; ROBBINS, Paul F; A ROSENBERG, Steven. 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. **Nature Immunology**, [S.L.], v. 18, n. 3, p. 255-262, mar. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/ni.3682>.
45. YARCHOAN, Mark; JOHNSON, Burles A.; LUTZ, Eric R.; LAHERU, Daniel A.; JAFFEE, Elizabeth M.. Targeting neoantigens to augment antitumour immunity. **Nature Reviews Cancer**, [S.L.], v. 17, n. 4, p. 209-222, 24 fev. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nrc.2016.154>.
46. WELLS, D. K. et al. Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. **Cell**, v. 183, n. 3, p. 818-834.e13, 29 out. 2020. <http://dx.doi.org/10.1016/j.cell.2020.09.015>.
47. ASSADIPOUR, Y. et al. Characterization of an Immunogenic Mutation in a Patient with Metastatic Triple-Negative Breast Cancer. **Clinical Cancer Research**, v. 23, n. 15, p. 4347–4353, 1 ago. 2017. <http://dx.doi.org/10.1158/1078-0432.ccr-16-1423>.
48. The problem with neoantigen prediction. **Nature Biotechnology**, [S.L.], v. 35, n. 2, p. 97-97, fev. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nbt.3800>.
49. VITIELLO, Antonella; ZANETTI, Maurizio. Neoantigen prediction and the need for validation. **Nature Biotechnology**, [S.L.], v. 35, n. 9, p. 815-817, set. 2017. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nbt.3932>.
50. SCHENCK, R. O. et al. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. **BMC Bioinformatics**, v. 20, n. 1, 22 maio 2019. <http://dx.doi.org/10.1186/s12859-019-2876-4>.
51. ANA VARELA COELHO et al. neoANT-HILL: an integrated tool for identification of potential neoantigens. v. 13, n. 1, 22 fev. 2020. <http://dx.doi.org/10.1186/s12920-020-0694-1>.

52. LANG, F. et al. NeoFox: annotating neoantigen candidates with neoantigen features. v. 37, n. 22, p. 4246–4247, 10 maio 2021. <https://doi.org/10.1093/bioinformatics/btab344>.
53. NIELSEN, M. et al. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. **Immunogenetics**, v. 57, n. 1-2, p. 33–41, 3 mar. 2005. <https://doi.org/10.1007/s00251-005-0781-7>.
54. LARSEN, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. **BMC Bioinformatics**, v. 8, n. 1, 31 out. 2007. <https://doi.org/10.1186/1471-2105-8-424>.
55. REYNISSON, B. et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. **Nucleic Acids Research**, v. 48, n. W1, p. W449–W454, 14 maio 2020. <http://dx.doi.org/10.1093/nar/gkaa379>.
56. SARKIZOVA, S. et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. **Nature Biotechnology**, v. 38, n. 2, p. 199–209, 16 dez. 2019. <https://doi.org/10.1038/s41587-019-0322-9>.
57. SCHRÖRS, B. et al. Multi-Omics Characterization of the 4T1 Murine Mammary Gland Tumor Model. **Frontiers in Oncology**, v. 10, 23 jul. 2020. <https://doi.org/10.3389/fonc.2020.01195>.
58. **Mus\_musculus\_BALB\_cJ - Ensembl genome browser 109**. Disponível em: <[https://www.ensembl.org/Mus\\_musculus\\_BALB\\_cJ/Info/Index](https://www.ensembl.org/Mus_musculus_BALB_cJ/Info/Index)>. Acesso em: 26 maio. 2023.
59. VAN ROSSUM, G. & DRAKE, F.L., 2009. **Python 3 Reference Manual**, Scotts Valley, CA: CreateSpace. <https://docs.python.org/3/reference/>.
60. GNU, P., 2007. Free Software Foundation. **Bash (3.2. 48)[Unix shell program]**. <https://www.gnu.org/software/bash/manual/bash.html>.
61. GEO. **GEO - NCBI**. Disponível em: <<https://www.ncbi.nlm.nih.gov/geo/>>.
62. LANG, F. et al. Identification of neoantigens for individualized therapeutic cancer vaccines. **Nature Reviews Drug Discovery**, v. 21, n. 4, p. 261–282, 1 fev. 2022. <https://doi.org/10.1038/s41573-021-00387-y>.
63. ROTH, A. et al. PyClone: statistical inference of clonal population structure in cancer. **Nature Methods**, v. 11, n. 4, p. 396–398, 1 abr. 2014. <https://doi.org/10.1038/nmeth.2883>.
64. JURTZ, V. et al. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. **The Journal of Immunology**, v. 199, n. 9, p. 3360–3368, 4 out. 2017. <https://doi.org/10.4049/jimmunol.1700893>.
65. **Altmetric – NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data**. Disponível em: <<https://www.altmetric.com/details/27228116/citations>>. Acesso em: 26 maio. 2023.
66. **THERMOFISHER - Mouse Haplotype Table**. Disponível em: <[https://assets.thermofisher.com/TFS-Assets/LSG/brochures/Mouse\\_Haplotype\\_Table.pdf](https://assets.thermofisher.com/TFS-Assets/LSG/brochures/Mouse_Haplotype_Table.pdf)>. Acesso em: 26 maio. 2023.
67. HEITMANN, J. S. et al. A COVID-19 peptide vaccine for the induction of SARS-CoV-2 T cell immunity. **Nature**, 23 nov. 2021. <https://doi.org/10.1038/s41586-021-04232-5>.