



UNIVERSIDAD  
SAN SEBASTIAN  
PROYECTO HUMANISTA CRISTIANO

FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO  
SEDE CONCEPCIÓN

# Mapeo *in silico* de regiones ricas en epítomos de células T en el proteoma de SARS-CoV-2: población chilena como caso de estudio

Tesis para optar al **Grado de Magíster en Innovación en Biociencias y  
Bioingeniería**

Profesor Guía: Dr. Patricio Oyarzún Cayo  
Estudiante: Juan I. Cisternas Ramos

## **AGRADECIMIENTOS**

A Marta Luisa Ramos Acuña, Juan Carlos Cisternas Mandujano, María Francisca Cisternas Ramos y María Luisa Cisternas Ramos por su amor incondicional, por su valiosa guía, energía e inspiración para trazar y recorrer nuevos rumbos. En ellos también está el refugio cuando el norte se disipa.

A los profesores del Magister y en especial al profesor Patricio Oyarzún por su valioso conocimiento y por forjarme en la profesión de la investigación (rubro que me apasiona) y la innovación.

A Mónica Araya por concederme el amor en tiempos de pandemia. En ella encontré un nuevo núcleo de apoyo y contención bañado en astucia y ternura que no dejan de sorprenderme hasta el día de hoy. Eternamente agradecido de ella y su familia por ser una parte importante en la superación de este proceso.

## ÍNDICE

<b>RESUMEN</b> .....	<b>1</b>
<b>ABSTRACT</b> .....	<b>2</b>
<b>1. INTRODUCCIÓN</b> .....	<b>3</b>
1.1. Enfermedades infecciosas emergentes .....	3
1.2. SARS-CoV-2.....	5
1.2.1. COVID-19: situación en Chile.....	7
1.3. Vacunas .....	9
1.3.1. Vacunas de primera generación.....	11
1.3.2. Vacunas de segunda generación.....	11
1.3.3. Vacunas de tercera generación.....	15
1.3.4. Vacunas basadas en epítomos .....	17
1.4. Sistema inmune.....	18
1.4.1. Inmunidad humoral.....	19
1.4.2. Inmunidad celular.....	19
1.4.3. Antígeno Leucocitario Humano (HLA).....	20
1.4.4. Polimorfismo de HLA.....	22
1.5. Inmunoinformática.....	24
1.5.1. Métodos pan-específicos .....	25
1.5.2. Predivac-3.0.....	26
<b>HIPÓTESIS</b> .....	<b>29</b>
<b>Objetivo general</b> .....	<b>29</b>
<b>Objetivos específicos</b> .....	<b>29</b>
<b>2. Estrategia experimental</b> .....	<b>30</b>
2.1. Construcción de base de datos de epítomos de células T .....	31
2.2. Predicción y mapeo <i>in silico</i> de epítomos de células T: Predivac-3.0 .....	33
2.2.1. Mapeo <i>in silico</i> de epítomos de células T.....	33
2.2.2. Aplicación poblacional: Detección de <i>hotspots</i> .....	33
2.3. Visualización epítomos putativos de células T .....	35
2.3.1. Análisis de epítomos de <i>hotspots</i> en SARS-CoV2 específicos para poblaciones en Chile.....	37
<b>3. RESULTADOS</b> .....	<b>39</b>
3.1. Construcción de base de datos EpiTex. ....	39
3.1.1. Epítomos de células T de EpiTex en proteínas estructurales.....	40
3.1.2. Epítomos de células T de EpiTex en proteínas codificadas por ORF.....	40
3.1.3. Restricciones HLA asociadas a los epítomos de células T de EpiTex .....	41

3.2.	Mapeo <i>in silico</i> de epítomos de SARS-CoV-2 .....	45
3.2.1.	Mapeo de epítomos de células T CD8+ .....	46
3.2.2.	Mapeo de epítomos de células T CD4+ .....	48
3.3.	Predicción de epítomos de células T considerando componente étnico en Chile.....	50
3.3.1.	Predicción de epítomos y <i>hotspots</i> de células T CD8+ .....	50
3.3.2.	Predicción de epítomos y <i>hotspots</i> de células T CD4+ .....	58
<b>4.</b>	<b>DISCUSIÓN</b> .....	<b>61</b>
4.1.	Variaciones entre predicciones de epítomos de células T CD4+ y CD8+ .....	61
4.2.	Predivac-3.0 y componente étnico en Chile .....	62
4.3.	<i>Hotspots</i> .....	64
4.3.1.	ORF .....	¡Error! Marcador no definido.
4.3.2.	Variabilidad genética de proteínas de SARS-CoV-2.....	67
<b>5.</b>	<b>CONCLUSIÓN</b> .....	<b>69</b>
<b>6.</b>	<b>BIBLIOGRAFÍA</b> .....	<b>71</b>

## ÍNDICE DE FIGURAS

Figura 1	Mapa de registro total de casos confirmados de contagio con SARS-CoV-2 adaptado del centro de recursos de coronavirus de la universidad de medicina Johns Hopkins ( <a href="https://coronavirus.jhu.edu/map.html">https://coronavirus.jhu.edu/map.html</a> ) actualizado para el 25 de Agosto de 2022.....	<b>4</b>
Figura 2	Esquema de proteoma y estructura de SARS-CoV-2 con proteínas estructurales y genoma de ARN extraído de los recursos de coronavirus de ViralZone, SIB Swiss Institute of Bioinformatics ( <a href="https://viralzone.expasy.org/9056">https://viralzone.expasy.org/9056</a> ).....	<b>6</b>
Figura 3	Dosis totales acumulativas de vacunas administradas por cada 100 personas a lo largo del tiempo desde el 13 de Diciembre de 2020 hasta el 23 de Noviembre de 2022, adaptado de “Our World in Data” (Mathieu et al., 2021). .....	<b>7</b>
Figura 4	Representación gráfica de estructuras de PfCSP y vacuna RTS,S (modificado de Lubanga et al., 2016). .....	<b>13</b>
Figura 5	Esquema gráfico de vacuna CYD-TDV (modificado de Wilken & Rimmelzwaan, 2020).....	<b>14</b>
Figura 6	Esquema del ARNm de la vacuna en las vacunas BionTech-Pfizer y Moderna modificado de los recursos de coronavirus de ViralZone, SIB Swiss Institute of Bioinformatics ( <a href="https://viralzone.expasy.org/9696">https://viralzone.expasy.org/9696</a> ).. .....	<b>16</b>
Figura 7	Inmunidad innata e inmunidad adaptativa extraído de Abbas et al, 2022. ....	<b>18</b>
Figura 8	Esquema de hendidura de unión de HLA-I y HLA-II adaptado de Zhang et al., 2012.....	<b>20</b>
Figura 9	Esquema de HLA en la superficie celular de APC presentando péptidos antigénicos a sus respectivos receptores de células T (extraído de: Sanchez-Trincado et al., 2017).....	<b>21</b>
Figura 10	Evolución anual del número de alelos de HLA descubiertos y anotados en la base de datos IPD-IMGT/HLA ( <a href="http://www.ebi.ac.uk/ipd/imgt/hla/">http://www.ebi.ac.uk/ipd/imgt/hla/</a> ) (Robinson et al., 2020).....	<b>23</b>
Figura 11	Posiciones SDR y criterios empleados para seleccionar estas posiciones en términos de polimorfismo y efecto electrostático calculado en la hendidura de unión, adaptado de Oyarzún et al., 2013. ....	<b>27</b>
Figura 12	Validación de predicciones de epítomos de células T CD8+ extraído de Oyarzún et al., 2021.....	<b>28</b>
Figura 13	Esquema de diagrama de flujo que representa los pasos seguidos por Predivac-3.0 para realizar un mapeo in silico de epítomos de células T de proteomas virales, extraído de Oyarzún et al., 2021.. .....	<b>34</b>
Figura 14	Esquema circular para visualización de epítomos de células T según clase de HLA.....	<b>36</b>
Figura 15	Esquema circular para visualización de epítomos de células T según clase de HLA y utilizando criterio poblacional. ....	<b>38</b>
Figura 16	Gráfico de barras de cantidad de epítomos de células T descubiertos experimentalmente en proteínas de SARS-CoV-2, ordenado de forma descendente. ....	<b>39</b>
Figura 17	Gráfico de barras de restricciones HLA asociadas a epítomos de células T descubiertos experimentalmente de SARS-CoV-2. ....	<b>42</b>

Figura 18	Visualización circular de mapeo de epítomos de células T CD8+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) realizado con Circos.....	<b>47</b>
Figura 19	Visualización circular de mapeo de epítomos de células T CD4+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) realizado con Circos.....	<b>49</b>
Figura 20	Visualización circular de mapeo de epítomos de células T CD8+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) para Chile y grupos étnicos “Mixed” y “Amerindian”, realizado con Circos. ....	<b>51</b>
Figura 21	Visualización circular de mapeo de epítomos de células T CD4+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) para Chile y grupos étnicos “Mixed” y “Amerindian”, realizado con Circos. ....	<b>59</b>
Figura 22	Esquema de arreglo genómico de SARS-CoV-2 extraído de Yadav et al., 2021.....	<b>65</b>
Figura 23	<i>Hotspots</i> de epítomos de células T CD8+ de ORF10 y ORF6 de SARS-CoV-2 para la población de Chile.....	<b>67</b>

## ÍNDICE DE TABLAS

Tabla 1	Reporte de cantidad de personas vacunadas en Chile actualizado para el 31 de Agosto de 2022 de acuerdo con el MINSAL. ....	<b>8</b>
Tabla 2	Organización de información filtrada presente EpiTex para epítomos de células T de SARS-CoV-2 mediante la construcción de un código de Python, pandas y SQL.....	<b>32</b>
Tabla 3	Cantidad de epítomos asociados a restricción HLA con respectiva cobertura poblacional en Chile de acuerdo con AFND ( <a href="http://www.allelefreqencies.net/">http://www.allelefreqencies.net/</a> ). ....	<b>44</b>
Tabla 4	Desglose de clústeres de epítomos de células T inmunodominantes de cobertura poblacional superior a 90% ( <i>hotspots</i> de células T CD8+) predichos por Predivac-3.0 en proteínas estructurales de SARS-CoV-2 para la población de Chile y Chile Mixed.....	<b>54</b>
Tabla 5	Desglose de clústeres de epítomos de células T CD8+ inmunodominantes de cobertura poblacional superior a 90% ( <i>hotspots</i> de células T CD8+) predichos por Predivac-3.0 en proteínas codificadas de ORF de SARS-CoV-2 para la población de Chile y Chile Mixed.....	<b>57</b>
Tabla 6	Desglose de clústeres de epítomos de células T CD4+ inmunodominantes de cobertura poblacional superior a 90% ( <i>hotspots</i> de células T CD8+) predichos por Predivac-3.0 en proteínas estructurales y codificadas de ORF de SARS-CoV-2 para la población de Chile y Chile Mixed.....	<b>60</b>
Tabla 7	Grupos étnicos registrados en poblaciones de Chile con loci HLA específico de la etnia.....	<b>62</b>
Tabla 8	Alelos HLA por continentes de acuerdo con AFND ( <a href="http://www.allelefreqencies.net/pop6001a.asp">http://www.allelefreqencies.net/pop6001a.asp</a> ). ....	<b>63</b>
Tabla 9	Poblaciones con componente étnico presentes en Chile de acuerdo con AFND....	<b>63</b>
Tabla 10	Selección de <i>hotspots</i> predichos por Predivac-3.0 para la población chilena, considerando características de densidad (clústeres), cobertura poblacional y coincidencia con epítomos descubiertos experimentalmente (EpiTex).....	<b>66</b>
Tabla 11	Estudio de variabilidad genética de proteínas de SARS-CoV-2 mediante el análisis de 329.942 secuencias presentes en GISAID extraído de Zelenova et al., 2021.....	<b>67</b>

## RESUMEN

Las enfermedades infecciosas emergentes (EIEs) representan una importante amenaza para la economía mundial y la salud pública. Factores como el aumento de la urbanización, globalización, el comercio internacional y/o el cambio climático aumentan la probabilidad de aparición de patógenos emergentes. Esto puede constatarse en la pandemia de COVID-19, así como en brotes previos del síndrome respiratorio agudo severo (SARS), síndrome respiratorio de Oriente Medio (MERS) o la reciente Viruela del mono. La vacunación es la estrategia más poderosa y efectiva para prevenir y controlar las EIEs, sin embargo, el desarrollo de vacunas contra virus emergentes enfrenta múltiples desafíos. Los genes HLA son conocidos por presentar el más alto nivel de diversidad en el genoma humano y son descritos como hiperpolimórficos. Las personas poseen diferentes conjuntos de alelos HLA con especificidades de ligando variables y frecuencias de expresión que difieren sustancialmente, incluso entre etnias, siendo un factor determinante frente a la susceptibilidad o protección de grupos étnicos frente a EIE. Esta investigación abordó el desarrollo de nuevas estrategias computacionales para acelerar el desarrollo de vacunas más seguras y eficientes, particularmente siguiendo el enfoque de vacunas basadas en epítomos, cuyo diseño evita el uso de componentes virales no esenciales y potencialmente tóxicos. La metodología utilizada se basó en el uso del método Predivac-3.0 y otras herramientas inmunoinformáticas para el descubrimiento de epítomos de células T y mapeo *in silico* de miles de péptidos, como un marco para abordar racionalmente la enorme diversidad de proteínas HLA considerando el componente étnico de proteínas HLA en poblaciones objetivo definidas. En esta tesis se investigó la presencia y distribución de epítomos de células T CD4+ y CD8+ en el proteoma de una variante de SARS-CoV-2 presente en Chile. Para el análisis y comparación de los resultados obtenidos con Predivac-3.0 se construyó una base de datos de epítomos de células T de SARS-CoV-2 determinados experimentalmente (denominada EpiTex). El repositorio construido almacenó 1.210 epítomos de células T de SARS-CoV-2, con 396 epítomos ubicados en proteínas estructurales y 814 epítomos en proteínas codificadas por ORF. En EpiTex también se almacenaron 1.571 restricciones HLA asociadas a estos 1.210 epítomos, de las cuales, 1.469 restricciones se encuentran repartidas en 35 alelos HLA clase I y 102 restricciones repartidas en 15 alelos HLA clase II. Predivac-3.0 identificó una serie de clústeres de epítomos de células T con potencial cobertura poblacional superior al 90% (*hotspots*), para las poblaciones de Chile y Chile Mixed. Chile Amerindian, a pesar de ser una población de componente étnico reconocido por Predivac-3.0, no presentó clústeres que cumplieran con el criterio para ser considerado *hotspots*. Las proteínas S, N, M, E, ORF1ab, ORF7a, ORF8, ORF10 y ORF6 presentaron potenciales *hotspots* que presentan una gran cantidad de epítomos presentes en EpiTex, dando lugar a un total de 22 secuencias enriquecidas en epítomos de células T fueron determinadas experimentalmente. Es posible concluir que los ORF6, ORF10 y E son regiones de interés para el desarrollo de una vacuna basada en epítomos escogidos especialmente para la población de Chile, considerando la distribución de los epítomos de células T en los *hotspots* predichos y la variabilidad reportada en las secuencias aminoacídicas de las proteínas de SARS-CoV-2. Estas regiones son potencialmente inmunodominantes para la población chilena y candidatos para etapas posteriores de validación experimental conducente al desarrollo de vacunas de nueva generación.

## ABSTRACT

Emerging infectious diseases (EIDs) represent a major threat to the global economy and public health. Factors such as increased urbanization, world travel, international trade, and/or climate change increase the likelihood of emergence of emerging pathogens. This can be seen in the SARS-CoV-2 pandemic, the recent outbreaks of severe acute respiratory syndrome (SARS), the Middle East respiratory syndrome (MERS) or the recent Monkeypox. Vaccination is the most powerful and effective strategy to prevent and control EIDs, however, the development of vaccines against emerging viruses faces multiple challenges. HLA genes are known to have the highest level of diversity in the human genome and are described as "hyperpolymorphic". People have different sets of HLA alleles with variable ligand specificities and expression frequencies that differ substantially even between ethnic groups, being considered as a factor against the vulnerability or protection of ethnic groups against EIDs. This research addressed the development of new computational strategies leading to quicker, safer and more efficient vaccines. Epitope-based vaccination allows for the rational design of immunogens based on short regions or peptides, which avoid non-essential and potentially toxic viral components. Immunoinformatics tools play an important role in accelerating epitope discovery and vaccine design through *in silico* mapping of thousands of peptides, as well as offering a framework to rationally address the enormous diversity of HLA proteins. In this research, the Predivac-3.0 tool was used to predict CD4+ and CD8+ T-cell epitopes of a variant of SARS-CoV-2 present in Chile. Predivac-3.0 is an immunoinformatics tool providing a computational framework for rational vaccine design based on multiple T-cell epitopes, while considering the ethnic background of HLA proteins in defined target populations. The presence and distribution of CD4+ and CD8+ T cell epitopes in the proteome of a SARS-CoV-2 variant present in Chile was investigated. For the analysis and comparison of the results obtained with Predivac-3.0, a database of experimentally determined SARS-CoV-2 T cell epitopes (called EpiTex) was built. This repository stored 1,210 SARS-CoV-2 T cell epitopes, with 396 epitopes located on structural proteins and 814 epitopes on ORF-encoded proteins. In EpiTex, 1,571 HLA restrictions associated with these 1,210 epitopes were also stored, of which 1,469 restrictions are distributed in 35 HLA class I alleles and 102 restrictions distributed in 15 HLA class II alleles. Predivac-3.0 identified a series of T cell epitope clusters with potential population coverage greater than 90% (hotspots), for the populations of Chile and Chile Mixed. Chile Amerindian, despite being a population of ethnic background recognized by Predivac-3.0, did not present clusters that met the criteria to be considered hotspots. The S, N, M, E, ORF1ab, ORF7a, ORF8, ORF10 and ORF6 proteins presented hotspots presenting a large number of EpiTex epitopes within them, giving a total of 22 experimentally determined T cell epitope-enriched amino acid sequences. It is possible to conclude that ORF6, ORF10 and E are promising regions for the development of vaccines based on T-cell epitopes specially selected for the Chilean population, by considering the distribution of epitopes in the predicted hotspots and the variability reported in the amino acid sequences of SARS-CoV-2 proteins. These regions are thus potentially immunodominant for the Chilean population and candidates for later stages of experimental validation leading to the development of new generation vaccines.

## 1. INTRODUCCIÓN

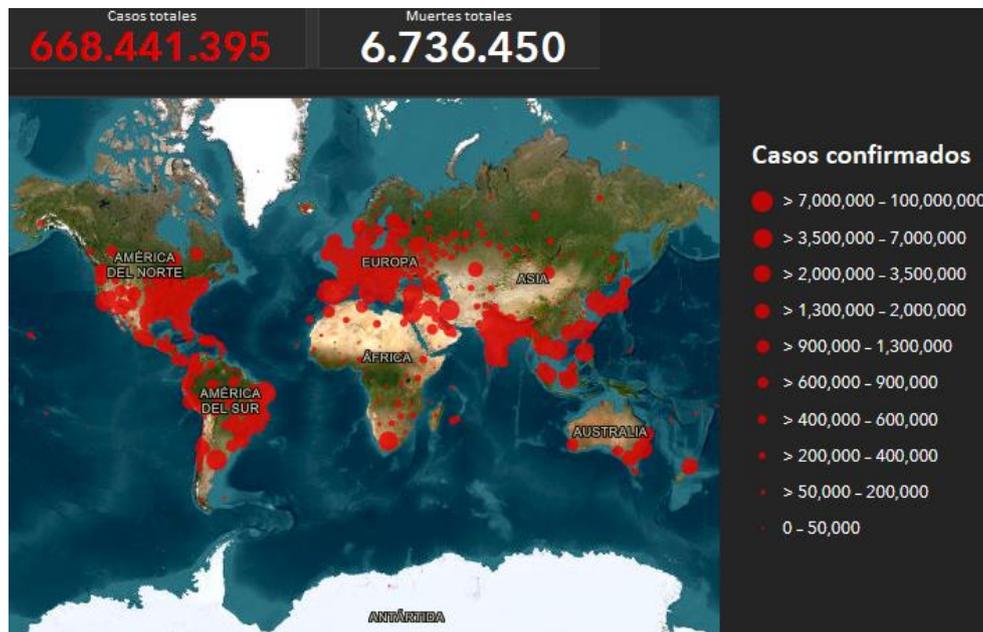
### 1.1. Enfermedades infecciosas emergentes

Las enfermedades infecciosas emergentes (EIEs) corresponden a enfermedades que han aparecido recientemente en una población o que han existido previamente, pero que rápidamente incrementan en incidencia y/o alcance geográfico (Morens et al., 2004). Algunas EIEs son causadas por especies de patógenos (organismos causantes de enfermedad) recién identificadas o patógenos que afectan a una nueva población. Las infecciones reemergentes como el sarampión o la tuberculosis resistente a los medicamentos también pertenecen a las EIEs (Tang et al., 2018). La infección se produce cuando un patógeno invade y comienza a crecer dentro de un hospedero. La enfermedad se produce si como consecuencia de la invasión y el crecimiento de un patógeno se altera la función tisular (National Institutes of Health (US), 2007).

A fines de diciembre en 2019, en Wuhan, provincia de Hubei, China, se reportaron los primeros casos de neumonía grave en adultos que sitúan el origen de la pandemia COVID-19 (coronavirus disease-2019) causada por una nueva variedad de  $\beta$ -coronavirus, denominado SARS-CoV-2 (Cui et al., 2019; Zepeda et al., 2020). Al igual que con SARS-CoV y MERS-CoV, este  $\beta$ -coronavirus está estrechamente relacionado con SARS-CoV-2 en murciélagos, quienes son considerados principales reservorios de  $\beta$ -coronavirus y especialmente de virus relacionados con SARS-CoV (Schlottau et al., 2020). Se considera a los murciélagos como reservorios ideales para virus, ya que son los únicos mamíferos que vuelan, lo que ayuda a la dispersión del virus; albergan en gran número, ayudando así a los ciclos de transmisión; algunos hibernan durante el invierno, proporcionando así un mecanismo para que los virus persistan entre temporadas; y los factores genéticos pueden desempeñar un papel en la capacidad de los murciélagos para albergar virus sin provocar una enfermedad clínica (Mackenzie et al., 2016). Actualmente, se postula que SARS-CoV-2 llegó a los humanos vía transmisión zoonótica por pangolines malayos (*Manis javanica*) importados ilegalmente a la provincia de Guangdong, quienes recibieron el virus a través de murciélagos. Los pangolines portan coronavirus (similares a SARS-CoV-2), cuyo dominio de unión a receptor (RBD) exhibe una fuerte similitud con RBD de SARS-CoV-2 (Andersen et al., 2020; Zhang et al., 2020).

De acuerdo con el Centro de Recursos de Coronavirus de la Universidad Johns Hopkins, para el 20 de Enero de 2023, SARS-CoV-2 registra una cantidad total de 668.441.395 casos confirmados, con un total de 6.736.450 muertes producto de COVID-19 (figura 1), demostrando

la gran necesidad de invertir en investigaciones para prevención, control de infección y manejo de EIE (Bourouiba, 2021).



**Figura 1** Mapa de registro total de casos confirmados de contagio con SARS-CoV-2 adaptado del centro de recursos de coronavirus de la universidad de medicina Johns Hopkins (<https://coronavirus.jhu.edu/map.html>) actualizado para el 20 de Enero de 2023.

La mayoría de las EIEs son zoonóticas, vale decir, se originan en la fauna silvestre por transmisión entre animales y humanos. Aunque la mayoría de estas enfermedades están asociadas a bacterias, los patógenos virales (especialmente los virus de ARN) representan una amenaza mayor debido a sus altas tasas de mutación y capacidad de adaptación a nuevos hospederos, incluido humanos (Jones et al., 2008). Es común que en brotes de EIEs exista un alto grado de desconocimiento sobre el patógeno, los mecanismos de transmisión y sobre la inmunología asociada a infección. Es decir, existe poco grado de conocimiento sobre la variabilidad antigénica y la respuesta inmunitaria a la infección viral, cuya información es necesaria para el desarrollo de vacunas contra un virus emergente o una enfermedad infecciosa que puede surgir en el futuro.

El desarrollo de nuevas EIEs puede verse favorecido por la mutación genética, recombinación viral genética, los cambios en la población de hospederos, ambientales y/o comportamiento humano, especialmente los relacionados con la globalización, movilidad y urbanización en un contexto de creciente globalización (Medicine, 2003). Esto ha facilitado las vías de transmisión de patógenos entre centros urbanos alrededor del mundo (Gubler, 2011). El

incremento de esta interconectividad hace que las EIEs sean una profunda amenaza latente a la salud humana, además del impacto sobre la economía y la sociedad en general (Halliday et al., 2017). El desarrollo de la pandemia COVID-19 y otros brotes previos de SARS, síndrome respiratorio de Oriente Medio (MERS) o influenza H5N1 y H1N1, son ejemplos que demuestran que actualmente existen factores o condiciones que dan forma a una época con tasas especialmente elevada de EIEs (Bourouiba, 2021).

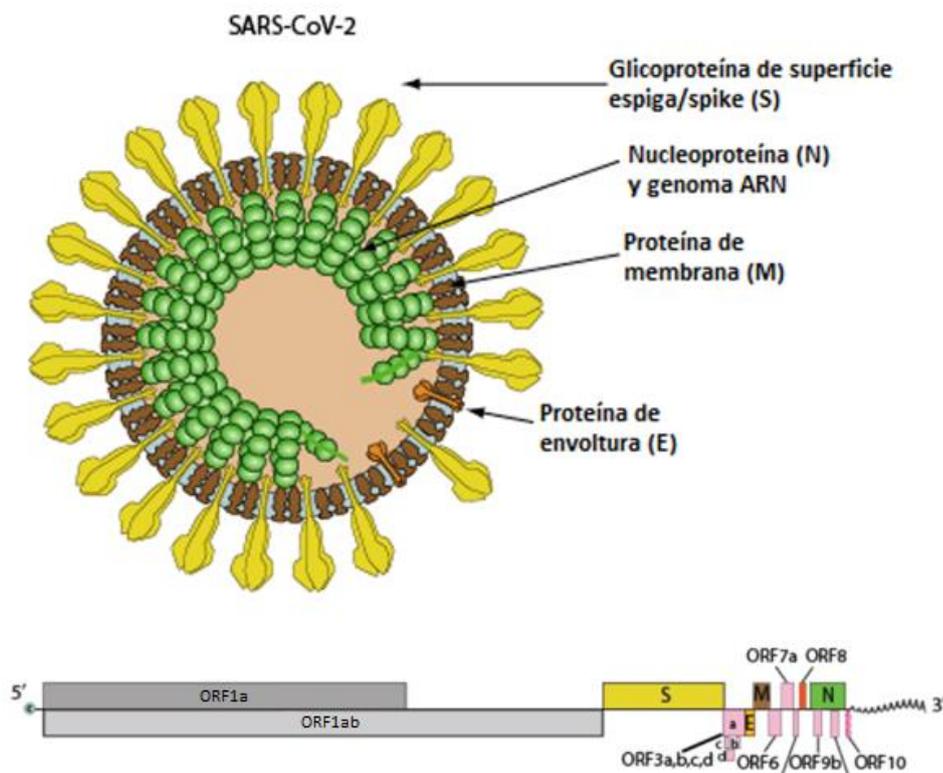
Las EIEs son típicamente difíciles de estudiar, en parte debido a su naturaleza intrínsecamente transitoria y repentina, lo que dificulta probar hipótesis y determinar estrategias de mitigación verdaderamente exitosas, especialmente, cuando los primeros casos de infectados implican un tamaño de muestra poblacional que no permite resultados estadísticos de alta relevancia (Sanchez-Mazas, 2020). Cuando un virus emerge en un área libre de infección, o en una nueva especie, es probable que la población hospedera sea susceptible y tenga poca o ninguna inmunidad preexistente al patógeno. La falta de inmunidad colectiva puede resultar en una diseminación rápida y en consecuencias más virulentas de la infección (García-Sastre & Mena, 2013).

Los sistemas de salud en todo el mundo no pudieron proporcionar servicios esenciales, incluso en países altamente desarrollados, contribuyeron a la transmisión de enfermedades durante epidemias de SARS-CoV-2 y MERS. Los países han comprometido recursos y esfuerzos hacia el fortalecimiento del sistema de salud, pero aún no han logrado consenso (Nuzzo et al., 2019). Por ejemplo, para el año 2020 aún persistían dudas similares a las discutidas durante la influenza de 1918 y la pandemia de SARS de 2003, tales como distancias de interacciones seguras, tipos de mascarillas o flujos de aire recomendables en espacios interiores (Bourouiba, 2021). Existe consenso que, para prevenir y controlar las EIEs es esencial brindar inmunidad protectora a la población a través de la vacunación como estrategia más eficaz y rentable (Wong & Qiu, 2018; García-Sastre & Mena, 2013).

## **1.2. SARS-CoV-2**

SARS-CoV-2 es el nombre de del betacoronavirus causante de la pandemia de COVID-19. Este virus pertenece al mismo género de coronavirus que SARS-CoV y MERS-CoV, ambos causantes de SARS y MERS en los años 2003 y 2012, respectivamente (Li et al., 2020). Ambos comparten una alta homología con coronavirus tipo SARS aislado de murciélagos, sugiriendo un origen común en este reservorio (Zhou et al., 2020).

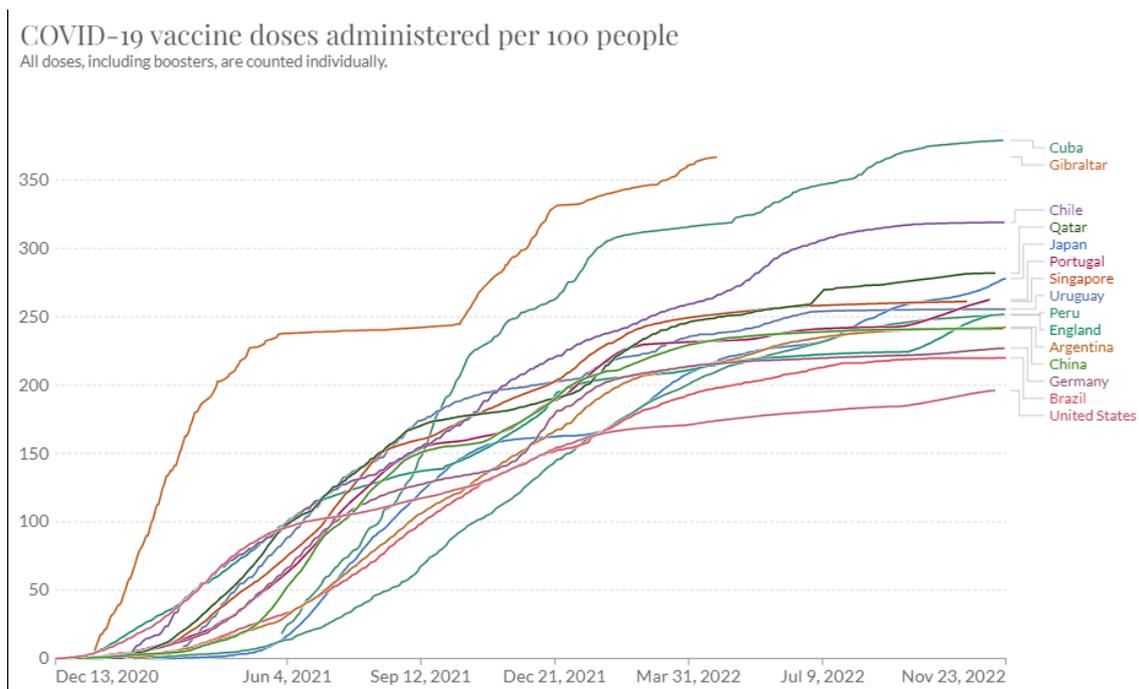
SARS-CoV-2 posee un genoma de 30kb de tamaño, el cual codifica una gran poliproteína no estructural (ORF1ab) que abarca 2/3 de la región genómica viral de SARS-CoV-2 que codifica proteínas no estructurales referidas como poliproteínas pp1a y pp1ab. ORF1ab pasa por clivaje proteolítico para generar 15 de las 16 proteínas no estructurales (NSP) (Yadav et al., 2021). Además, el genoma de SARS-CoV-2 abarca 4 proteínas estructurales y 5 proteínas accesorias (ORF3a, ORF6, ORF7, ORF8 y ORF9) (figura 2) (Chan et al., 2020; Ramaiah & Arumugaswami., 2020). Las cuatro proteínas estructurales consisten en (i) la proteína de membrana (M); (ii) la proteína de envoltura (E); (iii) la proteína de nucleocápside (N) y (iv) la glicoproteína de superficie (S). Esta última es responsable de interaccionar con enzima convertidora de angiotensina 2 (ACE2) para acceder a las células objetivo, lo cual se produce con una afinidad aproximadamente 10-20 veces mayor que SARS-CoV (Wrapp et al., 2020; Walls et al., 2020; Li et al., 2020; Cui et al., 2019). En el ser humano un 83% de las células que expresan ACE2 son células epiteliales alveolares tipo II, motivo por el cual los pulmones son un órgano objetivo de estos virus (Zhao et al., 2020).



**Figura 2** Esquema de proteoma y estructura de SARS-CoV-2 con proteínas estructurales y genoma de ARN extraído de los recursos de coronavirus de ViralZone, SIB Swiss Institute of Bioinformatics (<https://viralzone.expasy.org/9056>).

### 1.2.1. COVID-19: situación en Chile

Chile fue una de las naciones más exitosas del mundo en términos de implementación de un programa de vacunación contra COVID-19. Aproximadamente el 92% de la población chilena estaba completamente vacunada (dosis de refuerzo incluidas) en Abril de 2022 (Figura 3) (Toro-Ascuy et al., 2022). La primera vacuna aprobada en Chile para uso de emergencia fue la de Pfizer (BNT162b2) el 16 de Diciembre de 2020, en tanto que la segunda vacuna aprobada por las autoridades chilenas para uso de emergencia fue la vacuna CoronaVac con virión inactivado, el 20 de Enero de 2021. Poco después de la aprobación de la vacuna CoronaVac, la autorización de uso de emergencia de Oxford-AstraZeneca (ChAdOx1), Johnson & Johnson (Ad26.COV2), CanSino (Ad5-nCoV) y Sputnik V (GAM-COVIDVac) (Aguilera et al., 2021; Toro-Ascuy et al., 2022). En la tabla 1 se anexa el reporte de vacunación en Chile, actualizado para el 31 de Agosto de 2022.



**Figura 3** Dosis totales acumulativas de vacunas administradas por cada 100 personas a lo largo del tiempo desde el 13 de Diciembre de 2020 hasta el 23 de Noviembre de 2022, adaptado de “Our World in Data” (<https://ourworldindata.org/covid-vaccinations>) (Mathieu et al., 2021).

Región	Población Objetivo	1° Dosis	2° Dosis	Única	Refuerzo	4° Dosis	% Avance vacunación	% Cobertura vacunación	% Vacunación Refuerzo	% Vacunación 4 Dosis
<b>TOTAL</b>	<b>18.972.800</b>	<b>17.478.899</b>	<b>17.081.548</b>	<b>574.698</b>	<b>15.557.673</b>	<b>11.211.554</b>	<b>95,2%</b>	<b>93,1%</b>	<b>88,1%</b>	<b>72,1%</b>
Ignorada	0	44	42	0	41	26			97,6%	63,4%
Metropolitana de Santiago	7.934.198	6.906.610	6.752.336	224.902	6.180.846	4.572.253	89,9%	87,9%	88,6%	74,0%
Antofagasta	674.806	630.958	613.446	20.927	557.446	398.438	96,6%	94,0%	87,9%	71,5%
Biobío	1.615.246	1.517.688	1.488.441	43.867	1.362.987	989.612	96,7%	94,9%	88,9%	72,6%
Valparaíso	1.914.851	1.815.351	1.774.430	51.687	1.605.990	1.156.374	97,5%	95,4%	87,9%	72,0%
Arica y Parinacota	245.872	229.606	221.113	10.259	199.312	136.467	97,6%	94,1%	86,1%	68,5%
Coquimbo	818.247	784.518	766.322	24.294	691.690	488.378	98,8%	96,6%	87,5%	70,6%
Tarapacá	373.639	350.420	332.948	19.027	296.171	198.043	98,9%	94,2%	84,1%	66,9%
La Araucanía	983.931	944.252	923.336	28.983	822.135	549.563	98,9%	96,8%	86,3%	66,8%
Los Lagos	866.664	826.005	806.063	38.155	723.886	488.962	99,7%	97,4%	85,7%	67,5%
Maule	1.103.226	1.073.876	1.052.208	31.317	960.051	689.959	100,2%	98,2%	88,6%	71,9%
Magallanes y de la Antártica Chilena	174.128	159.471	156.909	15.020	157.061	118.368	100,2%	98,7%	91,4%	75,4%
Los Ríos	394.591	383.327	375.282	13.253	338.052	236.901	100,5%	98,5%	87,0%	70,1%
Atacama	304.367	300.549	291.804	7.241	259.762	177.944	101,1%	98,3%	86,9%	68,5%
Libertador General Bernardo O'Higgins	966.755	962.628	942.731	23.784	858.089	622.659	102,0%	100,0%	88,8%	72,6%
Ñuble	498.497	498.032	490.513	11.354	453.580	326.168	102,2%	100,7%	90,4%	71,9%
Aysén del General Carlos Ibáñez del Campo	103.782	95.564	93.624	10.628	90.574	61.439	102,3%	100,5%	86,9%	67,8%

**Tabla 1** Reporte de cantidad de personas vacunadas en Chile actualizado para el 31 de Agosto de 2022 de acuerdo con el MINSAL ([https://informesdeis.minsal.cl/SASVisualAnalytics/?reportUri=%2Freports%2Freports%2F9037e283-1278-422c-84c4-16e42a7026c8&sectionIndex=0&sso\\_guest=true&reportViewOnly=true&reportContextBar=false&sas-welcome=false](https://informesdeis.minsal.cl/SASVisualAnalytics/?reportUri=%2Freports%2Freports%2F9037e283-1278-422c-84c4-16e42a7026c8&sectionIndex=0&sso_guest=true&reportViewOnly=true&reportContextBar=false&sas-welcome=false)).

### 1.3. Vacunas

El control o eliminación de una enfermedad requiere la inducción de inmunidad protectora en una proporción alta de la población. Esto se logra mediante programas de inmunización capaces de inducir protección a largo plazo. La inmunidad a largo plazo es conferida por el mantenimiento de efectos inmunitarios específicos de antígeno y/o por la inducción de células de memoria inmunológica que pueden ser lo suficientemente eficientes y reactivarse rápidamente en efectos inmunitarios en caso de exposición a patógenos (Plotkin et al., 2012).

Las vacunas son preparaciones biológicas que inducen inmunidad adquirida activa a una enfermedad en particular. Típicamente consisten en una forma atenuada o muerta del organismo causante de enfermedad (patógeno) que estimula al sistema inmune, permitiéndole reconocer posteriormente al mismo patógeno para eliminarlo y guardar un registro de él para facilitar el reconocimiento y erradicación de cualquier microorganismo idéntico que encuentre después (Robinson, 2016). Los efectos inmunes inducidos por vacunas se manifiestan a través de la producción de anticuerpos (producidos por linfocitos B) capaces de unirse específicamente a una toxina o patógeno, además de otros efectos potenciales por parte de linfocitos T CD8+ y linfocitos T auxiliares CD4+. Los linfocitos T son esenciales para la inducción de anticuerpos de alta afinidad y memoria inmunológica (Rich et al., 2019). Los determinantes de la inducción de la memoria inmunológica, así como la contribución relativa de los anticuerpos persistentes y de la memoria inmunológica en la protección contra enfermedades específicas, son parámetros esenciales de la eficacia de las vacunas a largo plazo (Plotkin et al., 2012).

Las vacunas cumplen una contribución clave a la salud pública. A diferencia de medicamentos o fármacos cuyo beneficio está restringido al individuo que los consume, las vacunas tienen el potencial de tener efectos de largo alcance que abarcan la salud, bienestar general, desarrollo cognitivo y, en última instancia, la productividad económica (Doherty et al., 2016). El proceso completo para llevar una vacuna desde un laboratorio de investigación al público es largo, complejo y caro (Wong & Qiu, 2018). La fase inicial de investigación puede durar entre 1 a 5 años, seguido de fases clínicas y preclínicas que puede durar en promedio 15 a 20 años, antes de poder obtener una licencia (Artaud et al., 2019). Las fases preclínicas incluyen caracterización del producto, estudios de prueba de concepto (inmunogenicidad) y pruebas de seguridad en animales realizadas antes de introducir el producto en humanos (WHO,

2003). El propósito de un estudio preclínico en animales es asegurar los efectos farmacológicos y toxicológicos de una nueva vacuna, antes y en paralelo a un desarrollo clínico (Prescott & Lidster, 2017). Una vez aprobado, procede el desarrollo clínico para demostrar seguridad, inmunogenicidad y eficacia de la vacuna.

Las fases de evaluación clínica de una vacuna nueva se dividen generalmente en fase I, fase II y fase III. Estas pueden superponerse y múltiples ensayos de fase I o fase II pueden realizarse a medida se obtiene nueva información (Artaud et al., 2019). La fase I consta de una evaluación preliminar de la seguridad, inmunogenicidad y tolerabilidad de múltiples niveles de dosis de una vacuna candidato en un número pequeño de voluntarios adultos monitoreados. La fase II consiste en generar mayor información sobre seguridad e inmunogenicidad de la vacuna en una cantidad de participantes superior a los involucrados en la fase I. La fase II contiene 2 fases: IIa y IIb. En la fase IIa el estudio se enfoca en la inmunogenicidad, dosis y formulación. La fase IIb consiste en estudios diseñados para ensayos de eficacia de prueba de concepto (Artaud et al., 2019). Finalmente, los ensayos de fase III consisten en estudios terapéuticos confirmatorios en una población más grande y heterogénea, para demostrar uso clínico y establecer perfil de seguridad (Demets et al., 2010).

La investigación en vacunas contra EIEs tiene por finalidad proporcionar un marco teórico y experimental para acelerar el desarrollo de inmunidad protectora y de largo plazo para prevenir y controlar este tipo de enfermedades (García-Sastre & Mena, 2013), tal como se ha demostrado en la erradicación de la viruela, la reducción de polio, los progresos hacia la eliminación del tétano o en la reducción de la mortalidad a causa de sarampión (Serdobova & Kieny, 2006). Sin embargo, el desarrollo de una vacuna contra un patógeno emergente enfrenta varios desafíos. Aunque las estrategias de vacunación convencionales han sido fundamentales en el control y erradicación de algunas enfermedades infecciosas importantes, en muchos otros casos no logran alcanzar los niveles requeridos de inmunogenicidad, seguridad, protección cruzada a través de la variabilidad antigénica de los patógenos, o incluso exacerbar la enfermedad. Por este motivo, el desarrollo de vacunas para inmunizar contra patógenos emergentes es un campo de investigación activo para mejorar la respuesta a EIEs mediante nuevas aproximaciones más rápidas que incorporando tecnologías modernas y conceptos de diseño racional de inmunógenos (García-Sastre & Mena, 2013).

En base al tipo de material antigénico para la vacunación, se definen diferentes generaciones de vacunas (Alarcon et al., 1999).

### **1.3.1. Vacunas de primera generación**

Las vacunas de primera generación se basan en organismos completos, ya sea vivos atenuados o muertos (Alarcon et al., 1999). Tienen origen histórico en la primera vacuna desarrollada por Edward Jenner, quien inoculó viruela bovina contra la viruela en pacientes cuya mortalidad disminuyó de un 30-40% a un 1% (Minor, 2015). Si bien esta estrategia ha sido exitosa en casos como la erradicación de la viruela en 1980 (Fenner et al., 1988), la cercana erradicación global de la poliomielitis (Bart et al., 1996) y en el control del sarampión en gran parte del mundo (Griffin, 2018). Sin embargo, existen casos donde la seguridad y eficacia es cuestionable y pueden ocurrir eventos adversos asociados a reversión del patógeno. Por esta razón, una vacuna de primera generación contra el VIH o contra cualquier otro patógeno emergente que cause una enfermedad letal es problemática, ya que, al utilizar una vacuna viva o atenuada, no hay forma de controlar todo lo que puede implicar en el paciente (Minor, 2015). Por otra parte, también existe el riesgo de bioseguridad en la producción, ya que estas vacunas requieren el cultivo de grandes lotes de virus patógenos competentes para la replicación antes de la inactivación. Estos lotes pueden verse obstaculizados por factores tales como el cultivo difícil o imposible de los respectivos patógenos en condiciones *in vitro* o el requisito de un alto nivel de bioseguridad y laboratorios especializados para el cultivo (Rauch et al., 2018). El requisito de muestras virales para el desarrollo y fabricación de vacunas implica tiempo y la necesidad de instalaciones de alta contención (Sandbring & Koblenz, 2022). Por lo tanto, para abordar nuevos brotes de manera efectiva y rápida, se requieren enfoques nuevos y versátiles que sean independientes del cultivo completo de patógenos (Rauch et al., 2018).

### **1.3.2. Vacunas de segunda generación**

Las vacunas de segunda generación consisten en antígenos purificados (proteínas) identificados por ser inmunogénicos y que derivan del organismo, en vez de usar el mismo organismo completo (Alarcon et al., 1999; Boballa & Hook, 2016). Estos antígenos son menos reactogénicos (es decir, causan menos reacciones/efectos secundarios), produciendo bajos efectos adversos locales y sistémicos, pero también desventajas cuando se trata de estimular respuestas inmunes potentes y duraderas. Por este motivo, este tipo de vacunas son coadministradas con adyuvantes para activar y modular respuestas inmunes efectivas (Tandrup

et al., 2016), requiriendo administrar múltiples dosis de vacuna para desarrollar respuestas de protección duradera (Boballa & Hook, 2016).

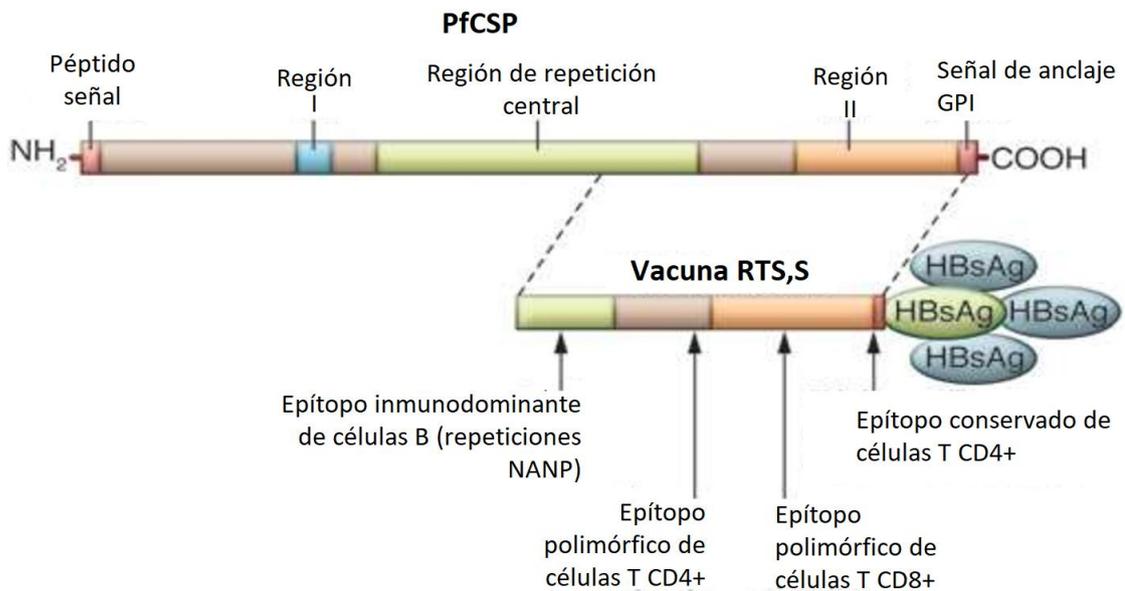
El descubrimiento de que antígenos superficiales del virus de la hepatitis B (HBsAg) que se encontraba en personas infectadas, son inmunogénicas y protectoras, pero no infecciosas, proporcionó la base para los esfuerzos por purificar estas partículas de la sangre de los portadores crónicos. Esto permitió la licencia de una vacuna derivada de plasma en Estados Unidos en 1981. Sin embargo, los productos derivados de la sangre humana se consideraban potencialmente peligrosos. Además, debido a que la vacuna dependía del suero humano, la fuente de antígeno era limitada. Estos obstáculos impulsaron la formulación de la primera vacuna recombinante mediante la clonación del gen de HBsAg en levadura (*Saccharomyces cerevisiae*) y en células de mamífero. Por lo tanto, debido a que se deriva de un gen, disminuye el riesgo de una posible contaminación con agentes extraños no detectados (Plotkin et al., 2012). Sin embargo, aunque las vacunas de proteínas recombinantes gatillan respuesta inmune humoral, la respuesta que gatillan en respuesta inmune celular es débil (Nascimento & Leite., 2012).

Las vacunas diseñadas como partículas similares a virus (VLPs) son consideradas vacunas de segunda generación. Las VLPs se basan en virus envueltos y no envueltos que se usan para inmunizar contra virus homólogos, o se diseñan mediante ingeniería genética para incorporar regiones de proteínas o epítopos de un patógeno diferente (García-Sastre & Mena, 2013). La ausencia de un genoma viral los convierte en vacunas más seguras, al no existir la posibilidad de reversión a virulencia. Las VLPs son estructuralmente diversas, teniendo una o múltiples proteínas de cápside, o una envoltura lipídica. Aunque no todos los virus son candidatos adecuados para VLP, aquellos que lo son, pueden generar una respuesta protectora sin requerir múltiples dosis de refuerzo, reduciendo significativamente el costo de la vacuna (Afrough et al., 2019).

#### **1.3.2.1. Vacuna RTS,S/AS01**

RTS,S/AS01 es una vacuna contra la malaria diseñada a base de VLPs como parte de un sistema adyuvante liposomal (AS01). La vacuna RTS,S/AS01 consta de dos componentes: (i) una proteína que contiene 18 copias de la repetición central de cuatro aminoácidos (NANP) de la proteína circumsporozoito principal (PfCSP) y (ii) el dominio C-terminal de la proteína PfCSP fusionada con el antígeno de superficie del virus de la hepatitis B (HbsAg) (Draper et al., 2018;

Laurens, 2020). El antígeno blanco, la proteína circumsporozoito, es una proteína de 412 aminoácidos, abundantemente asociada con la superficie del esporozoito, también expresada por formas hepáticas y exportada en el citoplasma de hepatocitos. La proteína PfCSP comprende una región N-terminal que contiene una secuencia de péptido señal y una región que se une a proteoglicanos de sulfato de heparina (región I) y tiene incrustada en su interior una secuencia conservada del sitio de escisión proteolítica de cinco aminoácidos (KLKQP); una región central que contiene repeticiones NANP; una región C-terminal (región II) que contiene un dominio similar a la trombospondina (TSP); y una secuencia canónica ancla de adición de glicosilfosfatidilinositol (GPI) (figura 4) (Kaslow & Biernaux, 2015). Esta proteína desempeña un papel central en la entrada del parásito en el tejido del hígado, regulando negativamente la inflamación mediada por NF-KB y creando así un nicho favorable para el parásito que permite el desarrollo de la etapa hepática (Vekemans et al., 2009; Oyen et al., 2017).



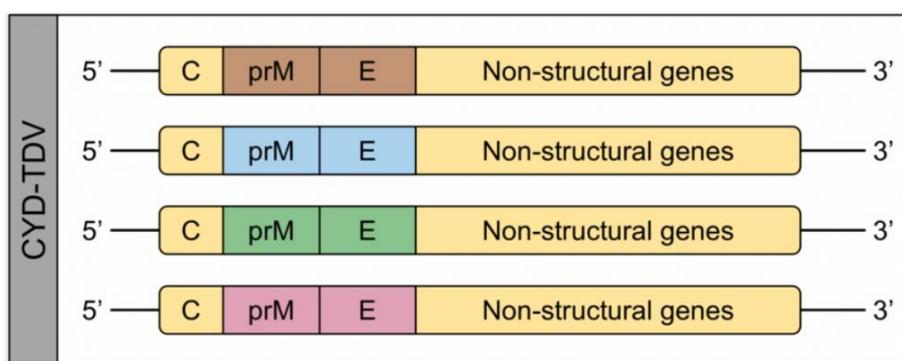
**Figura 4** Representación gráfica de estructuras de PfCSP y vacuna RTS,S (modificado de Lubanga et al., 2016). Los monómeros del HbsAg se auto ensamblan en VLP y aproximadamente el 25% de los monómeros HbsAg en RTS,S se recombinan genéticamente para el PfCSP truncado y sirven como transportadores de proteínas. El fragmento PfCSP en RTS,S contiene tres epítopos de células T conocidos: un epítipo de células T CD4+ altamente variable, un epítipo de células T CD8+ altamente variable y un epítipo de células T CD4+ conservado en el extremo C-terminal.

La vacuna RTS,S/AS01 ha sido la única en lograr eficacia protectora contra malaria en fase III de ensayos clínicos (Draper et al., 2018), donde redujo la incidencia de la enfermedad en cerca de un 51% (durante los primeros 14 meses) en niños que tenían entre 5 y 17 meses cuando recibieron la primera de tres dosis (Oyer et al., 2017). Sin embargo, la protección de RTS,S que

se consigue es parcial, disminuyendo con el tiempo y puede depender de la edad del paciente (Draper et al., 2018). Aun así, esta vacuna fue aprobada por la OMS en 2021, quien aboga por su integración en los programas de inmunización de rutina y las medidas existentes de control de la malaria (Nadeem et al., 2022)

### 1.3.2.2. Vacuna Dengvaxia (CYD-TDV)

El Dengue es un flavivirus identificado como una de las 10 amenazas para la salud mundial en 2019 por la OMS, convirtiéndose en un importante problema de salud pública durante las últimas décadas (Redoni et al., 2020). CYD-TDV (Dengvaxia) es la primera vacuna contra el Dengue en ser licenciada el 2015. Usando técnicas de ADN recombinante, el gen no estructural de la cepa de vacuna contra la fiebre amarilla 17D (CYD) se combinó con los genes estructurales de premembrana (prM) y envoltura (E) de cuatro serotipos de virus del Dengue (DENV) para inducir una respuesta inmune específica (figura 5). Dengvaxia carece de las principales proteínas inmunogénicas del virus del dengue que inducen respuestas celulares (NS3 y NS5), y las respuestas inmunitarias celulares CD4+ y CD8+ específicas inducidas por el virus de la fiebre amarilla tienen una reactividad cruzada limitada. Por lo tanto, Dengvaxia desencadena principalmente respuestas inmunitarias humorales en lugar de respuestas inmunitarias celulares (Bonaldo et al., 2014; Ma & Cheng, 2022). Aun cuando la vacuna incorpora cuatro virus quiméricos del dengue (figura 5) para proveer inmunidad balanceada contra todos los cuatro serotipos del virus, esta vacuna solo puede ser suministrada a niños de 9 años o más, dejando sin cobertura a individuos más jóvenes (Redoni et al., 2020).



**Figura 5** Esquema gráfico de vacuna CYD-TDV (adaptado de Wilken & Rimmelzwaan, 2020). CYD-TDV emplea la cepa de la vacuna YFV-17D (mostrada en amarillo) como base genética para la expresión de los genes prM y E de DENV1 (marrón), DENV2 (azul), DENV3 (verde) y DENV4 (rosa).

### **1.3.3. Vacunas de tercera generación**

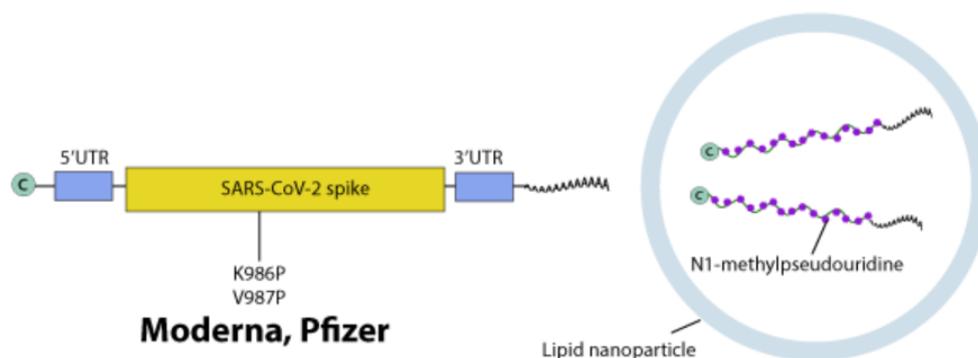
La tercera generación de vacunas son vacunas basadas en ácidos nucleicos basados en el material genético del organismo infeccioso (Alarcon et al., 1999). En el caso de las vacunas de ADN, éstas se basan en plásmidos bacterianos diseñados genéticamente y que codifican antígenos cuya expresión está bajo el control de promotores eucariotas de alta eficiencia (Rajčáni et al., 2005). A diferencia de las vacunas proteicas, las vacunas de ADN ofrecen la posibilidad de acceder al citoplasma de las células blanco. En este punto inducen la expresión del antígeno nativo (expresión endógena), optimizando el procesamiento y presentación al sistema inmune ya que permiten una eficiente interacción del antígeno con las proteínas de los antígenos leucocitarios humanos (HLA) (Li & Petrovsky, 2016). Por lo tanto, la formulación de este tipo de vacunas se hace de manera que desencadena la transcripción del gen utilizado para la vacuna en la propia maquinaria celular del hospedero (Silveira et al., 2021). Este enfoque induce respuestas inmunitarias tanto humorales como mediadas por células T (respuesta inmune celular) de manera eficiente (Silveira et al., 2021). La ventaja de las vacunas de ADN por sobre las vacunas tradicionales radica en su sencillo diseño mediante técnicas estándar de biología molecular, requiriendo la clonación de los antígenos y/o epítopos en un vector plasmidial, reduciendo costos y tiempo de producción (Li & Petrovsky, 2016).

En lo que respecta a seguridad, la clonación o síntesis de ácidos nucleicos en lugar de tener que purificar proteínas de patógenos, evita la necesidad de utilizar microorganismos patógenos en la fabricación de vacunas (Li & Petrovsky, 2016). Sin embargo, se demostró en varios estudios la persistencia a largo plazo de los plásmidos de ADN tras la inyección. En este contexto, la presencia de información genética ajena en el núcleo de las células transfectadas, plantea un riesgo adicional de integración genómica en los cromosomas del hospedero, lo que puede implicar mutagénesis y oncogénesis. Aún considerado como un riesgo pequeño, esto debe tenerse en consideración en sistemas con mayor absorción de ADN (Rauch et al., 2018).

Para SARS-CoV-2, la proteína S ha sido utilizada como antígeno para todas las vacunas de ARN que se probaron en ensayos clínicos (Silveira et al., 2021).

### 1.3.3.1. Vacunas de ARNm contra SARS-CoV-2

Las vacunas desarrolladas por Pfizer, BioNTech y Moderna usan una nueva tecnología basada en ARNm, cuya secuencia nucleica codifica para el antígeno proteico. Aunque el ARN es una molécula más inestable que el ADN, se han desarrollado nuevos diseños de vacuna para mejorar su estabilidad y la eficiencia de traducción de proteínas para mejorar la respuesta inmunitaria. La vacuna mRNA-1273 (Moderna Inc. Cambridge, MA, EE. UU.) fue la primera vacuna de ARNm diseñada contra COVID-19 y estableció un hito histórico con un tiempo récord de 63 días desde la etapa de diseño de la vacuna hasta los ensayos clínicos en humanos (Silveira et al., 2021). El diseño genético de la vacuna considera el gen de la proteína S (antígeno) codificado en un plásmido de ADN, el cual se transcribe *in vitro* mediante la ARN polimerasa T7 para generar ARNm. La molécula de ARNm está poliadenilada (polimerasa poliA) y protegida mediante enzimas de protección celular o del virus vaccinia. Luego, los ARNm se purifican, un paso que puede implicar la digestión con ADNasa para eliminar el ADN del plásmido original. La proteína contiene dos mutaciones de prolina en la subunidad 2 de la proteína S (K986P y V987P) para estabilizar el trímero (Heinz & Stiasny, 2021) (figura 6). Finalmente, el ARNm purificado se encapsula en una envoltura lipídica de nanoliposomas. Los ARNm encapsulados ingresan a cualquier célula o son eliminados por células inmunitarias como las células dendríticas. En ambos casos, el ARNm que ingresa al citoplasma se traduce y el antígeno se expresa, desencadenando una respuesta inmune (Heinz & Stiasny, 2021).



**Figura 6.** Esquema del ARNm de la vacuna en las vacunas BioNTech-Pfizer y Moderna modificado de los recursos de coronavirus de ViralZone, SIB Swiss Institute of Bioinformatics (<https://viralzone.expasy.org/9696>). El gen del antígeno está codificado en un plásmido de ADN que se transcribe *in vitro* por la ARN polimerasa T7 para generar ARNm. Las uridinas se reemplazan con N1-metilpseudouridina para estabilizar la molécula. El ARNm se protege y poliadenila (poliA polimerasa). La proteína S contiene dos mutaciones de prolina en S2 (K986P y V987P) para estabilizar el trímero.

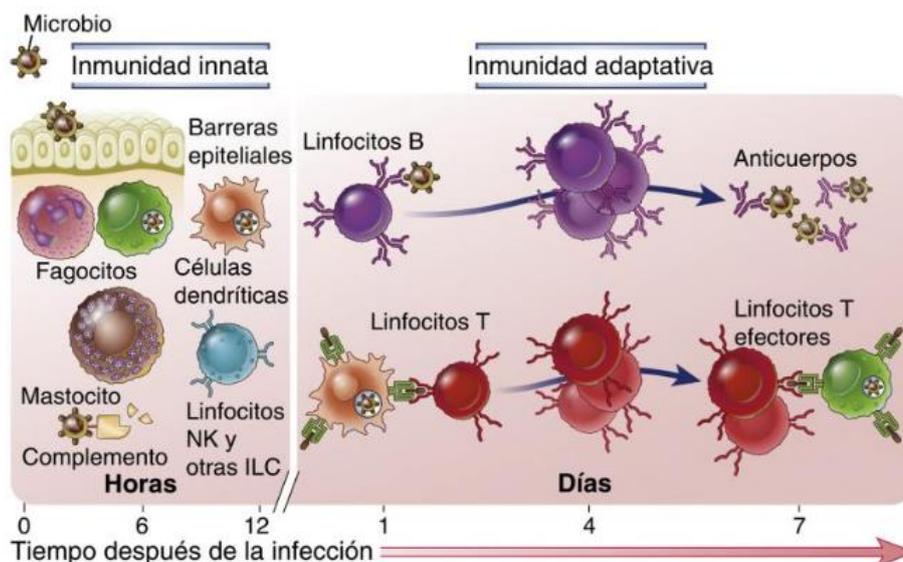
#### 1.3.4. Vacunas basadas en epítomos

La vacunación basada en epítomos es una estrategia de nueva generación que ofrece la posibilidad de un diseño racional basado en regiones proteicas cortas o péptidos, evitando componentes virales no esenciales y fragmentos potencialmente nocivos (Kar et al., 2020). Estas vacunas permiten inducir un gran repertorio de especificidades de células T para respuestas inmunitarias de células T restringidas por HLA; pueden diseñarse como epítomos de longitud mínima (que generalmente sufren de inmunogenicidad deficiente), o péptidos más largos compuestos de múltiples epítomos, ya sea basados en arreglos lineales o estructuras ramificadas/dendrimeras como péptidos antigénicos múltiples (Oyarzún & Kobe, 2016).

Las personas poseen diferentes conjuntos de alelos HLA con especificidades de ligando altamente variables (restricción por HLA) y frecuencias de expresión que difieren sustancialmente entre distintas etnias, y esto representa un problema importante para el diseño de vacunas basadas en epítomos (Oyarzún et al., 2021). En este contexto, las herramientas inmunoinformáticas juegan un rol clave, acelerando el descubrimiento de epítomos y el diseño de vacunas a través del mapeo *in silico* de epítomos a lo largo de todo un proteoma. Esto permite reducir el tiempo y costos involucrados en pruebas experimentales y abordar la variación genética en patógenos (Pan et al., 2020; Oyarzún & Kobe, 2016). Ejemplo de esto último, para abordar el problema de la variación genética del virus de inmunodeficiencia humana (VIH), se desarrolló el "método de mosaico" basado en algoritmos genéticos para generar secuencias de proteínas compuestas artificiales (proteínas de mosaico polivalentes) que están optimizadas para incluir una diversidad máxima de epítomos putativos de células T. Este método optimiza las secuencias de antígenos naturales para aumentar la reactividad cruzada de las respuestas de la vacuna para diversos aislados de VIH-1 (cepa más patogénica del VIH). Este enfoque ha sido capaz de lograr una cobertura de entre el 74 % y el 87 % de las secuencias de la proteína Gag del VIH-1, mientras que una sola proteína Gag natural logró solo una cobertura del 37 % al 67 %. Se ha demostrado que los antígenos mosaicos del VIH-1 se procesan y expresan por células T humanas *in vitro* y varios estudios de inmunogenicidad de prueba de concepto en primates no humanos han demostrado que los antígenos del mosaico del VIH codificados por vectores mejoran la profundidad y amplitud de las respuestas inmunitarias celulares, así como las respuestas de anticuerpos (NDhlovu et al., 2011; Oyarzún & Kobe, 2016; Rezaei et al., 2019).

#### 1.4. Sistema inmune

El sistema inmunológico es una organización de células y moléculas con funciones especializadas en la defensa contra infecciones. Hay dos tipos fundamentalmente diferentes de respuesta a la invasión de patógenos: La respuesta inmune innata y la respuesta inmune adaptativa (Delves & Roitt, 2000). El sistema innato responde casi de inmediato a los patógenos y las células dañadas. Sus receptores son específicos frente a estructuras que son comunes de patógenos relacionados y no distinguen las diferencias sutiles que hay entre ellos. Las exposiciones repetidas inducen respuestas inmunitarias innatas idénticas. Los principales componentes de la inmunidad innata son: 1) barreras físicas y químicas (epitelio y sustancias químicas antimicrobianas producidas en las superficies epiteliales); 2) células fagocíticas (neutrófilos, macrófagos, células dendríticas, mastocitos, linfocitos NK (natural killer) y otras células linfocíticas innatas), y 3) proteínas sanguíneas, como miembros del sistema de complemento y otros mediadores de la inflamación (figura 7) (Abbas et al., 2022).



**Figura 7** Esquema del proceso de activación del sistema inmune innato y adaptativo (extraído de Abbas et al, 2022). La inmunidad innata proporciona la defensa inicial contra las infecciones. La inmunidad adaptativa aparece después y requiere la activación de los linfocitos. La cinética de ambas respuestas son aproximaciones y pueden variar según la infección.

La respuesta inmunitaria adaptativa está mediada por linfocitos, los cuales expresan receptores muy diversos capaces de reconocer un enorme número de antígenos. Hay dos poblaciones principales de linfocitos, conocidos como linfocitos B y linfocitos T, los cuales median diferentes tipos de respuestas inmunitarias adaptativas (Abbas et al., 2022).

#### **1.4.1. Inmunidad humoral**

La inmunidad humoral corresponde a la respuesta inmunitaria mediada por linfocitos B (células B). La respuesta inmune contra antígenos proteicos requiere señales activadoras (cooperadoras) de los linfocitos T CD4+ (células T CD4+). Las células B que reconocen antígenos, proliferan y se diferencian en células plasmáticas que secretan diferentes clases de anticuerpos con funciones distintas. Algunas células plasmáticas secretoras de anticuerpos migran a la médula ósea o los tejidos mucosos donde viven durante años, produciendo cantidades bajas de anticuerpos, los que proporcionan protección inmediata si el patógeno vuelve a infectar al individuo (linfocitos B de memoria) (Abbas et al., 2022).

#### **1.4.2. Inmunidad celular**

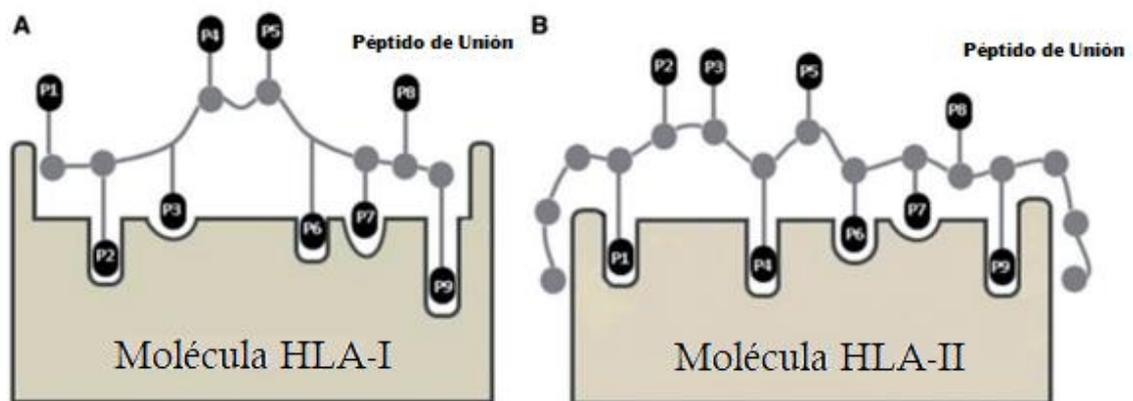
Este sistema es crítico para la defensa contra patógenos intracelulares. Los linfocitos T (células T) reconocen los antígenos de patógenos que son procesados y presentados por células especializadas denominadas células presentadoras de antígenos (APC). Existen diferentes subpoblaciones de células T que colaboran con fagocitos para eliminar los patógenos o a las células infectadas. Las células T poseen receptores que son moléculas de membrana diferentes de los anticuerpos, pero de similar estructura. Las células T tienen una especificidad restringida frente a los antígenos; reconociendo péptidos derivados de patógenos que están unidas a proteínas del HLA (epítomos). Por ello, estos linfocitos reconocen y responden a antígenos asociados a células, pero no a antígenos solubles (Abbas et al., 2022).

Las poblaciones mejor definidas de células T son la de células T cooperadoras CD4+ (auxiliares) y la de células T citotóxicas CD8+, siendo CD4 y CD8 correceptores presentes en las superficies de células T que aumentan el reconocimiento de péptidos presentados por HLA-II y HLA-I, respectivamente (James et al., 2022). Los linfocitos T auxiliares actúan por las citocinas que secretan y las moléculas de membrana que activan a otras células para la eliminación de patógenos. Los linfocitos T CD8+ producen moléculas para la eliminación directa de células infectadas del hospedador. Algunas células T actúan inhibiendo respuestas inmunitarias y se conocen como linfocitos T reguladores (Abbas et al., 2022).

### 1.4.3. Antígeno Leucocitario Humano (HLA)

Las proteínas HLA son complejos proteicos transmembranales que juegan un rol central en la respuesta inmune y en la inmunidad inducida por vacunas (Dendrou et al., 2018), las cuales se encuentran codificadas por un conjunto de 21 loci codificadores de proteínas ubicados en la región 6p21 del genoma humano (Shiina et al., 2009). Estas proteínas se encuentran principalmente involucradas en la presentación de péptidos derivados de patógenos para ser reconocidas por células T, lo cual ocurre en la superficie celular de las APC. Dicha presentación puede desencadenar la activación de las células T y la consiguiente inducción de una respuesta inmune de tipo celular (Reynisson et al., 2020).

Los genes HLA clásicos (A, B, C, DR, DQ y DP) se categorizan en clase I y clase II. La hendidura de unión de HLA clase I donde se acomoda todo el péptido de unión, tiene sus extremos cerrados, mientras que en proteínas HLA clase II esta región presenta sus extremos abiertos para la unión de péptidos más largos, donde una pequeña parte central del péptido se une a la hendidura y los extremos terminales del péptido se orientan hacia fuera de la hendidura (figura 8).

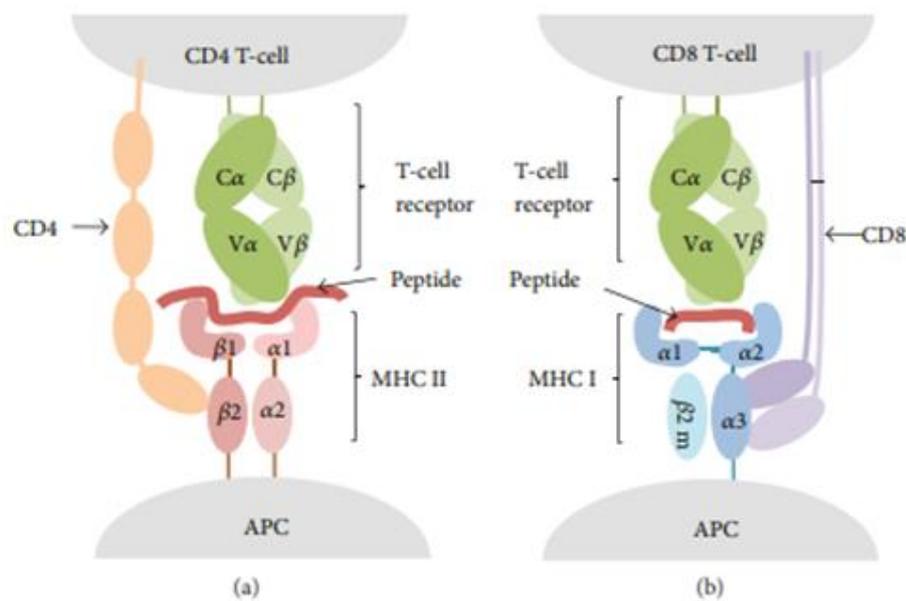


**Figura 8** Esquema de hendidura de unión de HLA-I y HLA-II adaptado de Zhang et al., 2012. Péptidos de unión se unen a (A) HLA-I donde se ajusta un número en la hendidura, y (B) HLA-II donde un péptido más largo se ajusta parcialmente en la hendidura.

Las moléculas clase I (A, B, C) están compuestas de una cadena  $\alpha$  unida no covalentemente a un pequeño polipéptido de  $\beta$ 2-microglobulina. Sus dominios  $\alpha$ 1 y  $\alpha$ 2 forman el sitio de unión a péptidos donde se unen ligandos que gatillan una respuesta inmune de células T CD8+, llamados epítopos de células T CD8+. Estos epítopos consisten en proteínas virales de 8 a 10

aminoácidos que son presentados a linfocitos citotóxicos CD8+ procediendo al mecanismo de respuesta inmune celular. La respuesta mediada por células T CD8+ es crítica en la respuesta a infecciones virales. Además, la generación de células T CD8+ de memoria que son específicas de virus se mantienen en el hospedero y proporcionan protección contra subsecuentes infecciones. Por tanto, la inducción combinada de células T CD8 específicas del virus y anticuerpos puede proporcionar una inmunidad protectora óptima. (Schmidt & Varga 2018).

Las moléculas de HLA clase II (DR, DQ, DP) son moléculas heterodímeras compuestas de una cadena  $\alpha$  y una  $\beta$ , en donde los dominios  $\alpha 1$  y  $\beta 1$  forman el sitio de unión a péptidos, en donde se unen péptidos de 13 a 25 aminoácidos (epítomos de células T CD4+) para ser presentado a linfocitos T CD4+ colaboradores que poseen capacidad para ayudar a células B a producir anticuerpos, inducir macrófagos a mejorar actividad microbicida, reclutar neutrófilos, eosinófilos y basófilos a sitios de infección e inflamación, y producir citocinas y quimiocinas para construir la panoplia completa de la respuesta inmune (figura 9) (Zhu & Paul, 2008) (Parham, 2014).



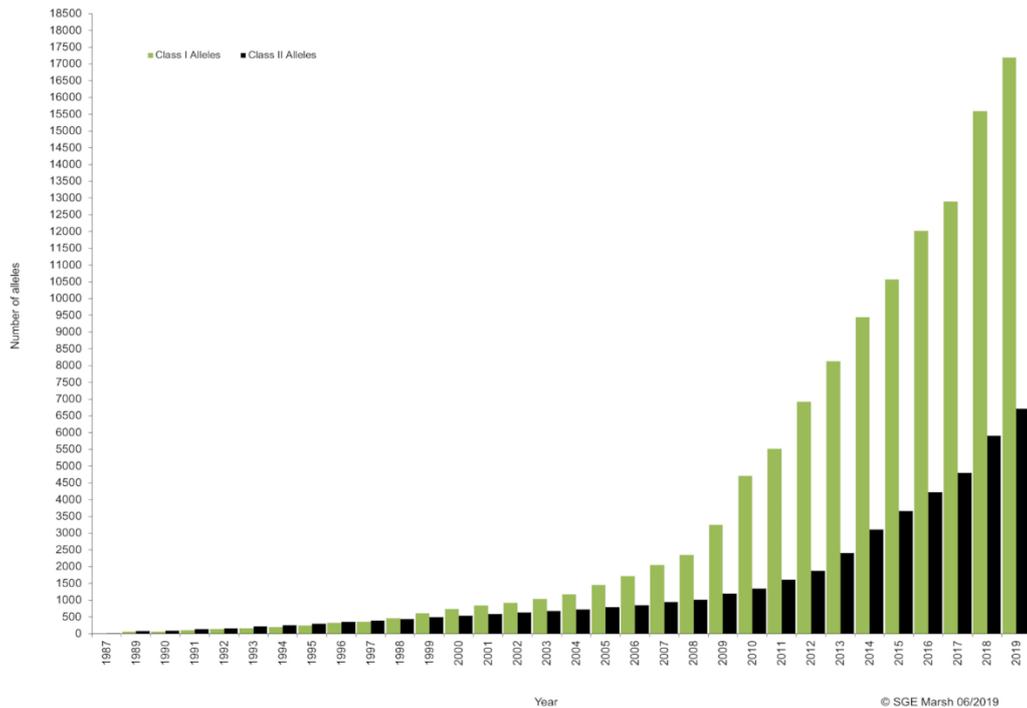
**Figura 9** Esquema de proteínas HLA en la superficie de células APC, presentando péptidos antigénicos a sus respectivos receptores de células T (extraído de: Sanchez-Trincado et al., 2017). (a) Los péptidos en HLA clase II se componen de 13 a 25 aminoácidos y sus extremos sobresalen del sitio de unión a péptido conformado por los dominios  $\alpha 1$  y  $\beta 1$ . Estos péptidos son presentados a receptores de células T CD4+ auxiliares. (b) Los péptidos en HLA clase I se componen de 8 a 10 aminoácidos y no sobresalen del sitio de unión a péptido conformado por los dominios  $\alpha 1$  y  $\alpha 2$ . Estos péptidos son presentados a receptores de células T CD8+.

#### 1.4.4. Polimorfismo de HLA

Los genes HLA son conocidos por presentar el más alto nivel de diversidad del genoma humano y son descritos como “hiperpolimórficos”, con miles de diferentes alelos reportados que se diferencian en combinaciones de polimorfismos de nucleótido simple (SNPs) (Sanchez-Mazas, 2020; Robinson et al, 2020). La característica distintiva de esta diversidad radica en el elevado polimorfismo que afecta a las regiones que codifican los sitios de unión a péptidos (dominios  $\alpha 1$  y  $\alpha 2$ , para HLA clase I, y dominios  $\alpha 1$  y  $\beta 1$  para HLA clase II) (Robinson et al, 2020). Esta diversidad se explica por la ventaja que confiere a individuos heterocigotos, mejorando el repertorio de células T y así la protección contra patógenos gracias a una mayor variedad de péptidos que pueden expresarse en comparación con individuos homocigotos (Shiina et al., 2009).

Múltiples alelos de HLA tienen el potencial de conferir protección o susceptibilidad a un determinado patógeno, sin embargo, la hipervariabilidad genómica de HLA dificulta realizar predicciones sobre los efectos en relación a una determinada respuesta inmune (Rezza & Weaver, 2019). Esta variabilidad ha demostrado ser un factor significativo para el estudio de eficacia e inmunogenicidad de vacunas (Nielsen et al., 2018; Hertz et al., 2013). Por este motivo es que existe una necesidad de comprender cómo los individuos y grupos poblacionales reaccionan contra diferentes tipos de infecciones tomando en cuenta su genotipo HLA, ya que distintos alelos pueden responder y presentar diferentes regiones y epítomos (epítomos promiscuos) (Hertz et al., 2013; Alicia, 2020).

La base de datos de inmunopolimorfismo alojada en el Instituto Europeo de Bioinformática (<https://www.ebi.ac.uk/ipd/index.html>) fue desarrollada en 2003 con el propósito de sistematizar la información internacional sobre los polimorfismos en genes del sistema inmune y en particular del sistema HLA. Visitada en 20 de Enero de 2023, la plataforma registra 25.228 alelos para HLA clase I y 10.592 alelos para HLA clase II (<https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/>), registrando descubrimientos de nuevos alelos con crecimiento exponencial (figura 10) (Robinson et al, 2020).



**Figura 10** Evolución anual del número de alelos de HLA descubiertos y anotados en la base de datos IPD-IMGT/HLA (<http://www.ebi.ac.uk/ipd/imgt/hla/>). Las barras verdes corresponden a alelos HLA clase I y barras negras corresponden a alelos HLA clase II. (Robinson et al., 2020).

#### 1.4.4.1. Componente étnico: factor de susceptibilidad o protección

Se ha señalado que la susceptibilidad y/o resistencia a un patógeno, no sólo es observable en individuos, sino también en grupos étnicos (Pan et al, 2020). Un grupo étnico consistente en un conjunto de personas que comparten composición genética, identidad social/cultural y patrones de comportamiento (Bonham et al., 2018). Es posible constatar en relación a la relación entre etnicidad frente al riesgo de contagio de COVID-19 que la mayoría de los reportes recopilados de Reino Unido y Estados Unidos indican un impacto significativo de COVID-19 en comunidades negras, asiáticas y minorías étnicas (BAME, por sus siglas en inglés), en contraste con comunidades blancas, donde las comunidades BAME presentan un índice de contagio y mortalidad significativamente mayor (Pan et al., 2020).

#### 1.4.4.2. Inmunodominancia y *Hotspots* de epítomos

En una respuesta inmune antiviral las células T reconocen solo una pequeña fracción del total de epítomos potenciales y a veces solo a uno. Más específicamente, menos del 3% de estos

péptidos posibles pueden unirse a moléculas HLA lo suficientemente fuerte como para generar una respuesta inmune. Esto implica que una pequeña fracción de péptidos antigénicos tiene la capacidad de montar respuestas inmunitarias. Este(os) epítipo(s) se les denomina inmunodominante (Nowak et al., 1995; Assarson et al., 2008; Zhang et al., 2012; Afrough et al., 2019).

Se define un *hotspot* de epítomos como una región potencialmente inmunodominante que contiene una alta densidad de epítomos que son presentados por diferentes alelos HLA, y que están potencialmente presentes en múltiples individuos de una población (Hertz et al., 2013). Las regiones ricas en epítomos específicas para una población conforman “clústeres de epítomos”, los cuales pueden determinar patrones de inmunodominancia. Estas regiones permiten seleccionar epítomos candidatos para el diseño de una vacuna (Scholzen et al., 2019) (Oyarzún et al., 2021).

### **1.5. Inmunoinformática**

La inmunoinformática corresponde a un área de especialización de la bioinformática, cuyo objetivo es modelar y desarrollar métodos predictivos de interacciones entre antígenos, epítomos y receptores del sistema inmune (Evans, 2008). Una vez es identificado un antígeno, mediante herramientas inmunoinformáticas, es posible predecir qué región de este antígeno puede generar una respuesta inmunogénica por sí solo (María et al., 2017). La aplicación de métodos computacionales se ha aplicado al estudio de estructuras de anticuerpos, células B, células T, alérgenos, predicciones de unión a HLA, modelamiento de epítomos, análisis de redes inmunes y el desarrollo de vacunas (Potocnakova et al., 2016). Estas herramientas permiten acelerar el descubrimiento de epítomos y el diseño de vacunas a través de mapeo *in silico* de miles de péptidos, además de ofrecer un marco para tratar racionalmente la enorme diversidad de proteínas HLA, reduciendo tiempo y costos que involucrarían ensayos experimentales (Oyarzún et al., 2015).

Se han reportado métodos inmunoinformáticos se han desarrollado para guiar la selección de epítomos de células T considerando la fracción de individuos potencialmente cubiertos por vacunas basadas en epítomos. Como ejemplo, existen herramientas para determinar conjuntos de epítomos óptimos con respecto a la inmunogenicidad en un individuo específico o una población objetivo (e.g., un determinado grupo étnico). Al respecto, la herramienta OptiTope se basa en un marco matemático para la selección de un conjunto óptimo de péptidos para vacunas

basadas en epítomos. Para esto, de un conjunto dado de epítomos candidatos, OptiTope escoge un subconjunto que, de todos los subconjuntos que cumplan ciertos requisitos para una vacuna efectiva (por ejemplo, tolerancia a la mutación, cobertura de población), predigan la máxima inmunogenicidad en la población objetivo. Luego, dado un conjunto de epítomos y un conjunto de alelos HLA, la inmunogenicidad del conjunto de epítomos está compuesta de la inmunogenicidad de sus componentes con respecto a los diferentes alelos del HLA. Esto produce una interpretación matemática de la inmunogenicidad general como la suma ponderada de la inmunogenicidad de los epítomos. (Toussaint & Kohlbacher, 2009).

Por otra parte, Pepvac es un método que puede optimizar la formulación de vacunas multi-epítomo con amplia cobertura poblacional, considerando la unión de péptidos a un conjunto de moléculas HLA clase I como base para predecir epítomos de células T. Esta optimización se logra a través de la predicción de péptidos que se unen a varias moléculas HLA con especificidad de unión a péptidos similar (supertipos). La frecuencia fenotípica combinada de los supertipos utilizados por PEPVAC (A2, A3, A24, B7 y B15) es >95 % para las cinco principales etnias estadounidenses (negra, caucásica, hispana, nativa americana y asiática). Con esto, al apuntar a estos supertipos con predicciones de epítomos, postula proporcionar potencialmente una cobertura poblacional  $\geq 95$  %, independiente de la etnia. (Reche & Reinherz, 2005).

### **1.5.1. Métodos pan-específicos**

Los métodos pan-específicos utilizan datos experimentales de múltiples alelos HLA como datos de entrada para predecir especificidades de unión no solo de los alelos previamente caracterizados sino también de nuevos alelos para los cuales no existe información experimental disponible (Reynisson et al., 2020). Al respecto, las hendiduras de unión de proteínas HLA son distintas dependiendo del alelo, lo que dificulta definir el núcleo de unión óptimo de un péptido dado y con ello la predicción de unión de péptidos a moléculas HLA (Zhang et al., 2012). El principio detrás de los métodos pan-específicos es establecer conexiones entre diferentes alelos HLA en términos de especificidades de unión, en virtud de lo cual se pueden capturar las especificidades vinculantes del alelo conocido y otro desconocido. En principio los métodos pan-específicos tiene la capacidad de predecir los sitios de unión de cualquier alelo toda vez que la secuencia proteica sea conocida (Zhang et al., 2009; Liu et al., 2017; Nielsen et al., 2007).

### 1.5.2. Predivac-3.0

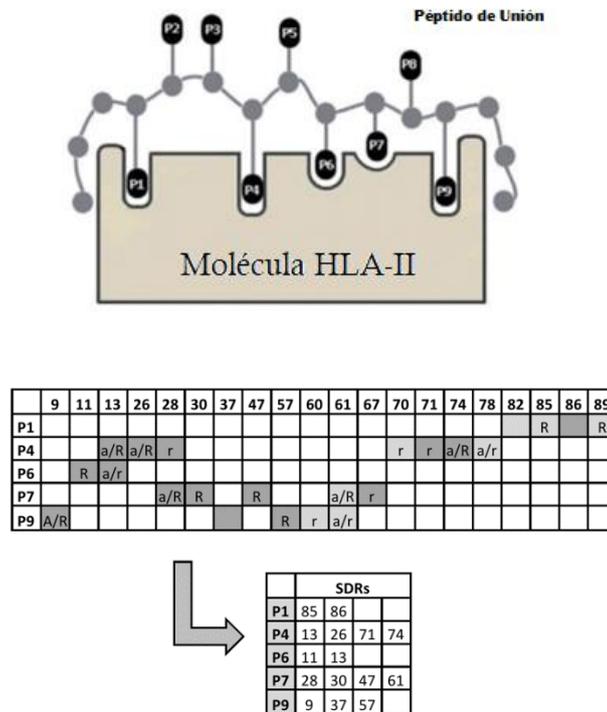
Predivac-3.0 es una herramienta computacional con enfoque en poblaciones objetivo para el descubrimiento de epítomos de células T y clústeres de epítomos de células T, tomando en consideración el componente étnico de proteínas HLA. Este método es de interés para el diseño de vacunas de nueva generación contra patógenos emergentes, ya que las distribuciones geográficas de los virus zoonóticos y poblaciones que necesitan vacunas suelen estar bien definidas para epidemias regionales. Por lo tanto, en teoría es posible determinar la distribución de las proteínas HLA de las poblaciones que son potencialmente susceptibles a patógenos emergentes. Predivac-3.0 también tiene potencial para contribuir a la comprensión de respuestas inmunes inducidas por vacunas en contextos poblacionales, aportando información que contribuya al diseño racional de inmunógenos basados en epítomos considerando variaciones a nivel étnico (Oyarzún et al., 2021).

Predivac-3.0 consta de un módulo principal que consulta a una base de datos (PredivacDB) especialmente diseñada de residuos determinantes de especificidad (SDRs, por sus siglas en inglés) de HLA clase I y II, asociados con secuencias de proteínas HLA con sitios de unión de péptidos de alta afinidad, y una base de datos de frecuencias de alelos HLA disponible en “Allele Frequency Net Database” (AFND). PredivacDB contiene 26.068 péptidos que representan 77 alelos HLA clase I (23.373 péptidos) y 29 alelos HLA clase II (2.695 péptidos) exportados y filtrados de IEDB. Predivac-3.0 predice la unión de péptidos al establecer una correlación entre los SDRs de la proteína HLA de consulta y los SDR asociados con las proteínas HLA de especificidad conocida. El proceso implica los siguientes pasos: (i) Se identifican los SDRs para cada una de las posiciones de unión en la secuencia de la proteína HLA de consulta; (ii) Se consulta PredivacDB y se calculan las frecuencias y los pesos de los aminoácidos para secuencias de péptidos asociadas con alotipos que comparten SDR similares con la proteína de consulta en cada posición de unión; y (iii) Una matriz de peso de posición (PWM) se construye en base a los datos vinculantes, que consta de 20 columnas (aminoácidos) × 5 filas (sitios de unión) (figura 11), donde los valores de la matriz se calculan mediante la siguiente fórmula:

$$w_{ij} = \log(p_{a,i}/p_a)$$

Donde “ $p_{a,i}$ ” es la probabilidad de observar el aminoácido “a” en la posición “i” del péptido ( $i=0, \dots, 8$ ), y “ $p_a$ ” es la probabilidad de observar el aminoácido “a” en el modelo de fondo. La

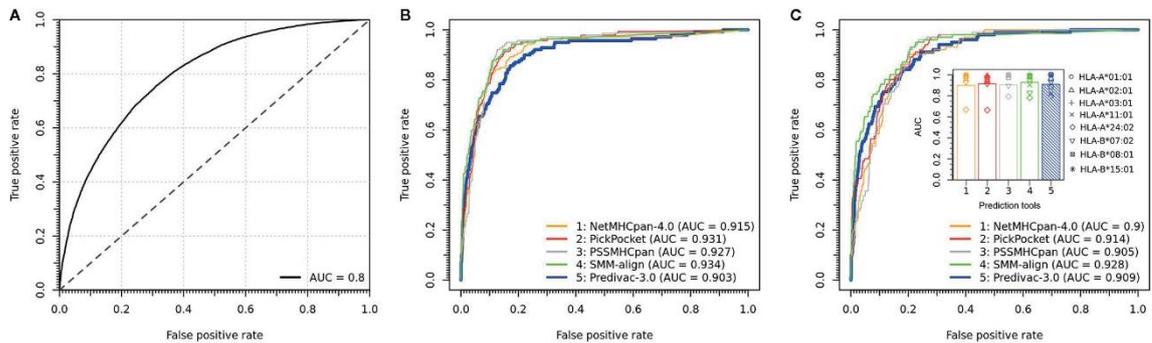
frecuencia de un residuo en una posición  $i$  en el péptido se estima usando pseudorecuentos (Oyarzún et al., 2013).



**Figura 11** Posiciones SDR y criterios empleados para seleccionar estas posiciones en términos de polimorfismo y efecto electrostático calculado en la hendidura de unión, adaptado de Oyarzún et al., 2013. Las columnas corresponden a las posiciones de aminoácidos de contacto más conservadas identificadas en las estructuras cristalinas HLA clase II analizadas. Las filas son las posiciones de unión a péptidos (P1, P4, P6, P7 y P9). Las celdas en gris son posiciones altamente polimórficas. En términos de modelado electrostático, las celdas con letras A/a corresponden a posiciones responsables del anclaje, mientras que R/r corresponde a posiciones responsables del reconocimiento. Mayúsculas corresponde a posiciones con un efecto más significativo. Las celdas de colorp gris oscuro son posiciones polimórficas y las celdas de color gris claro son posiciones no polimórficas.

El área bajo la curva característica operativa del receptor (AUC), es un gráfico de la sensibilidad frente a la tasa de falsos positivos (1 - especificidad). El AUC proporciona una indicación de la precisión de un método de predicción, donde un AUC = 1 corresponde a predicciones perfectas y un AUC = 0,5 refleja predicciones aleatorias. El rendimiento predictivo de Predivac-3.0 se midió en términos del AUC. Mediante validación cruzada con exclusión de un alelo, utilizando un conjunto de datos de 17.425 aglutinantes peptídicos de alta afinidad, restringidos por 46 alelos HLA clase I con 25 o más ligandos peptídicos presentes en la base de datos de Predivac, el programa demostró una elevada capacidad en identificación de epítomos

de células T CD8+ (AUC ~0,8). Además, Predivac-3.0 se comparó con métodos pan-específicos de última generación, donde demostró una precisión comparable con la de estos métodos (figura 12). Por otra parte, Predivac-3.0, demostró precisión y sensibilidad para la identificación *in silico* de epítomos de células T CD8+ específicos del VIH-1 que son inmunodominantes en la población japonesa. El enfoque de esta herramienta permite guiar el descubrimiento de epítomos de acuerdo con las frecuencias alélicas de HLA en una población étnica específica (Oyarzún et al., 2021)



**Figura 12** Validación de predicciones de epítomos de células T CD8+ extraído de Oyarzún et al., 2021. (A) Rendimiento predictivo general de Predivac-3.0 medido mediante validación cruzada de exclusión. Se compara Predivac-3.0 frente a métodos pan-específicos (NetMHCpan 4.0, Pickpocket, PSSMHCpan y SMM-align) utilizando (B) un conjunto de datos de epítomos de células T CD8+ validados experimentalmente derivados del recurso de análisis IEDB (conjunto de datos IEDB) y (C) un conjunto de datos de ligandos peptídicos de alta afinidad derivados del repositorio de Dana-Farber para el aprendizaje automático en inmunología (conjunto de datos DFRMLI).

## **HIPÓTESIS**

Regiones ricas en epítomos en el proteoma viral del SARS-CoV-2 son específicas para grupos HLA en la población chilena con un componente étnico definido.-

### **Objetivo general**

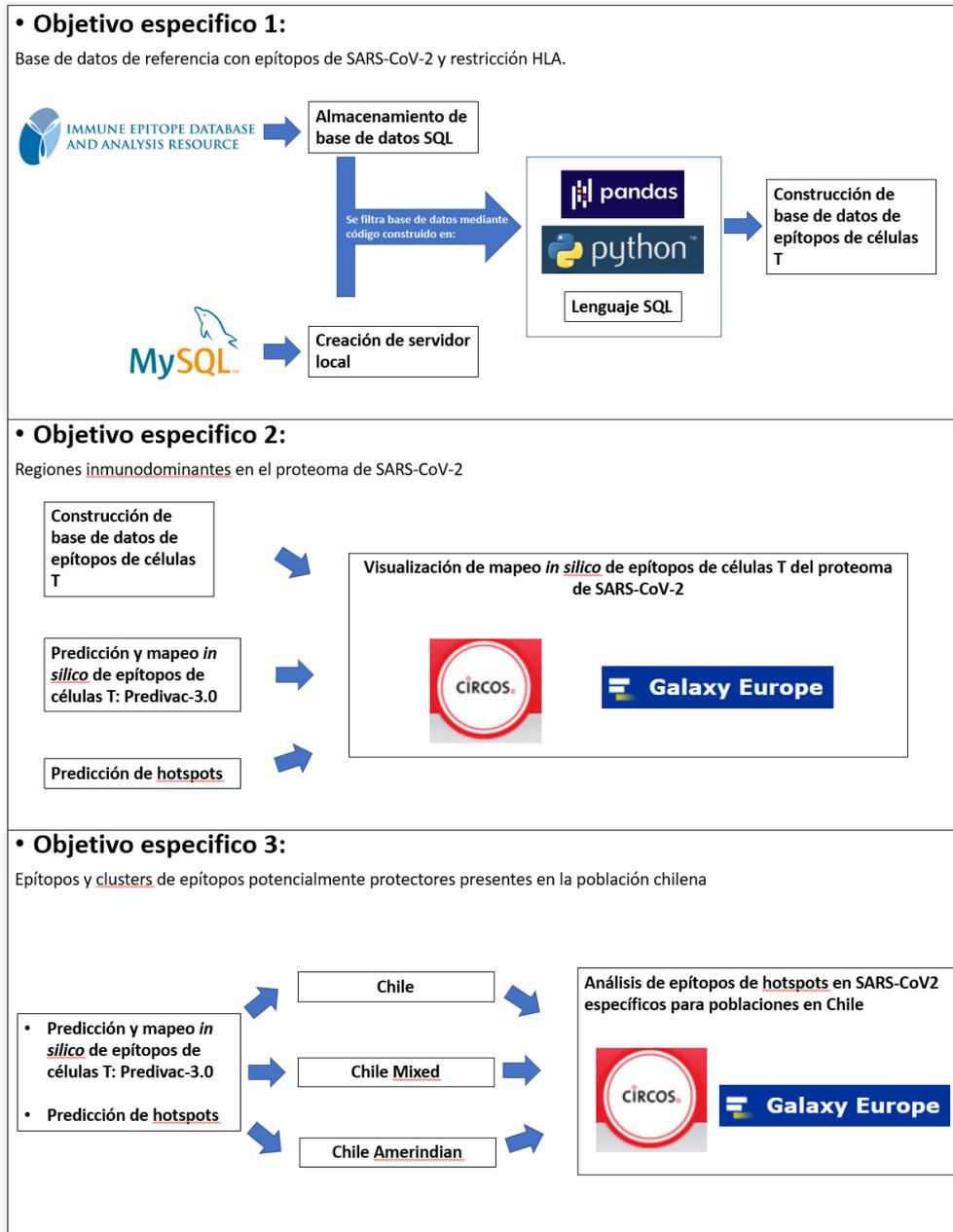
Determinar en el proteoma de SARS-CoV-2 regiones ricas en epítomos de células T CD8+ y CD4+ específicos para la población chilena.-

### **Objetivos específicos**

1. Construir base de datos de referencia con epítomos de SARS-CoV-2 determinados experimentalmente y restricción HLA conocida.
2. Identificar clústeres de epítomos de células T potencialmente inmunodominantes en el proteoma de SARS-CoV-2.
3. Seleccionar epítomos y regiones en SARS-CoV-2 que sean de interés para el desarrollo de vacunas dirigidas a la población chilena.

## 2. Estrategia experimental

A continuación, se presenta un esquema con la estrategia experimental y metodologías utilizadas en el marco de la investigación:



## 2.1. Construcción de base de datos de epítomos de células T

Las secuencias de epítomos se obtuvieron desde el portal IEDB (<https://www.iedb.org/>), el cual posee una completa base de datos de epítomos y recursos de análisis (Vita et al., 2018). De este repositorio se obtuvo un archivo que almacena sus datos de forma relacional en formato SQL (Structured Query Language) encontrado bajo el nombre de “SQL Statement Export” ([https://www.iedb.org/database\\_export\\_v3.php](https://www.iedb.org/database_export_v3.php)). Luego se instaló la plataforma MySQL (<https://dev.mysql.com/downloads/mysql/>). MySQL es una plataforma de código abierto que permite el acceso y trabajo con base de datos que utilizan lenguaje SQL, un lenguaje diseñado para administrar y recuperar información de bases de datos relacionales. Mediante la herramienta “MySQL Workbench” se generó un servidor local con los parámetros que vienen por defecto, en el que se incorporó la base de datos extraída de IEDB. Se utilizó este servidor para facilitar el acceso y filtrado de la información necesaria de IEDB mediante la generación de códigos en lenguaje de programación Python (Van Rossum & Drake, 1995), la herramienta de análisis y manipulación de datos de código abierto “pandas” (McKinney, 2010) y lenguaje SQL. La función de este código fue filtrar, ordenar y almacenar epítomos de células T corroborados experimentalmente del proteoma de SARS-CoV-2 presentes en IEDB, en una base de datos en formato “.xlsx”, categorizando la información en columnas de acuerdo con la tabla 2 por: proteína, organismo, posición inicial de secuencia, posición final de secuencia, hospedero, restricción HLA, artículo, experimento, respuesta y categoría. Las columnas organismo, hospedero y categoría se utilizaron para corroborar que la información correspondía a epítomos de células T de SARS-CoV-2. Esta base de datos fue denominada como “EpitEx”.

ID	SECUENCIA	PROTEÍNA	ORGANISMO	POSICIÓN INICIAL	POSICIÓN FINAL	HOSPEDERO	RESTRICCIÓN HLA	ARTÍCULO	EXPERIMENTO	RESPUESTA	CATEGORÍA
7168194	VTLAILTALR	envelope protein [Severe acute respira	SARS-CoV2	29	38	human (Homo sapiens)	HLA-A*11:01	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7168184	NIVNVSLVK	envelope protein [Severe acute respira	SARS-CoV2	45	53	human (Homo sapiens)	HLA-A*03:01	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
8259628	SLVKPSFYV	envelope protein [Severe acute respira	SARS-CoV2	50	58	human (Homo sapiens)	HLA-A*02:01	Robust T Cell Immunity in Convalescent Individuals with Asymptom multimer/tetramer		qualitative binding	T Cell
7444074	LVKPSFYVY	envelope protein [Severe acute respira	SARS-CoV2	51	59	human (Homo sapiens)	HLA-B*15:01, HLA-C*07:02	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod ELISPOT, multimer/tetramer		IFNg release, qualitative binding	T Cell
7168214	KPSFYVYSRV	envelope protein [Severe acute respira	SARS-CoV2	53	62	human (Homo sapiens)	HLA-B*07:02	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7168204	FVYSRVKKNL	envelope protein [Severe acute respira	SARS-CoV2	56	65	human (Homo sapiens)	HLA-A*24:02	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7444400	VVYSRVKKNL	envelope protein [Severe acute respira	SARS-CoV2	57	65	human (Homo sapiens)	HLA-C*06:02	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod multimer/tetramer		qualitative binding	T Cell
7358374	GTITVEELK	membrane glycoprotein [Severe acute	SARS-CoV2	6	14	human (Homo sapiens)	HLA-A*11:01, HLA-A*68:01	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 biological activity, multimer/tetr activation, qualitative binding		IFNg release, TNFa release, activati	T Cell
7168195	GTITVEELKK	membrane glycoprotein [Severe acute	SARS-CoV2	6	15	human (Homo sapiens)	HLA-A*11:01	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7168175	KLLEQWNLV	membrane glycoprotein [Severe acute	SARS-CoV2	15	23	human (Homo sapiens)	HLA-A*02:01	Robust T Cell Immunity in Convalescent Individuals with Asymptom ELISPOT, ICS, biological activity, IFNg release, TNFa release, activati		IFNg release, TNFa release	T Cell
7161974	FLFLTWICL	membrane glycoprotein [Severe acute	SARS-CoV2	26	34	human (Homo sapiens)	HLA-A*02:01	Suboptimal SARS-CoV-2-specific CD8<sup>+</sup> T cell response i ICS		IFNg release, TNFa release	T Cell
7358297	FAYANRNRF	membrane glycoprotein [Severe acute	SARS-CoV2	37	45	human (Homo sapiens)	HLA-B*15:01, HLA-B*35:01	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod biological activity, multimer/tetr activation, qualitative binding		qualitative binding	T Cell
7444339	AYANRNRF	membrane glycoprotein [Severe acute	SARS-CoV2	38	45	human (Homo sapiens)	HLA-A*24:02	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod multimer/tetramer		qualitative binding	T Cell
7444009	AYANRNRF	membrane glycoprotein [Severe acute	SARS-CoV2	38	46	human (Homo sapiens)	HLA-A*24:02	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod multimer/tetramer		qualitative binding	T Cell
7442827	NVFAFPFTI	ORF10 protein [Severe acute respirator	SARS-CoV2	5	13	human (Homo sapiens)	HLA-A*02:01	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell
7168240	NVFAFPFTIY	ORF10 protein [Severe acute respirator	SARS-CoV2	5	14	human (Homo sapiens)	HLA-B*15:01	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7443026	AFPFYIYSL	ORF10 protein [Severe acute respirator	SARS-CoV2	8	16	human (Homo sapiens)	HLA-A*24:02	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell
7168219	FPFYIYSLLL	ORF10 protein [Severe acute respirator	SARS-CoV2	9	18	human (Homo sapiens)	HLA-B*07:02	SARS-CoV-2-derived peptides define heterologous and COVID-19-in ELISPOT		IFNg release	T Cell
7442888	SLVPGFNEK	orf1ab polyprotein [Severe acute resp	SARS-CoV2	3	11	human (Homo sapiens)	HLA-A*03:01, HLA-A*11:01	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell
7444170	VLSEARQHL	orf1ab polyprotein [Severe acute resp	SARS-CoV2	38	46	human (Homo sapiens)	HLA-A*02:01	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod multimer/tetramer		qualitative binding	T Cell
7444377	HLKDGTCGL	orf1ab polyprotein [Severe acute resp	SARS-CoV2	45	53	human (Homo sapiens)	HLA-B*08:01	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod multimer/tetramer		qualitative binding	T Cell
7442788	GLVEVEKGV	orf1ab polyprotein [Severe acute resp	SARS-CoV2	52	60	human (Homo sapiens)	HLA-A*02:01	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell
7443080	RTAPHGVMV	orf1ab polyprotein [Severe acute resp	SARS-CoV2	77	85	human (Homo sapiens)	HLA-B*07:02	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell
7168211	APHGHVMVEL	orf1ab polyprotein [Severe acute resp	SARS-CoV2	79	88	human (Homo sapiens)	HLA-B*07:02	SARS-CoV-2 genome-wide T cell epitope mapping reveals immunod ELISPOT, multimer/tetramer		IFNg release, qualitative binding	T Cell
7442789	VMVELVAEL	orf1ab polyprotein [Severe acute resp	SARS-CoV2	84	92	human (Homo sapiens)	HLA-A*02:01	SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-1 multimer/tetramer		qualitative binding	T Cell

**Tabla 2** Ejemplo de la estructura de datos presente EpiEx para epítomos de células T de SARS-CoV-2, cuya base de datos fue implementada mediante código de Python, pandas y SQL. Columnas corresponden a características identificadas de epítomos de células T. Filas corresponden a casos ejemplo de información presente en EpiEx.

## **2.2. Predicción y mapeo *in silico* de epítomos de células T: Predivac-3.0**

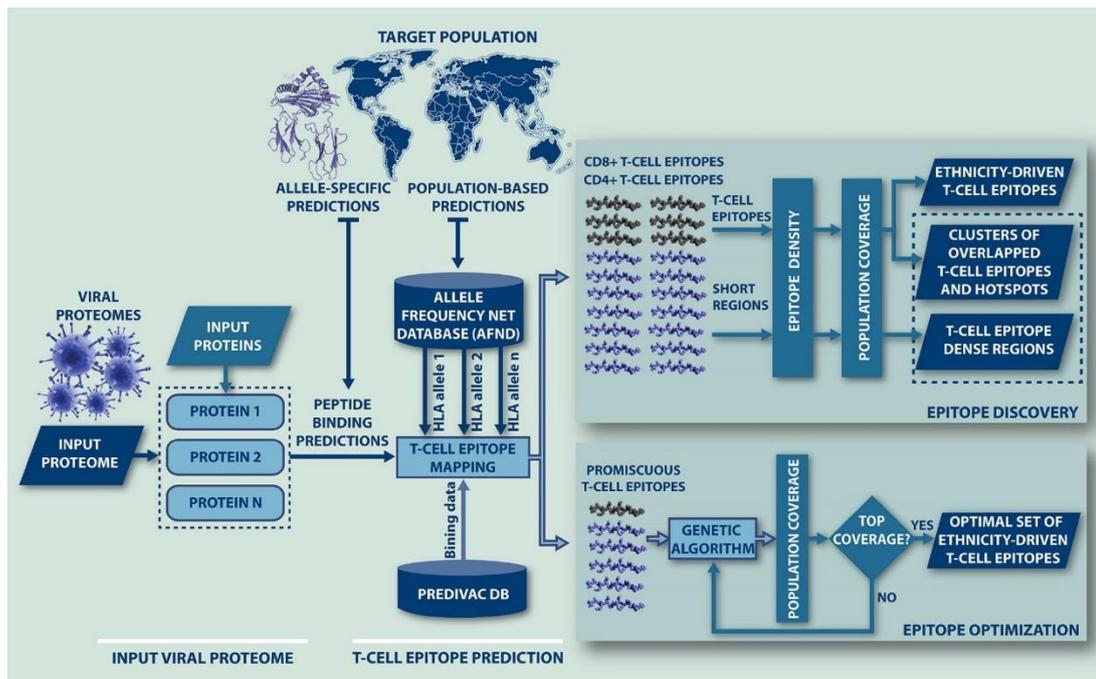
### **2.2.1. Mapeo *in silico* de epítomos de células T**

Se utilizó la herramienta inmunoinformática Predivac-3.0, la cual requiere que el usuario proporcione el proteoma de consulta que consta de un archivo FASTA con múltiples secuencias. Luego, selecciona la población objetivo, obteniendo los datos de alelo HLA para esa población y extrayendo automáticamente la información de SDR de las proteínas HLA de consulta para realizar un mapeo *in silico* de epítomos de células T. Los polimorfismos de HLA se manejan favoreciendo la selección de epítomos de células T CD4+ y CD8+ ampliamente reconocidos por alelos HLA (epítomos promiscuos) e inmunodominantes que tienen la capacidad de desencadenar potencialmente una amplia respuesta inmune en una alta proporción de individuos que expresan moléculas distintas de HLA (Oyarzún et al., 2021).

El mapeo de epítomos de células T se llevó a cabo mediante el análisis de secuencias de proteínas de consulta en segmentos superpuestos noaméricos (péptidos), a cada uno de los cuales se le asigna una puntuación de unión utilizando una matriz de peso de posición (técnica de ventana deslizante). Las puntuaciones brutas son normalizadas por Predivac-3.0 en un rango de 1 a 100 mediante una transformación lineal, teniendo en cuenta las puntuaciones de péptidos teóricas mínimas y máximas que se pueden obtener de la matriz de peso. Un límite recomendado para discriminar los péptidos que se unen de los que no lo hacen es 60 (Oyarzún et al., 2013).

### **2.2.2. Aplicación poblacional: Detección de *hotspots***

Predivac-3.0 requiere que el usuario ingrese la secuencia del proteoma de consulta (o proteínas individuales) y establecer la población objetivo (país o región geográfica), sobre la cual el programa extrae del portal AFND (Allele Frequency Net Database) los datos de frecuencia de alelos HLA clase I y clase II disponibles para muestras poblacionales por país/región. Luego, el programa ofrece la opción de implementar dos tipos de predicciones: (i) Epitope Discovery y (ii) Epitope Optimization. En el primer caso, el programa genera un ranking de epítomos de células T en función de su cobertura poblacional, o alternativamente permite detectar regiones ricas en presencia de epítomos (clústeres). En el segundo caso, el programa implementa un algoritmo genético que explora y optimiza las combinaciones de epítomos de células T para maximizar la cobertura de la población (Oyarzún et al., 2021) (figura 13).



**Figura 13** Esquema de diagrama de flujo que representa los pasos seguidos por Predivac-3.0 para realizar un mapeo *in silico* de epítomos de células T de proteomas virales, extraído de Oyarzún et al., 2021. El proteoma es analizado de acuerdo con las proteínas que lo componen, escaneando la secuencia en búsqueda de epítomos de células T considerando contexto HLA de la población objetivo que se está estudiando, generando un ranking de epítomos de células T en función de su cobertura poblacional (Epitope Discovery) y optimizando las combinaciones de epítomos de células T para maximizar la cobertura de la población (Epitope Optimization).

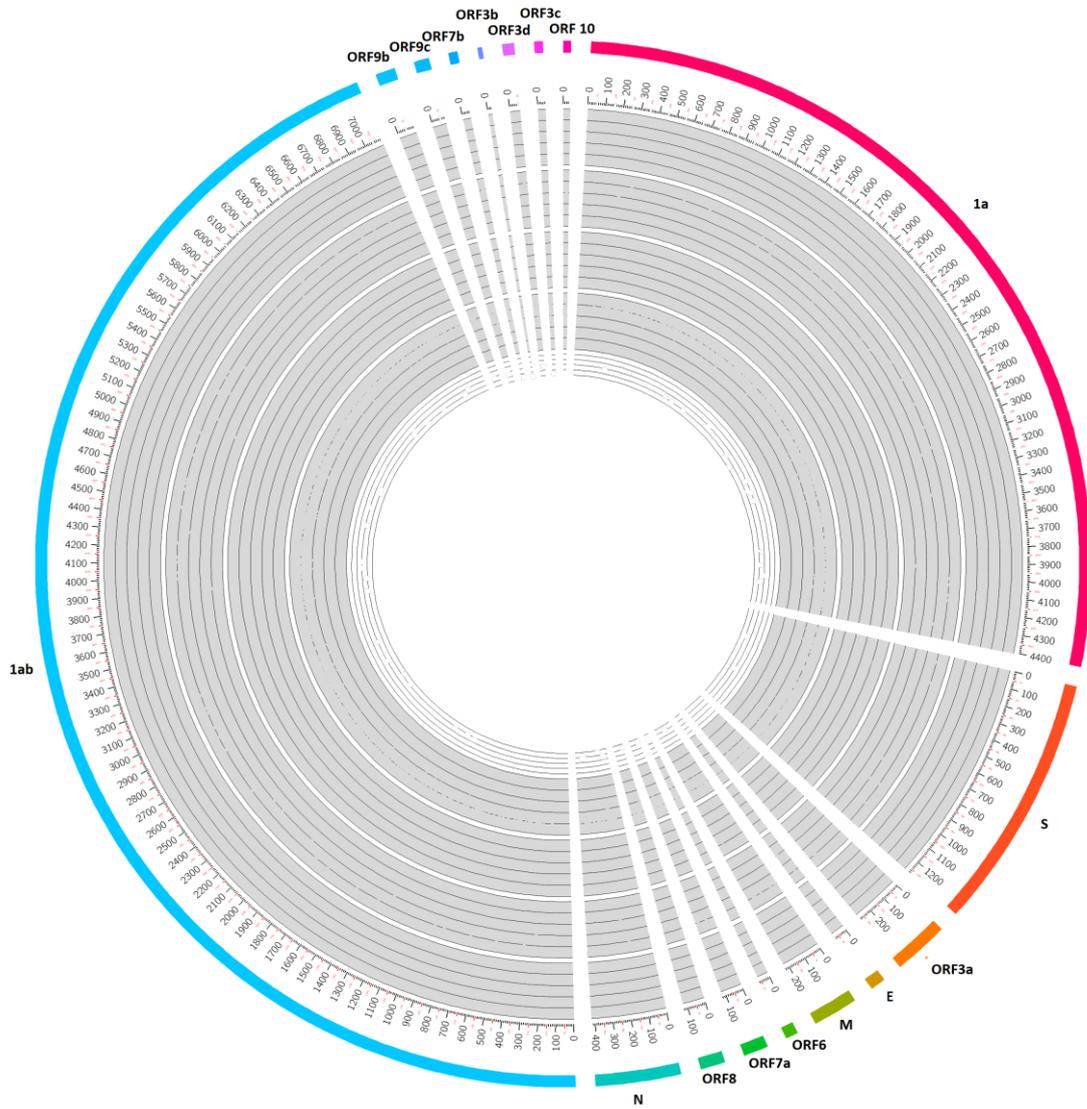
Se aplicó Predivac-3.0 para escanear el proteoma de SARS-CoV-2 con el fin de detectar regiones de alta densidad de epítomos de células T y asociados con una alta cobertura de población en Chile. Este proceso fue realizado a través del modo “Epitope Discovery”, el cual entrega un listado de epítomos y datos de restricción de HLA. Se implementó un algoritmo a este listado para detectar clústeres de epítomos superpuestos (en regiones 9-mer) y regiones enriquecidas de epítomos en marcos de tamaño definido por el usuario (regiones 30-mer, por defecto). El programa selecciona los clústeres de epítomos que cumplan con los siguientes criterios: (i) densidad de epítomos  $\geq 90\%$  de la cantidad máxima de epítomos superpuestos (en base al proteoma), y (ii) cobertura poblacional  $\geq 20\%$  de la población objetivo. La significancia estadística de estos clústeres es determinada mediante un procedimiento de simulación consistente en seleccionar aleatoriamente (1000 veces) un 10% de regiones del mismo tamaño, utilizando el promedio de densidad de epítomos de cada simulación como la distribución de epítomos para calcular valores p. Las regiones que tengan un valor  $p < 0,001$  son consideradas como clústeres (Oyarzún et al., 2021). Luego, los clústeres superpuestos se fusionan y las regiones más reactivas (inmunodominantes) con cobertura de población  $\geq 90\%$ , fueron

denotadas como *hotspots*. Oyarzún et al., 2021 utiliza el criterio de cobertura de población  $\geq 80\%$  para denotar *hotspots*.

### 2.3. Visualización epítomos putativos de células T

Para la visualización de posición y densidad de epítomos de células T de EpiTex y Predivac-3.0 a lo largo del proteoma de SARS-CoV-2, se utilizó la herramienta de visualización Circos (<http://circos.ca/>) (Krzywinski et al., 2009) disponible en la plataforma web Galaxy (<https://usegalaxy.eu/>) (Afgan et al., 2018). Galaxy es una plataforma basada en la web de código abierto que permite el análisis de grandes conjuntos de datos biomédicos, como los presentes en genómica, proteómica y metabolómica. Circos es una de las herramientas de las que dispone Galaxy. Circos utiliza un diseño de ideograma circular para facilitar la visualización de las relaciones entre pares de posiciones mediante el uso de cintas, que codifican la posición, el tamaño y la orientación de los elementos genómicos relacionados.

El esquema circular construido presentó gráficamente la información mediante un modelo de 6 anillos concéntricos. El anillo más externo refiere a la secuencia aminoacídica de las proteínas que componen el proteoma de SARS-CoV-2. Los siguientes 4 anillos concéntricos están segmentados en quintiles y el anillo concéntrico más interno está segmentado en cuartiles. En los 5 anillos segmentados se graficó la información en el siguiente orden de afuera hacia adentro: **Anillo 1)** Posición de epítomos de células T registrados en EpiTex. El anillo está segmentado en quintiles donde cada quintil corresponde a 10 restricciones HLA asociadas a epítomo para HLA clase I y 5 restricciones HLA asociadas a epítomo para HLA clase II. **Anillo 2)** Posición de epítomos de células T predichos por Predivac-3.0. El anillo está segmentado en quintiles donde cada quintil corresponde a 10 restricciones HLA asociadas a epítomo para HLA clase I y 5 restricciones HLA asociadas a epítomos para HLA clase II. **Anillo 3)** Densidad de epítomos de células T superpuestos a lo largo del proteoma de SARS-CoV-2 registrados en EpiTex. El anillo está segmentado en quintiles donde cada quintil corresponde a 2 epítomos para ambos HLA. **Anillo 4)** densidad de epítomos de células T superpuestos a lo largo del proteoma de SARS-CoV-2 predichos por Predivac-3.0. El anillo está segmentado en quintiles donde cada quintil corresponde a 2 epítomos para ambos HLA. **Anillo 5)** clústeres de epítomos de células T de alta cobertura poblacional. El anillo está segmentado en cuartiles donde cada cuartil corresponde a 25% de cobertura (figura 14).



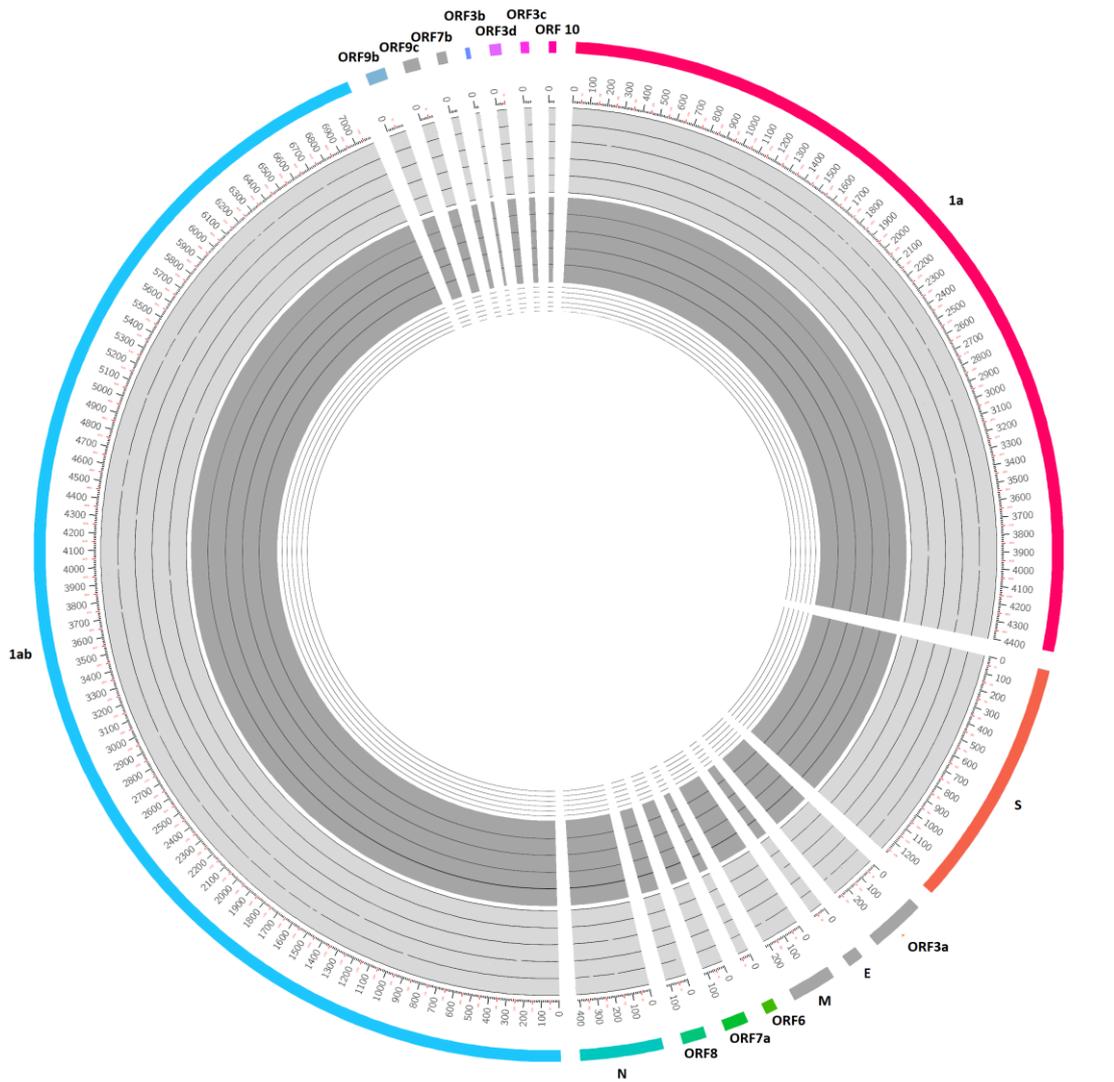
**Figura 14** Esquema circular para visualización de epítomos de células T según clase de HLA. Anillo más externo corresponde a las proteínas presentes en proteoma de SARS-CoV-2. Números indican posición de secuencia con respecto a la proteína a la que pertenece.

### 2.3.1. Análisis de epítomos de *hotspots* en SARS-CoV2 específicos para poblaciones en Chile

Para la evaluación y comparación de los perfiles de inmunidad entre las poblaciones detectadas en Chile, se construyeron esquemas circulares en Circos de 4 anillos concéntricos. El anillo más externo refiere a la secuencia aminoacídica de las proteínas que componen el proteoma de SARS-CoV-2. Los siguientes 2 anillos concéntricos están segmentados en quintiles, y el anillo más interno está segmentado en cuartiles. Tal como se presenta en la figura 15, los anillos representan diferentes capas de información sobre los epítomos predichos por Predivac-3.0. De afuera hacia adentro, el **anillo 1** entrega información sobre la posición de epítomos de células T para las poblaciones objetivo, el **anillo 2** corresponde a la densidad de epítomos de células T para las poblaciones objetivo. **Anillo 3**, clústeres de epítomos de células T de alta cobertura poblacional para las poblaciones objetivo.

Con el fin de analizar los resultados predichos en Predivac-3.0, se realizó una comparación de las secuencias y posiciones en el proteoma de SARS-CoV-2 con epítomos previamente identificados mediante métodos experimentales (en EpiTex). Para esto se construyeron tablas con las secuencias de los *hotspots* detectados por Predivac-3.0 y los epítomos de células T presentes en EpiTex y predichos por Predivac-3.0 ubicados dentro de estas regiones. A modo de sugerir *hotspots* como candidatos para el desarrollo de una vacuna, a partir de estas tablas se identificaron aquellos *hotspots* que cumplieran con las siguientes condiciones:

- I) concentran una cantidad mayor a 5 epítomos presentes en EpiTex de la respectiva proteína y
- II) Epítomos que concentran todos los epítomos descubiertos de EpiTex de la respectiva proteína.

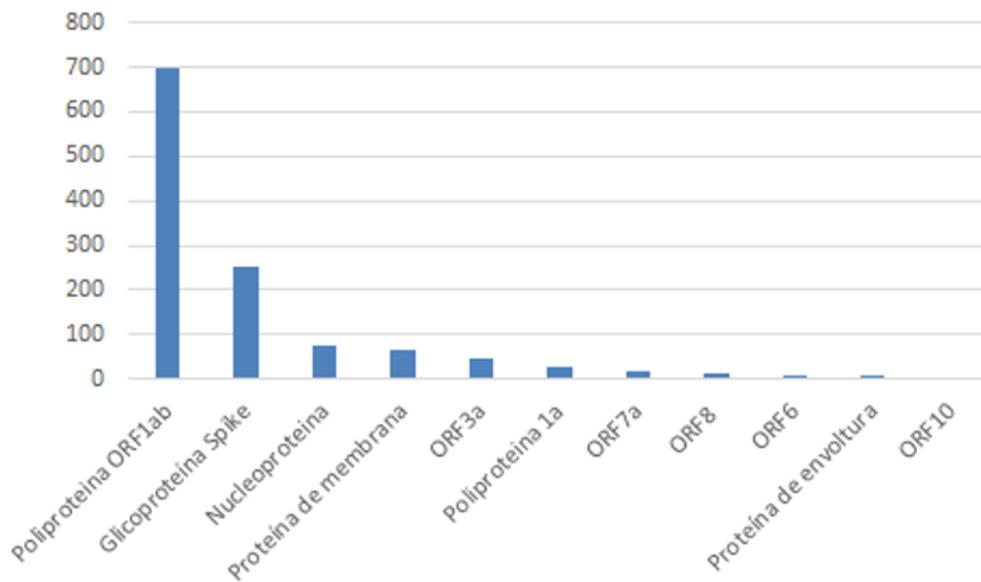


**Figura 15** Esquema circular para visualización de epítomos de células T según clase de HLA y utilizando criterio poblacional. Anillo más externo corresponde a las proteínas presentes en proteoma de SARS-CoV-2. Números indican posición de secuencia con respecto a la proteína a la que pertenece.

### 3. RESULTADOS

#### 3.1. Construcción de base de datos EpiEx.

La base de datos de IEDB, para el 2016, registraba cerca de 260.000 epítomos en su repositorio a través de una literatura de más de 18.000 referencias curadas, actualizándose continuamente hasta la fecha. Mediante un código construido en lenguaje “SQL” y “Python”, utilizando la librería “pandas”, se construyó un documento Excel en el que se almacenaron 1.210 epítomos de células T descubiertos experimentalmente para SARS-CoV-2 (IEDB ID: 2697049) en humanos, denominado como la base de datos “EpiEx”. Estos epítomos corresponden a todos los epítomos de células T descubiertos experimentalmente presentes en IEDB, los que están repartidos en 396 epítomos descubiertos para proteínas estructurales y 814 epítomos descubiertos para proteínas codificadas por ORF (Open Reading Frame) (figura 16).



**Figura 16** Gráfico de barras de cantidad de epítomos de células T descubiertos experimentalmente en proteínas de SARS-CoV-2, ordenado de forma descendente. Eje Y corresponde a cantidad de epítomos de EpiEx y Eje X corresponde a proteínas de SARS-CoV-2. Un total de 1.210 epítomos repartidos en 696 para ORF1ab, 246 para glicoproteína de superficie S, 75 para fosfoproteína de nucleocápside, 67 para glicoproteína de membrana, 47 para proteína codificada por ORF3a, 26 para poliproteína 1a, 16 para proteína codificada por ORF7a, 14 para proteína codificada por ORF8, 10 para proteína codificada por ORF6, 8 para proteína de envoltura y 5 para proteína codificada por ORF10.

### **3.1.1. Epítomos de células T de EpiTex en proteínas estructurales**

Las proteínas estructurales de SARS-CoV-2 son proteínas codificadas por genes ubicados dentro de la región que precede al extremo 3' del genoma. En SARS-CoV-2, estas proteínas constan de las proteínas S, N, M y E. La proteína S es la responsable de reconocer el receptor del hospedero para unirse a él, fusionándose con su membrana. Las investigaciones para el desarrollo de vacunas de tercera generación se han enfocado en esta proteína de 1.237 aminoácidos. De las proteínas estructurales, es la que concentra la mayor cantidad de epítomos de células T presentes en EpiTex con 246 epítomos. La proteína N es una proteína de 419 aminoácidos que se expresa en muestras hospederas durante la etapa temprana de infección; posee 3 dominios altamente conservados consistentes en un dominio N-terminal, un dominio de unión a ARN o región *linker*, y un dominio C-terminal. 75 epítomos de células T de EpiTex fueron los asociados a esta proteína. La proteína M es una proteína de 220 a 260 que se asocia con otras proteínas estructurales virales, facilitando el ensamblaje molecular de partículas virales. En esta proteína se registran 67 epítomos de células T de EpiTex. Finalmente, la proteína E es una proteína de membrana pequeña compuesta de un dominio N-terminal hidrofóbico y una cadena en el extremo C-terminal. Es la proteína más pequeña de las proteínas estructurales; posee entre 76 y 109 aminoácidos y juega un rol mayor en la patogénesis, ensamblaje del virus y liberación. Esta proteína es la que menos epítomos de células T de EpiTex registra con un total de 8 epítomos.

### **3.1.2. Epítomos de células T de EpiTex en proteínas codificadas por ORF**

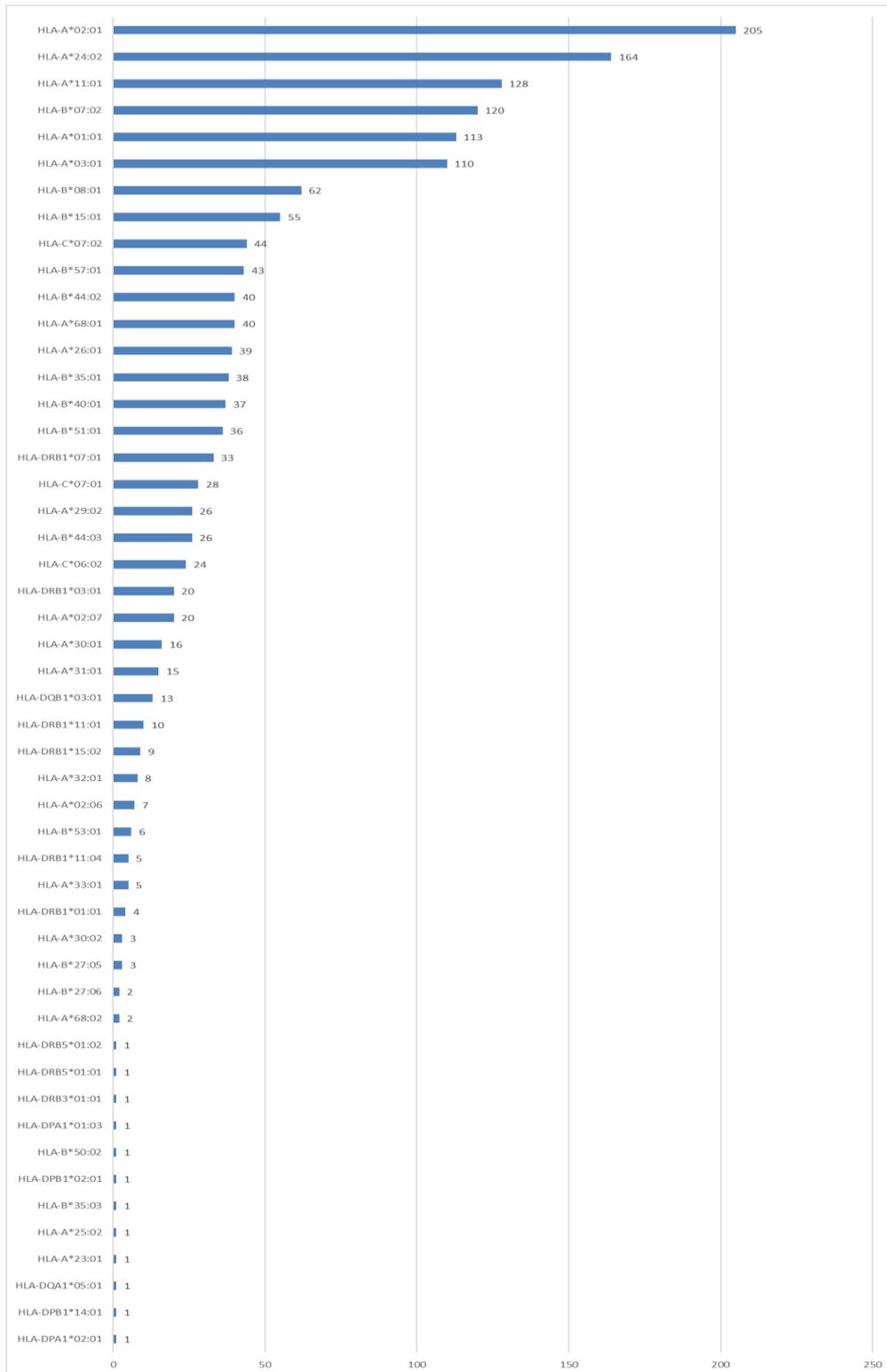
Las proteínas codificadas por ORF concentran cerca del 67% de los epítomos de células T de EpiTex, con ORF1ab abarcando cerca del 86% de ese segmento. ORF1ab (también referido como ORF1a y ab) abarca 2/3 de la región genómica viral de SARS-CoV-2 y registra un total 696 epítomos de células T de EpiTex. El gen ORF1a codifica para la poliproteína no estructural 1a que corresponde a las proteínas no estructurales NSP1 hasta NSP11. Se han detectado 26 epítomos de células T de EpiTex en ORF1a.

Con excepción de ORF1ab y ORF1a, todas las ORF forman parte de proteínas estructurales. Las proteínas accesorias codificadas de ORF3a, ORF6, ORF7a, ORF8 y ORF10, son proteínas codificadas por genes estructurales, que, junto con las proteínas estructurales y no estructurales, son las responsables de los aspectos funcionales y estructurales de SARS-CoV-2. ORF3a se encuentra ubicado entre los genes de S y E, y es la más grande de las proteínas

accesorias de SARS-CoV-2 con 274 aminoácidos. Esta proteína registra 47 epítomos de células T de EpiTex. ORF6 codifica para una proteína asociada a membrana de 61 aminoácidos y posee 10 de los epítomos de células T EpiTex. ORF7a codifica para una proteína de 122 aminoácidos y registra 16 de los epítomos de células T de EpiTex. ORF8 codifica una proteína de 121 aminoácidos a la que se le ha detectado interacciones con HLA-I, ayudando en la evasión inmune; registra 14 de los epítomos de células T de EpiTex. Finalmente, ORF10 codifica la proteína que menor epítomos de células T de EpiTex registra con 5 epítomos a lo largo de sus 38 aminoácidos. ORF10 es expresado exclusivamente por SARS-CoV-2 y se postula que jugaría un rol en la supresión de la respuesta inmune innata antiviral.

### **3.1.3. Restricciones HLA asociadas a los epítomos de células T de EpiTex**

La promiscuidad de un epítomo de célula T es el fenómeno de cuando un epítomo puede ser reconocido por más de un alelo HLA. De los 1.210 epítomos de EpiTex, se detectó un total de 1.571 restricciones HLA asociados a los epítomos de EpiTex, ordenados en un gráfico de epítomos por alelo HLA de forma descendente (figura 17). Para determinar la cantidad de restricciones HLA, se construyó un código en lenguaje “SQL” y “Python”, utilizando la biblioteca “pandas”, con lo que se organizó la información en un documento que entrega la cantidad de restricciones por alelo HLA. Estas 1.571 restricciones se reparten en 1.469 restricciones repartidas en 35 alelos HLA clase I, con HLA\*02:01 registrando la mayor cantidad de restricciones asociado a 205 epítomos de EpiTex; mientras que para HLA clase II hay 102 restricciones repartidas en 15 alelos, con HLA-DRB1\*07:01 registrando la mayor cantidad de restricciones asociado a 33 epítomos de EpiTex (Tabla 3).



**Figura 17** Gráfico de barras de restricciones HLA asociadas a epítomos de células T descubiertos experimentalmente de SARS-CoV-2. Eje Y corresponde a alelos HLA, y eje X corresponde a cantidad de restricciones asociados a epítomos de EpiTex. EpiTex registró un total 1.571 restricciones para 35 alelos HLA clase I y 15 alelos HLA clase II.

La tabla 3 organiza los alelos HLA con restricciones asociadas a epítomos de EpiTex en forma descendente y separados por clase I y II, junto con la frecuencia asociada de estos alelos a poblaciones con componente étnico en Chile de acuerdo con AFND presentes en la tabla 1. Los grupos étnicos detectados de estas poblaciones consisten en “Amerindian” (Isla de Pascua, Huilliche y Mapuche), “Mixed” (Chile Santiago Mixed), “Caucasoid” (Chile Santiago pop 2) e “Hispanic” (Chile Santiago). El grupo “Amerindian” consta de gente indígena de Islas del Caribe, América del Sur, Central y del Norte, antes de la llegada de los europeos. El grupo “Mixed” consiste en poblaciones cuya mezcla no fue posible de categorizar en otra categoría. El grupo “Caucasoid” consiste en individuos descendientes de Europa del Este y Europa del Oeste. El grupo “Hispanic” consiste en individuos descendientes de españoles o latinoamericanos.

Los cuadros en verde de la tabla 3 señalan la población que posee la mayor frecuencia del respectivo alelo HLA, con respecto al resto de las poblaciones en Chile ubicadas en columnas. Los alelos en negritas y cursivas presentes en la tabla 3 consisten en alelos que no estuvieron presente para ninguna de las poblaciones con componente étnico en Chile; para HLA clase I, esto ocurre con los alelos HLA-A\*02:07, HLA-B\*27:06, HLA-A\*25:02 y HLA-B\*50:02; para HLA clase II, esto ocurre con los alelos HLA-DPA1\*01:03, HLA-DPA1\*02:01, HLA-DPB1\*02:01, HLA-DPB\*14:01, HLA-DRB3\*01:01, HLA-DRB5\*01:01 y HLA-DRB5\*01:02.

A pesar de compartir el mismo grupo étnico que la población Mapuche y de Isla de Pascua, la población Huilliche (“Chile Huilliche” y “Chile Huilliche NA-DHS\_14 (G)”) no registra frecuencia alguna en los alelos HLA con restricciones asociadas a epítomos de EpiTex. Las poblaciones Mapuche, Isla de Pascua y Santiago Mixed son las que presentan más información con respecto a frecuencia de alelos que las poblaciones “Chile Santiago pop 2” y “Chile Santiago”. Estas 2 poblaciones corresponden a grupos étnicos “Hispanic” y “Caucasoid”, y solo se observan frecuencias en alelos HLA de clase II.

Clase HLA	Alelo HLA	Cantidad	Frecuencia de alelos en poblaciones de Chile (AFND)				
			Chile Easter Island (Amerindian)	Chile Mapuche (Amerindian)	Chile Santiago Mixed (Mixed)	Chile Santiago pop 2 (Hispanic)	Chile Santiago (Caucasoid)
HLA I	HLA-A*02:01	205	9,50%	<b>22,38%</b>	16,33%	-	-
	HLA-A*24:02	164	<b>35,80%</b>	3,08%	8,34%	-	-
	HLA-A*11:01	128	<b>38,10%</b>	1,54%	1,00%	-	-
	HLA-B*07:02	120	-	5,38%	<b>6,19%</b>	-	-
	HLA-A*01:01	113	-	<b>6,15%</b>	5,66%	-	-
	HLA-A*03:01	110	-	<b>8,46%</b>	6,72%	-	-
	HLA-B*08:01	62	-	<b>7,69%</b>	2,02%	-	-
	HLA-B*15:01	55	-	<b>3,08%</b>	1,00%	-	-
	HLA-C*07:02	44	-	-	<b>13,97%</b>	-	-
	HLA-B*57:01	43	-	<b>3,08%</b>	2,02%	-	-
	HLA-A*68:01	40	-	<b>35,38%</b>	2,02%	-	-
	HLA-B*44:02	40	<b>7,10%</b>	1,54%	3,04%	-	-
	HLA-A*26:01	39	-	-	<b>3,04%</b>	-	-
	HLA-B*35:01	38	-	<b>6,15%</b>	1,51%	-	-
	HLA-B*40:01	37	<b>19,00%</b>	-	1,00%	-	-
	HLA-B*51:01	36	-	<b>8,46%</b>	3,56%	-	-
	HLA-C*07:01	28	-	-	<b>5,13%</b>	-	-
	HLA-A*29:02	26	-	-	<b>5,13%</b>	-	-
	HLA-B*44:03	26	-	2,38%	<b>3,56%</b>	-	-
	HLA-C*06:02	24	-	-	<b>5,13%</b>	-	-
	<b>HLA-A*02:07</b>	20	-	-	-	-	-
	HLA-A*30:01	16	-	-	<b>0,50%</b>	-	-
	HLA-A*31:01	15	-	<b>6,15%</b>	-	-	-
	HLA-A*32:01	8	-	<b>1,54%</b>	0,50%	-	-
	HLA-A*02:06	7	-	-	<b>0,50%</b>	-	-
	HLA-B*53:01	6	-	-	<b>1,00%</b>	-	-
	HLA-A*33:01	5	-	<b>3,85%</b>	0,50%	-	-
	HLA-A*30:02	3	-	0,77%	<b>3,04%</b>	-	-
	HLA-B*27:05	3	-	<b>3,08%</b>	1,51%	-	-
	HLA-A*68:02	2	-	-	<b>1,00%</b>	-	-
<b>HLA-B*27:06</b>	2	-	-	-	-	-	
HLA-A*23:01	1	-	0,77%	<b>1,51%</b>	-	-	
<b>HLA-A*25:02</b>	1	-	-	-	-	-	
HLA-B*35:03	1	-	<b>0,77%</b>	-	-	-	
<b>HLA-B*50:02</b>	1	-	-	-	-	-	
<b>TOTAL</b>	<b>1469</b>						
HLA II	HLA-DRB1*07:01	33	-	<b>9,23%</b>	5,13%	-	9,70%
	HLA-DRB1*03:01	20	-	<b>6,15%</b>	5,14%	-	5,90%
	HLA-DQB1*03:01	13	-	<b>31,54%</b>	18,76%	25,00%	25,00%
	HLA-DRB1*11:01	10	-	3,08%	<b>4,08%</b>	-	3,24%
	HLA-DRB1*15:02	9	-	0,77%	<b>1,51%</b>	-	1,09%
	HLA-DRB1*11:04	5	-	0,77%	2,02%	-	<b>3,30%</b>
	HLA-DRB1*01:01	4	-	1,54%	1,51%	-	<b>2,33%</b>
	<b>HLA-DPA1*01:03</b>	1	-	-	-	-	-
	<b>HLA-DPA1*02:01</b>	1	-	-	-	-	-
	<b>HLA-DPB1*02:01</b>	1	-	-	-	-	-
	<b>HLA-DPB1*14:01</b>	1	-	-	-	-	-
	HLA-DQA1*05:01	1	-	-	-	<b>16,90%</b>	<b>16,90%</b>
	<b>HLA-DRB3*01:01</b>	1	-	-	-	-	-
<b>HLA-DRB5*01:01</b>	1	-	-	-	-	-	
<b>HLA-DRB5*01:02</b>	1	-	-	-	-	-	
<b>TOTAL</b>	<b>102</b>						

**Tabla 3** Cantidad de epítomos asociados a restricción HLA con respectiva cobertura poblacional en Chile de acuerdo con AFND (<http://www.allelefrequencias.net/>). De las 1.571 restricciones, 1.469 están asociados a HLA clase I y 102 están asociados a HLA clase II. En columna "Alelos HLA", los alelos en negritas y cursiva corresponden a alelos que no están presente en ninguna de las poblaciones identificadas en Chile de acuerdo con AFND. Cuadros en verde corresponde a la población con mayor porcentaje de frecuencia de alelos. Poblaciones de grupos étnicos "Hispanic" y "Caucasoid" solo presentan frecuencia en alelos HLA de clase II. La población Huilliche ("Chile Huilliche" y "Chile Huilliche NA-DHS\_14 (G)") no registra frecuencia alguna en los alelos HLA con restricciones asociadas a epítomos de EpitEx, por lo que no se muestran en la tabla.

El portal AFND proporciona una lista completa de la distribución de frecuencias de los alelos más prevalentes para los loci HLA-A, -B, -C y -DBR1 a nivel mundial (<http://www.allelefreqencies.net/top10dist.asp>). Comparando esta información con lo almacenado en EpiTex, de 35 alelos HLA clase I, 4 no corresponden a alelos prevalentes a nivel global: HLA-A\*33:01, HLA-B\*27:06, HLA-A\*25:02 y HLA-B\*50:02, mientras que de 6 restricciones HLA clase II asociadas a loci -DBR1, todos corresponden al conjunto de alelos más prevalentes.

### 3.2. Mapeo *in silico* de epítomos de SARS-CoV-2

Se utilizó el programa Circos (<http://circos.ca/>) para visualizar datos e información en un diseño circular, permitiendo explorar las relaciones entre objetos o posiciones. Esta herramienta está presente en Galaxy (<https://galaxyproject.eu/>), una plataforma abierta basada en la web para la investigación biológica computacional accesible, reproducible y transparente. La interfaz que ofrece Galaxy para trabajar con Circos permite trabajar con el visualizador de forma online, lo que evita el procedimiento de instalación de Circos en el equipo, además de facilitar las opciones que ofrece este software para la visualización. Fue mediante estos recursos que se visualizó el mapeo *in silico* de epítomos de células T CD4+ y CD8+ de EpiTex en el proteoma de SARS-CoV-2, además del mapeo *in silico* de los epítomos de células T CD4+ y CD8+ para SARS-CoV-2 predichos por Predivac-3.0. La visualización graficada en esquemas circulares incluye clústeres y *hotspots* de epítomos de células T CD4+ y CD8+.

Se construyeron 2 esquemas para visualizar el mapeo *in silico* de epítomos de células T CD8+ (HLA clase I) (figura 18) y epítomos de células T CD4+ (HLA clase II) (figura 19), además de la densidad de estos epítomos y *hotspots* de epítomos a lo largo de las proteínas presentes en el proteoma de SARS-CoV-2, presente en la base de datos exhaustiva de secuencias proteicas e información de anotaciones Uniprot (Uniprot ID: UP000464024; <https://www.uniprot.org/proteomes/UP000464024>) (Morgat et al., 2020). Uniprot permite la obtención de proteomas en formato FASTA, un formato de documento de texto que permite el análisis de secuencias en herramientas de procesamiento de texto y lenguajes de programación como Python.

### 3.2.1. Mapeo de epítomos de células T CD8+

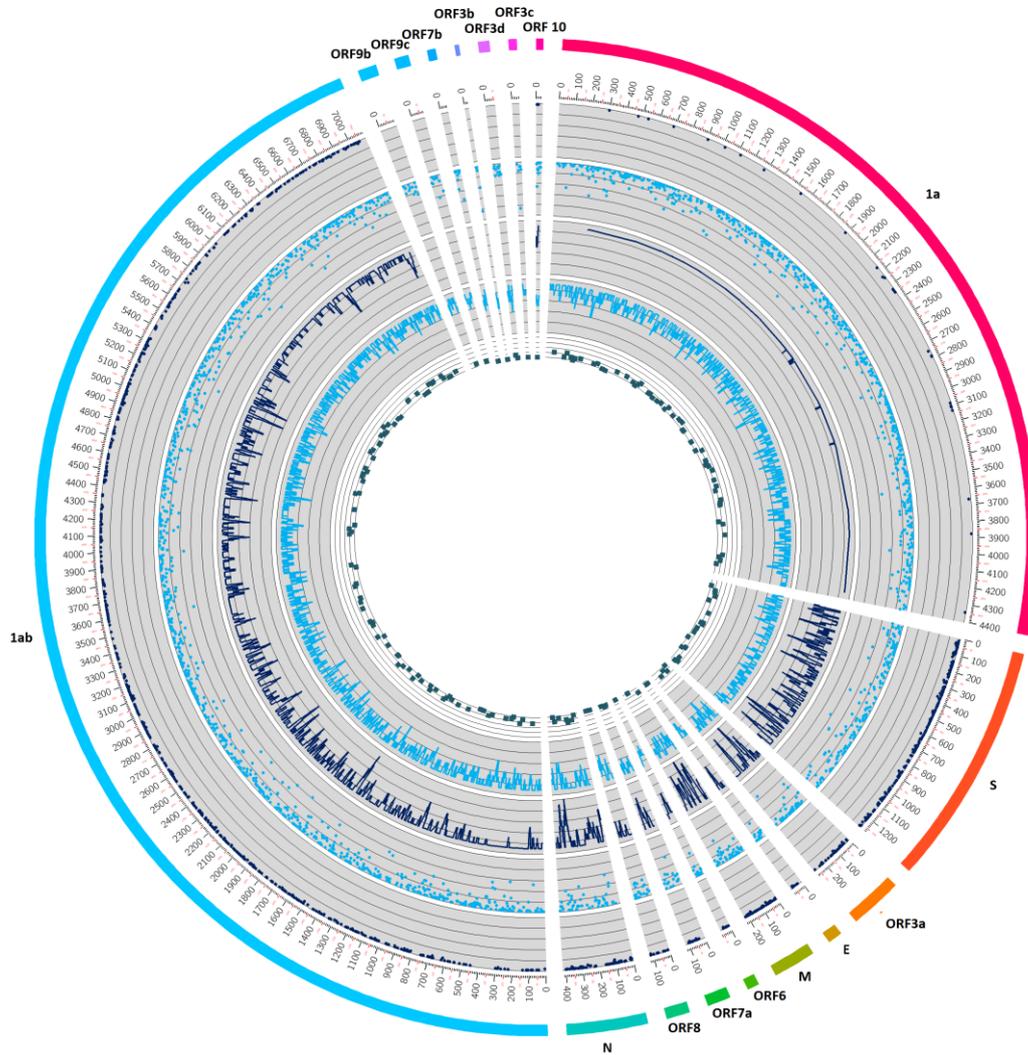
La respuesta inmune mediada por células T CD8+ es crítica frente a infecciones virales y puede ser independiente de la respuesta inmune mediada por células T CD4+. Los epítomos involucrados en esta respuesta constan de péptidos de 8 a 10 aminoácidos y son presentados por HLA de clase I. A fin de comparar la información internacional con la situación en Chile, se graficó en la figura 18 los resultados de mapeo de epítomos de células T CD8+ de SARS-CoV-2 presentes en la base de datos EpiTex, además de las predicciones realizadas en Predivac-3.0 para epítomos de células T CD8+ en Chile como población objetivo.

El anillo más externo comprende las proteínas presentes en el proteoma de SARS-CoV-2. De afuera hacia adentro, los primeros 2 anillos del esquema (seguido del anillo de las proteínas) abordan la posición de los epítomos de células T de EpiTex y Predivac-3.0 junto con la cantidad de restricciones HLA asociadas al epítomo. En este segmento se aprecia no solo la diferencia de la presencia de epítomos a lo largo del proteoma de SARS-CoV-2, sino que la cantidad de restricciones HLA asociadas a los epítomos es mayor en el segmento que comprende Predivac-3.0 (anillo 2) que EpiTex (anillo 1). La ausencia de epítomos de células T descubiertos experimentalmente se observa en las proteínas codificadas por ORF3c, ORF3d, ORF3b, ORF7b, ORF9c y ORF9b. Estas son proteínas accesorias que se producen a partir de 5 genes codificantes ORF (ORF3a, ORF6, ORF7a, ORF7b y ORF8) (Yadav et al., 2021).

Los anillos 3 y 4 abordan la densidad de epítomos a lo largo de las proteínas del proteoma de SARS-CoV-2 de EpiTex y Predivac-3.0, respectivamente. Se observa una concentración considerable de epítomos de células T de EpiTex en las proteínas estructurales de SARS-CoV-2, a diferencia de la densidad observada de los resultados obtenidos de Predivac-3.0 (anillo 3), los que presentan una densidad menor que lo observado en EpiTex para las proteínas estructurales. La proteína S presenta zonas que llegan a los 9-10 epítomos para EpiTex, mientras que para Predivac-3.0, esta densidad no llega a superar los 4 epítomos a lo largo de toda la secuencia de la proteína S. Similar es el caso para las proteínas M, N, y las proteínas codificadas por ORF3a y ORF7a. ORF1a presenta una baja densidad de epítomos de la mano de poca cantidad de epítomos de células T que se observa en anillo 1. ORF1ab presenta epítomos de células T y densidad considerable a lo largo de toda su secuencia tanto para EpiTex como para Predivac-3.0.

De acuerdo con Predivac-3.0, la cantidad de clústeres de epítomos de células T CD8+ cuya cobertura es superior al 90% (*hotspots*) de la población de Chile, comprenden: ORF1ab (32),

ORF1a (19), S (4), N (3), ORF7a (3), ORF8 (2), M (1), ORF10 (1), ORF3a (1), ORF3b (1), ORF3c (1), ORF3d (1), ORF6 (1), ORF7b (1).

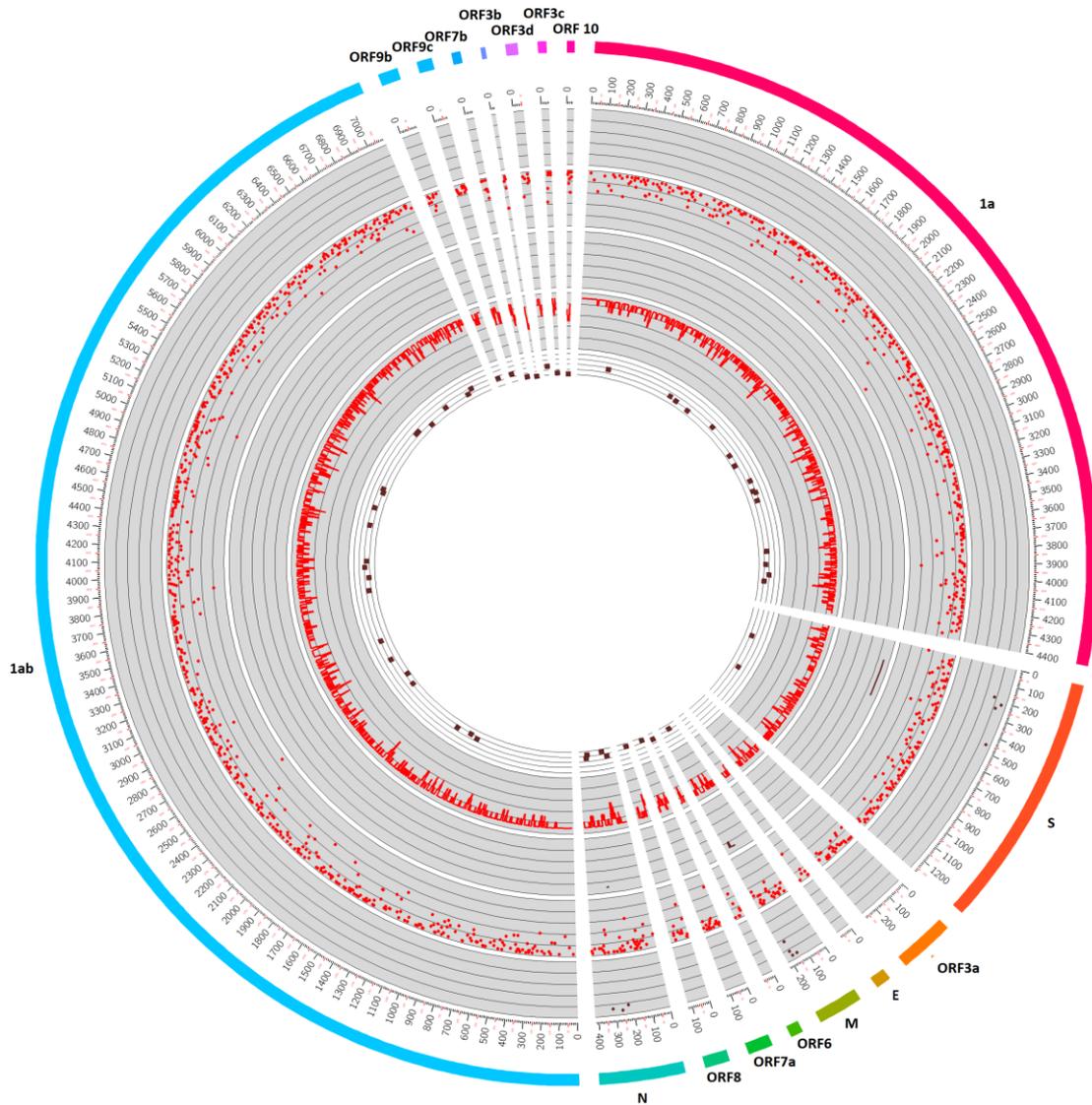


**Figura 18** Visualización circular de mapeo de epítomos de células T CD8+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) realizado con Circos. Anillo más externo comprende a las proteínas del proteoma de SARS-CoV-2. Los anillos concéntricos siguientes de afuera hacia adentro corresponden a: **anillo 1**, posición de epítomos CD8+ de EpitEx (puntos azules), cada quintil del anillo corresponde a 10 restricciones HLA asociados al epítomo, **anillo 2**, posición de epítomos CD8+ predichos por Predivac-3.0 (puntos celestes), cada quintil del anillo corresponde a 10 restricciones HLA asociados al epítomo, **anillo 3**, densidad de epítomos superpuestos CD8+ de EpitEx (línea celeste), cada quintil del anillo corresponde a 2 epítomos, **anillo 4**, densidad de epítomos superpuestos CD8+ predichos por PREDIVAC (línea azul), cada quintil del anillo corresponde a 2 epítomos, **anillo 5**, clústeres de epítomos CD8+ (puntos azules) de alta cobertura poblacional, cada cuartil del anillo corresponde a 25% de cobertura.

### 3.2.2. Mapeo de epítomos de células T CD4+

La figura 19 presenta en un gráfico circular, los resultados de mapeo de epítomos de células T CD4+ de SARS-CoV-2 presentes en la base de datos EpiTex, los cuales consisten en péptidos de 13 a 25 aminoácidos presentados por HLA clase II. También presenta las predicciones realizadas en Predivac-3.0 para epítomos de células T CD4+ con Chile como población objetivo.

En comparación con la información de epítomos de células T CD8+ EpiTex, la información de epítomos de células T CD4+ descubiertos experimentalmente es considerablemente menor. La poca información disponible de epítomos de células T CD4+ en proteoma de SARS-CoV-2 se concentra en las proteínas estructurales S, M y N, contrastando con los resultados de la predicción realizada por Predivac-3.0 para epítomos de células T CD4+, en donde se encuentra información disponible en todas las proteínas del proteoma de SARS-CoV-2. De igual manera, las predicciones de Predivac-3.0 para *hotspots* de epítomos de células T CD4+ entrega menor cantidad de resultados que los observados en los epítomos de células T CD8+. Los clústeres de epítomos de células T CD4+ que cubren más del 90% de la población de Chile, comprenden: ORF1a (1), N (1), ORF10 (1), ORF3b (1), ORF3c (1), ORF7b (1). ORF3a es la única proteína del proteoma de SARS-CoV-2 que no presenta clústeres de epítomos de células T CD4+ para Chile, de acuerdo a los resultados de Predivac-3.0.



**Figura 19** Visualización circular de mapeo de epítomos de células T CD4+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) realizado con Circos. Anillo más externo comprende a las proteínas del proteoma de SARS-CoV-2. Los anillos concéntricos siguientes de afuera hacia adentro corresponden: **anillo 1**, posición de epítomos CD4+ de **EpiTex** (puntos burdeos); cada quintil del anillo corresponde a 5 restricciones HLA asociados al epítomo, **anillo 2**, posición de epítomos CD4+ predichos por Predivac-3.0 (puntos rojos); cada quintil del anillo corresponde a 5 restricciones HLA asociados al epítomo, **anillo 3**, densidad de epítomos superpuestos CD4+ de **EpiTex** (línea burdeo); cada quintil del anillo corresponde a 2 epítomos, **anillo 4**, densidad de epítomos superpuestos CD4+ predichos por Predivac-3.0 (línea roja); cada quintil del anillo corresponde a 2 epítomos, **anillo 5**, clústeres de epítomos CD4+ (cuadrados rojos) de alta cobertura poblacional; cada cuartil del anillo corresponde a 25% de cobertura.

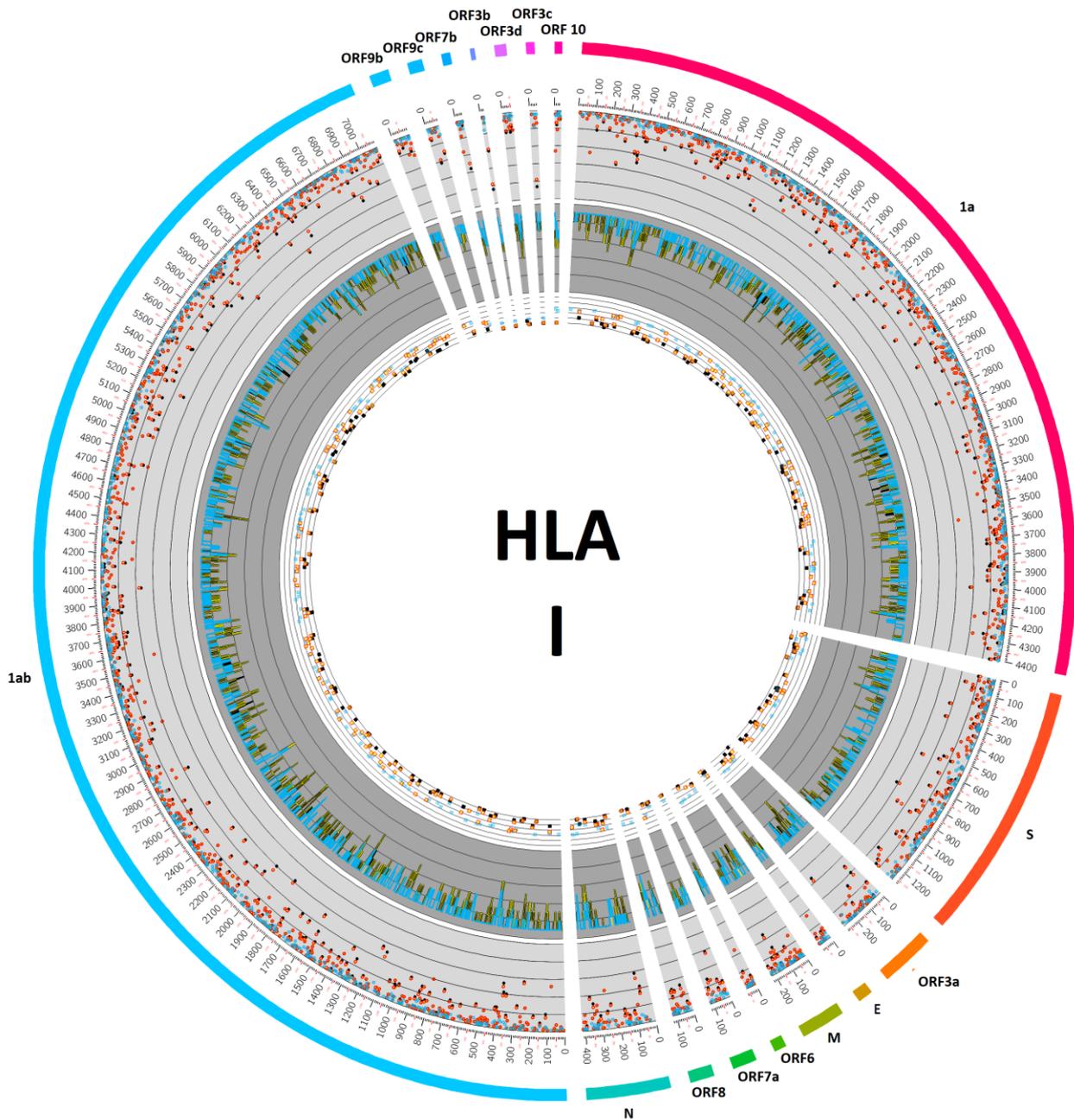
### 3.3. Predicción de epítomos de células T considerando componente étnico en Chile

De acuerdo con AFND, las poblaciones con componente étnico en Chile abarcan los grupos étnicos “Mixed”, “Amerindian”, “Hispanic” y “Caucasoid”. Predivac-3.0 registra resultados solo para los grupos étnicos “Mixed” y “Amerindian” en Chile, las que fueron designadas como población objetivo para la consulta en Predivac-3.0 junto con la población chilena. Utilizando Circos mediante Galaxy se construyeron esquemas circulares de 4 anillos concéntricos para el estudio de cantidad, densidad, clústeres y *hotspots* de epítomos de células T CD4+ y CD8+ para las poblaciones de Chile, Chile Mixed y Chile Amerindian. Se organizó la información en tablas para abordar las secuencias de los *hotspots* de epítomos de células T CD4+ y CD8+ predichos por Predivac-3.0 cuya cobertura poblacional es superior al 90%, la posición de estos *hotspots* y la cantidad de epítomos de células T dentro de ellos tanto de las predicciones de Predivac-3.0 como de los presentes en EpiTex. El criterio utilizado para determinar los epítomos de células T de EpiTex situados dentro de los *hotspots* predichos por Predivac-3.0, consistió en considerar todos aquellos epítomos cuya secuencia estuviera completa dentro del *hotspot*. Es decir, si el epítomo de EpiTex posee 1 residuo o más fuera de la secuencia del *hotspot*, éste no fue considerado.

#### 3.3.1. Predicción de epítomos y *hotspots* de células T CD8+

La figura 20 presenta en un gráfico circular el mapeo de epítomos de células T CD8+ con información sobre posición, densidad y clústeres de epítomos de células T CD8+ predichos por Predivac-3.0 para la población de Chile, Chile Mixed y Chile Amerindian. Se incluyeron clústeres de epítomos de células T CD8+ cuya cobertura poblacional es mayor a 30% debido a la baja cobertura observada en la población Chile Amerindian, a modo de tener una perspectiva más completa de la situación con respecto a esa población. Esta baja cobertura es tal que ningún clúster de epítomos de células T CD8+ superó el criterio de tener una cobertura poblacional mayor al 90%, indicando ausencia absoluta a lo largo del proteoma de SARS-CoV-2 de *hotspots* para la población de Chile Amerindian. La población Chile Mixed posee un perfil muy similar a la población de Chile, en lo que respecta a la posición de epítomos y densidad de epítomos de células T CD8+. Esto se infiere a través del anillo de la figura 20 que aborda la posición de los epítomos y la cantidad de restricciones HLA asociadas a esos epítomos (anillo 1), donde se observa que epítomos ubicados a la misma altura para Chile y Chile Mixed poseen una cantidad de restricciones HLA muy similar. Lo mismo se observa en el anillo que aborda la densidad de

los epítomos de células T CD8+. Es en los *hotspots* de epítomos donde se observa mayor diferencia entre Chile y Chile Mixed.



**Figura 20** Gráfico circular de mapeo de epítomos de células T CD8+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) para Chile y grupos étnicos “Mixed” y “Amerindian”, realizado con Circos. **Anillo 1**, posición de epítomos CD8+ para Chile puntos negros), grupo étnico “Mixed” (puntos amarillos con bordes rojos) y grupo étnico “Amerindian” respectivamente (puntos celestes), **anillo 2**, corresponden a densidad de epítomos CD8+ para Chile (línea negra), grupo étnico “Mixed” (línea amarilla) y grupo étnico “Amerindian” (línea celeste), **anillo 3**, clústeres de epítomos de células T CD8+ de alta cobertura poblacional para Chile (cuadrados negros), grupo étnico “Mixed” (cuadrados amarillos de bordes rojos) y grupo étnico “Amerindian” (cuadrados celestes).

### 3.3.1.1. Proteínas estructurales

En la tabla 4 se desglosa la información de los *hotspots* de epítomos de células T CD8+ presentes en proteínas estructurales de SARS-CoV-2, donde se muestra la secuencia, posición y cobertura de los *hotspots*, cantidad de restricciones HLA asociado a los epítomos de células T CD8+ dentro del *hotspot* predichos por Predivac-3.0, además de la cantidad de epítomos de células T CD8+ de Predivac-3.0 y EpiTex ubicados dentro de la región del *hotspot*, para las poblaciones de Chile y Chile Mixed, ya que Chile Amerindian no registra *hotspots* a lo largo de las proteínas estructurales de SARS-CoV-2. Dentro de los *hotspots* expuestos en la tabla 4, se señalan en naranja aquellos *hotspots* que presentan una cantidad de epítomos de células T CD8+ de EpiTex superior a 5 epítomos. Chile y Chile Mixed comparten una similitud considerable en varias proteínas de SARS-CoV\_2 respecto a los *hotspots* de células T CD8+ presentes en cada población.

La proteína S es la proteína estructural con más epítomos de EpiTex con un total de 246 epítomos de EpiTex, teniendo una considerable densidad de epítomos a lo largo de su secuencia. Aun así, solo 4 *hotspots* fueron los detectados en esta proteína, los que contienen 15 epítomos de EpiTex. Uno de estos *hotspots* posee 8 epítomos de EpiTex en su interior. Para el caso de Chile Mixed, son 2 los *hotspots* presentes para la proteína S, los que abarcan 18 epítomos repartidos en 11 y 7 epítomos de células T CD8+ para ambos *hotspots*. Los *hotspots* de Chile y Chile Mixed no comparten posición en la secuencia de la proteína S.

La proteína N con 75 epítomos de EpiTex, presenta 3 *hotspots* para la población de Chile, los que poseen un total de 7 epítomos de células T CD8+ de EpiTex en su interior; uno de estos *hotspots* abarca 6 de los 7 epítomos. De igual manera son 3 los *hotspots* encontrados en la proteína N para Chile Mixed, los que abordarían una cantidad mayor de epítomos que la población de Chile con 21 epítomos de células T CD8+ de EpiTex. Uno de estos *hotspots* posee un total de 13 de los 21 epítomos descubiertos, el que además compartiría secuencia y posición con el *hotspot* destacado de la población de Chile.

En la proteína estructural M de 67 epítomos de EpiTex, existe 1 *hotspot* tanto para la población de Chile como de Chile Mixed, los que compartirían una posición muy similar en la secuencia de la proteína. La diferencia recae en la cantidad de epítomos de células T CD8+ presente en cada *hotspot*, donde la población de Chile contendría 10 epítomos de células T CD8+

de EpiTex, mientras que para Chile Mixed, la cantidad de epítomos de células T CD8+ de EpiTex en su interior, son 13.

La proteína E que contiene 8 epítomos de EpiTex, posee 1 *hotspot* sólo para la población Chile Mixed, que abarca un tamaño considerable dentro de la secuencia de la proteína E. Este *hotspot* contiene 7 de los 8 epítomos de EpiTex.

Proteína	Población	Secuencia	Posición residuos		Cobertura Poblacional	Alelos HLA	N° de epítomos		
			Inicio	Fin			Predivac	IEDB	Total IEDB
S	Chile	SGTNGTKRFDNPVLPFNDGVYFA	71	93	98,97%	36	5	4	15
		GYLQPRFTLLKYENGTITDAVDCALDPLSETK	278	300	91,06%	38	6	1	
		SALLAGTITSGWTFGAGAALQIP	875	897	99,11%	47	5	2	
		<b>KGYHLMSFPQSAPHGVVFLHVTY</b>	<b>1045</b>	<b>1067</b>	<b>96,81%</b>	<b>35</b>	<b>5</b>	<b>8</b>	
	Chile Mixed	<b>HSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFA</b>	<b>49</b>	<b>93</b>	<b>97,66%</b>	<b>44</b>	<b>9</b>	<b>11</b>	18
		<b>AYYVGYLQPRFTLLKYENGTITDAVDCALDPLSETK</b>	<b>264</b>	<b>307</b>	<b>92,03%</b>	<b>42</b>	<b>10</b>	<b>7</b>	
N	Chile	PNNTASWFTALTQHG	46	60	96,24%	43	4	1	7
		RQKRTATKAYNVTQ	259	272	93,87%	36	3	0	
		<b>AIKLDKDPNFKDQVILLNKHIDAYKTFPPT</b>	<b>336</b>	<b>366</b>	<b>97,82%</b>	<b>43</b>	<b>9</b>	<b>6</b>	
	Chile Mixed	MSDNGPQNQRNAPRITFGGSDSTGSNQNGERSGARS	0	37	99,15%	64	10	4	21
		RPQGLPNNTASWFTALTQHG	41	60	90,91%	41	4	4	
		<b>GMSRIGMEVTPSGTWLTYTGAIKLDKDPNFKDQVILLNKHIDAYKTFPPT</b>	<b>316</b>	<b>366</b>	<b>98,57%</b>	<b>56</b>	<b>13</b>	<b>13</b>	
M	Chile	<b>LTRP LLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATS</b>	<b>129</b>	<b>173</b>	<b>99,84%</b>	<b>54</b>	<b>12</b>	<b>10</b>	<b>10</b>
	Chile Mixed	<b>ESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLG A</b>	<b>135</b>	<b>181</b>	<b>92,84%</b>	<b>50</b>	<b>12</b>	<b>13</b>	<b>13</b>
E	Chile Mixed	<b>LLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKNLNSS</b>	<b>27</b>	<b>68</b>	<b>97,63%</b>	<b>47</b>	<b>12</b>	<b>7</b>	<b>7</b>

**Tabla 4** Desglose de clústeres de epítomos de células T inmunodominantes de cobertura poblacional superior a 90% (*hotspots* de células T CD8+) predichos por Predivac-3.0 en proteínas estructurales de SARS-CoV-2 para la población de Chile y Chile Mixed. Se destacan en naranja los *hotspots* de células T CD8+ que contienen más de 5 epítomos de EpiEx (columna IEDB) en su interior (bajo el criterio de que estos epítomos no tengan ningún residuo fuera de la secuencia del *hotspot*). Columna “Alelos HLA” corresponde a la cantidad de restricciones HLA asociados a los epítomos de células T CD8+ ubicados dentro del *hotspot*, de acuerdo con los resultados de Predivac-3.0. Columna “Secuencia” corresponde a la secuencia aminoacídica de los *hotspots*. No se encontraron *hotspots* para población de Chile Amerindian.

### 3.3.1.2. Proteínas codificadas por ORF

En la tabla 5 se desglosa la información de los *hotspots* de epítomos de células T CD8+ presentes en proteínas codificadas por ORF de SARS-CoV-2, donde se muestra la secuencia, posición y cobertura de los *hotspots*, la cantidad de restricciones HLA asociado a los epítomos de células T CD8+ dentro del *hotspot* predichos por Predivac-3.0, la cantidad de epítomos de células T CD8+ de Predivac-3.0 y EpitepDB ubicados dentro del *hotspot* para las poblaciones de Chile y Chile Mixed. Por otra parte, Chile Amerindian no registra *hotspots* a lo largo de las proteínas codificadas por ORF de SARS-CoV-2. Dentro de los *hotspots* expuestos en la tabla 5, se señalan en naranja aquellos *hotspots* que presentan una cantidad de epítomos de células T CD8+ de Epitep superior a 5 epítomos, y en verde aquellos *hotspots* que poseen la totalidad de epítomos de células T CD8+ de Epitep descubiertos para la respectiva proteína. Chile y Chile Mixed comparten una similitud considerable en varias proteínas de SARS-CoV-2 respecto a los *hotspots* de células T CD8+ presentes en cada población.

El ORF1ab registra un total de 32 *hotspots* para la población de Chile dentro de los cuales están ubicados un total de 81 epítomos de células T CD8+ de Epitep de los 696 epítomos de células T de Epitep presentes en ORF1ab; 4 de estos *hotspots* poseen más de 5 epítomos de células T CD8+ de Epitep en su interior. Para el caso de Chile Mixed, son 10 los *hotspots* detectados en ORF1ab, los que poseen un total de 41 epítomos de células T CD8+ de Epitep en su interior. De estos *hotspots*, 3 poseen más de 5 epítomos de células T CD8+ de Epitep en su interior. Ambas poblaciones comparten similitud en posición y secuencia con 2 *hotspots* que poseen más de 5 epítomos de células T CD8+ de Epitep en su interior. Uno de estos *hotspots* posee la misma secuencia y posición para ambas poblaciones, y se encontraría ubicado desde el residuo 1627 al 1654 de ORF1ab.

El ORF1a posee una cantidad de 19 *hotspots* para la población de Chile; a pesar de la cantidad de *hotspots*, ninguno cumple el criterio de superar la cantidad de 5 epítomos de células T CD8 de Epitep en su interior, puesto que sólo se encuentran 2 epítomos de células T CD8+ de EpitepDB en estos *hotspots*, del total de 26 epítomos presentes en ORF1a. Para Chile Mixed son 6 los *hotspots* detectados en ORF1a, en los cuales no hay epítomos de Epitep en su interior.

El ORF7a, que posee 16 epítomos de Epitep, contiene 3 *hotspots* para la población de Chile y 2 *hotspots* para la población Chile Mixed. Ambas poblaciones contendrían dentro de sus

*hotspots* un total de 8 epítomos de células T CD8+ de EpiTex, en donde ambas poblaciones también poseen el mismo *hotspot* que contiene 7 epítomos de células T CD8+ de EpiTex.

El ORF8 contiene 14 epítomos de células T de EpiTex. Presenta 2 *hotspot* de cobertura cercana al 100% para la población de Chile, los que abarcarían un total de 8 epítomos de epítomos de células T de EpiTex, sin embargo, ningún *hotspot* cumpliría el criterio de poseer más de 5 epítomos de EpiTex en su interior. En Chile Mixed se detectó 1 *hotspot* que contendría en su interior 8 epítomos para células T CD4+ y CD8+ de EpiTex.

El ORF6 contiene 10 epítomos de células T de EpiTex y presenta 1 *hotspot* para cada población (Chile y Chile Mixed). El *hotspot* presente en la población de Chile contiene todos los epítomos de células T de EpiTex, además de cubrir al 100% de la población. El *hotspot* presente en Chile Mixed comparte una posición similar con el *hotspot* presente en la población de Chile, con la única diferencia en el residuo donde comienza la secuencia del *hotspot*. El *hotspot* de Chile Mixed contiene 7 epítomos de células T CD8+ de EpiTex.

El ORF10 contiene 5 epítomos de células T de EpiTex. Presenta 1 *hotspot* para la población de Chile y Chile Mixed, los que compartirían la misma secuencia con la leve diferencia de que el *hotspot* de Chile termina en el residuo 45, mientras que el de Chile Mixed termina en el residuo 47. Ambos *hotspot* abarcan la totalidad de los epítomos de células T de EpiTex presentes en ORF10 y cubrirían aproximadamente la totalidad de ambas poblaciones.

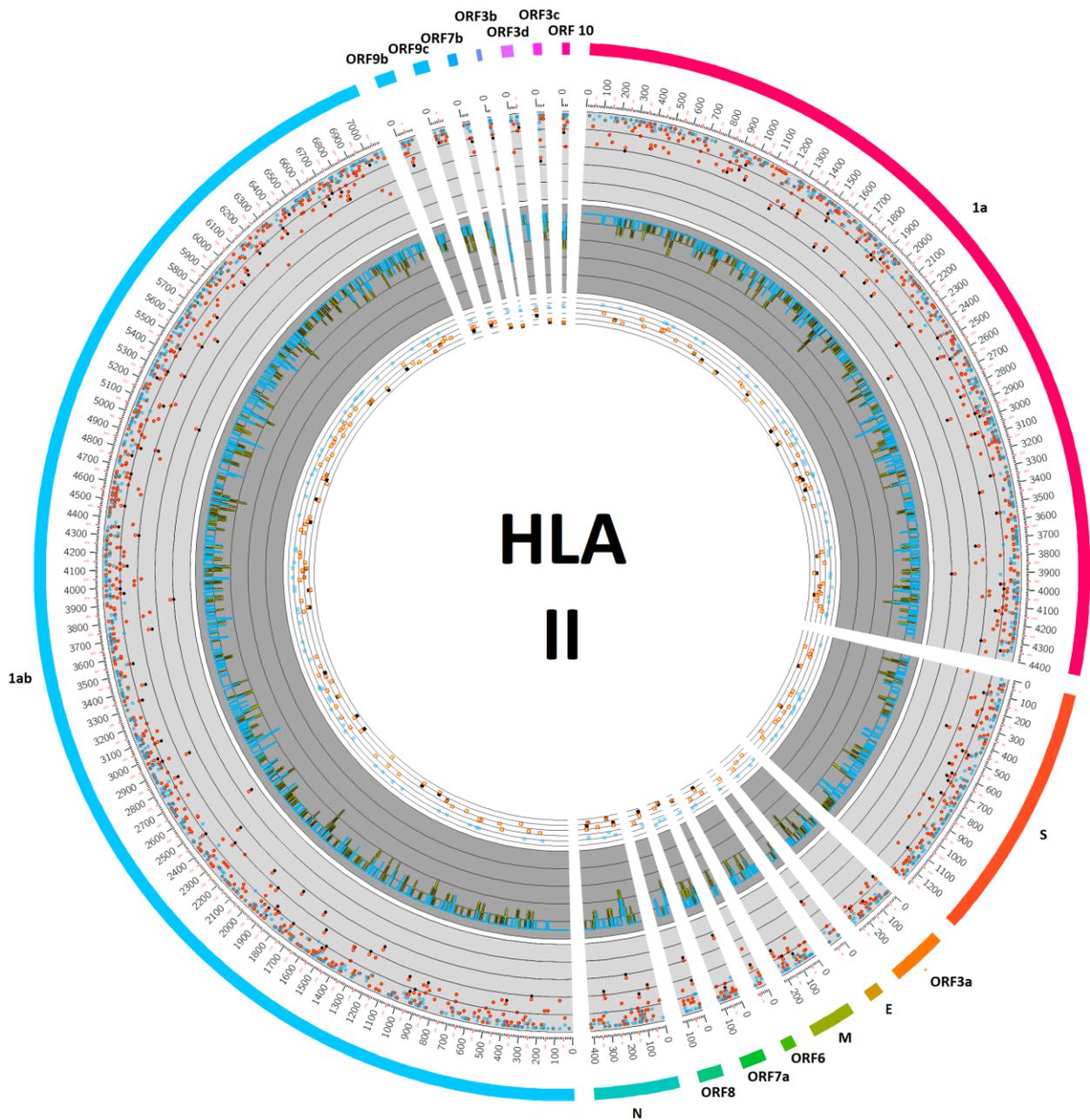
El ORF3a contiene la mayor cantidad de epítomos de células T CD8+ de EpiTex de las proteínas accesorias con 47 epítomos. Sin embargo, solo posee 1 *hotspot* para la población de Chile, la que contendría 3 epítomos de células T CD8+ en su interior.

Proteína	Población	Secuencia	Posición residuos		Cobertura Poblacional	Alelos HLA	N° de epitopos				
			Inicio	Fin			Predivac	IEDB	Total IEDB		
ORF1ab	Chile	WYTERSEKVELQTPFEIKLAKK	240	262	92,07%	39	6	1	81		
		YCPACHNSEVGEHSLAEYHNS	369	391	97,60%	38	6	0			
		ESVQTFKLVNKFILA	673	687	90,85%	27	5	2			
		ALAPNMVMVNTNFTLKGGAAPT	801	823	98,25%	44	5	2			
		<b>ITFELDERIDKLVNEKCSAYTVELGTEVNEFACVADAVIKT</b>	<b>841</b>	<b>882</b>	<b>99,54%</b>	<b>50</b>	<b>10</b>	<b>6</b>			
		ITFLKDDAPYVGVQEGVLTA	1276	1298	91,18%	28	7	1			
		NYITTFPGQLNGYVEEAKT	1324	1344	98,28%	34	4	1			
		TSNPTTFHLDGEVITFDNKLTL	1538	1560	93,84%	27	6	4			
		<b>LRVFAEYHYHTDPSFLGRYMSALNHTK</b>	<b>1627</b>	<b>1654</b>	<b>94,49%</b>	<b>31</b>	<b>7</b>	<b>8</b>			
		PITDVFYKENSYTT	1862	1876	97,07%	43	4	2			
		RCLWSTKPVETSNSFDVLKSED	2010	2032	92,23%	38	5	3			
		VLMNSLGMPSYCTGYREGYLNST	2252	2274	98,67%	37	5	1			
		NVTIATYCTGSI	2275	2285	97,00%	41	3	0			
		AGSTFISDEVARDLSLQFKRPIN	2460	2482	92,03%	31	5	2			
		<b>KMFDAIYNTFSSTFNVPMEKLTAVATAEELAKNVSNDNV</b>	<b>2589</b>	<b>2629</b>	<b>99,97%</b>	<b>59</b>	<b>11</b>	<b>10</b>			
		RIMTWLDMVDTLSLGFKDKDCVM	3662	3684	94,56%	32	5	1			
		ATEVPANSTVLSFCFAVDAAKAYDYLA	4257	4285	98,14%	49	6	4			
		NLLKDCPAVAKHDFEK	4480	4495	90,00%	25	4	2			
		SFGLPVRKIFVDGVPVSTGYHF	4717	4741	96,35%	44	6	5			
		SAGFFPNKWKARLYDSMSYEDQDA	4893	4918	96,51%	48	6	2			
		ECAQVLESMVMC	5050	5061	99,25%	45	4	0			
		IERFVSLAIDAYPLTKHPNQEYADVHLYLQYI	5248	5282	99,97%	56	10	5			
		SYGIATVREVL	5472	5482	95,53%	28	3	0			
		RELHLSWEVGPFRPPL	5485	5500	94,97%	28	4	0			
		KFKVNSTLEQYFCTVINA	5669	5686	92,95%	39	4	0			
		POIGVREFTFRNPAWRKAVIFSPINSONAVASKILGL	5815	5852	97,90%	37	8	2			
		LQFTSLEIPRRNVAATLQAEVNTGLFKDCSKV	5909	5939	95,84%	40	7	2			
		APTHLSVDTKFTEGLCVDPDIPGPKDMTYRRLISMGMGFKMNYQ	5948	5990	100,00%	62	10	2			
		<b>LYLDAYNNMISAGFSLWVYKQFDTYNLWNTFRLQSLNVAFNV</b>	<b>6418</b>	<b>6461</b>	<b>98,93%</b>	<b>46</b>	<b>11</b>	<b>11</b>			
		INNVTYTKVDGVDLEFNKTTLPVAVFELW	6479	6510	98,72%	43	8	2			
		LTLAVPNNMR	6855	6864	92,31%	25	3	0			
		APGTAVLRQWLPTG	6877	6890	92,83%	24	3	0			
Chile Mixed	REHEHEIAWYTERSEKVELQTPFEIKLAKKDFTFNG	232	268	91,16%	40	8	1	41			
	<b>ITFELDERIDKLVNEKCSAYTVELGTEVNEFACVV</b>	<b>841</b>	<b>875</b>	<b>95,48%</b>	<b>42</b>	<b>8</b>	<b>6</b>				
	VLMNSLGMPSYCTGYREGYLNST	2252	2274	96,15%	36	5	1				
	<b>KMFDAIYNTFSSTFNVPMEKLTAVATAEELAKNVSNDNV</b>	<b>2589</b>	<b>2629</b>	<b>98,51%</b>	<b>55</b>	<b>11</b>	<b>10</b>				
	ATEVPANSTVLSFCFAVDAAKAYDYLA	4257	4285	93,90%	46	6	4				
	SAGFFPNKWKARLYDSMSYEDQDA	4893	4918	92,50%	47	6	2				
	<b>GTLMIERFVSLAIDAYPLTKHPNQEYADVHLYLQYIRKL</b>	<b>5244</b>	<b>5283</b>	<b>97,53%</b>	<b>52</b>	<b>11</b>	<b>8</b>				
	DLYDLQFTSLEIPRRNVAATLQAEVNTGLFKDCSKV	5904	5939	90,70%	41	8	2				
	APTHLSVDTKFTEGLCVDPDIPGPKDMTYRRLISMGMGFKMNYQ	5948	5990	98,72%	58	10	2				
	LYLDAYNNMISAGFSLWVYKQFDTYNLWNTFRLQSLNVAFNV	6418	6461	90,00%	44	11	5				
	ORF1a	Chile	WYTERSEKVELQTPFEIKLAKKDFTFNG	240	264	92,44%	40		6	0	2
			YCPACHNSEVGEHSLAEYHNS	369	391	97,49%	37		6	0	
ESVQTFKLVNKFILA			673	687	90,85%	27	5	0			
ALAPNMVMVNTNFTLKGGAAPT			801	823	98,07%	44	5	0			
ITFELDERIDKLVNEKCSAYTVELGTEVNEFACVADAVIKT			841	882	99,65%	51	10	0			
ITFLKDDAPYVGVQEGVLTA			1276	1298	91,18%	28	7	1			
NYITTFPGQLNGYVEEAKT			1324	1344	97,93%	33	4	0			
TTFHLDGEVITFDNKLTL			1542	1560	92,40%	28	6	0			
EGKTFVLPNDLRLVFAEYHYHTDPSFLGRYMSALNHTK			1614	1654	96,89%	34	9	0			
AANFCALLAYCNK			1707	1720	94,26%	26	3	0			
VEAVMYMGTLSEYQFKGVQVCTCGGQ			1765	1792	90,28%	29	7	0			
PITDVFYKENSYTT			1862	1876	97,62%	45	4	0			
RCLWSTKPVETSNSFDVLKSED			2010	2032	92,23%	38	5	0			
VLMNSLGMPSYCTGYREGYLNST			2252	2274	98,67%	37	5	0			
NVTIATYCTGSI			2275	2285	97,00%	41	3	0			
AGSTFISDEVARDLSLQFKRPIN			2460	2482	97,26%	35	5	0			
Chile Mixed		KMFDAIYNTFSSTFNVPMEKLTAVATAEELAKNVSNDNV	2589	2629	99,76%	58	11	0	0		
		RIMTWLDMVDTLSLGFKDKDCVM	3662	3684	93,96%	30	5	1			
		ATEVPANSTVLSFCFAVDAAKAYDYLA	4257	4285	98,14%	49	6	0			
		REHEHEIAWYTERSEKVELQTPFEIKLAKKDFTFNG	232	268	98,51%	55	11	0			
		DTVIEQVQKSVNITFELDERIDKLVNEKCSAYTVELGTEVNEFACVV	828	875	96,15%	45	10	0			
		TYLCIDGALLTKSEYKGPITDVFYKENSYTT	1844	1876	96,15%	36	5	0			
		VLMNSLGMPSYCTGYREGYLNST	2252	2274	93,90%	46	6	0			
		KMFDAIYNTFSSTFNVPMEKLTAVATAEELAKNVSNDNV	2589	2629	91,39%	41	8	0			
ORF7a		Chile	MKILFLALITLATELYHYQECVGRGTLLV	9	31	95,93%	32	5	0	8	
			PCSSGTYEGNSPFH	34	47	92,08%	39	3	1		
		<b>VKHVYQLRARSVPKLFIRQEEVQELYSPI</b>	<b>71</b>	<b>100</b>	<b>98,56%</b>	<b>44</b>	<b>6</b>	<b>7</b>			
		Chile Mixed	MKILFLALITLATELYHYQECVGRGTLLVKEPCSSGTYEGNSPFH	9	47	95,86%	59	8	1		
<b>VKHVYQLRARSVPKLFIRQEEVQELYSPI</b>			<b>71</b>	<b>100</b>	<b>93,74%</b>	<b>42</b>	<b>6</b>	<b>7</b>			
ORF8		Chile	QECSLQSCIQHQPPYVDDPCPHFYSKWY	18	46	99,19%	45	7	3	8	
		Chile Mixed	IQYIDIGNYVSLPFTINCQEPKLGSLVRCFSFYEDFLEYHDDRVLV	66	118	98,88%	57	13	8		
ORF10		Chile	MGYINVFAPFTYISLLCRMNSRNIAQDVVNF	0	35	100,00%	68	10	5	5	
	Chile Mixed	MGYINVFAPFTYISLLCRMNSRNIAQDVVNFNL	0	37	99,15%	64	10	5			
ORF6	Chile	MFHLVDFQVIAEILLIMRTFKVSIWNLDIYIILIKLSKLTENKY	0	49	99,98%	61	12	10	10		
	Chile Mixed	LIMRTFKVSIWNLDIYIILIKLSKLTENKYS	16	49	97,98%	47	9	7			
ORF3a	Chile	FVRATATIPIQASLPFGWLIV	28	48	98,94%	42	5	3	3		

**Tabla 5** Desglose de clústeres de epitopos de células T CD8+ inmunodominantes de cobertura poblacional superior a 90% (*hotspots* de células T CD8+) predichos por Predivac-3.0 en proteínas codificadas de ORF de SARS-CoV-2 para la población de Chile y Chile Mixed. Se destacan en naranja los *hotspots* de células T CD8+ que contienen más de 5 epitopos de EpiTex (columna IEDB) en su interior y en verde se destacan los *hotspots* de células T CD8+ que contienen todos los epitopos de células T CD8+ de EpiTex en la respectiva proteína (bajo el criterio de que estos epitopos no tengan ningún residuo fuera de la secuencia del *hotspot*). Columna “Alelos HLA” corresponde a la cantidad de restricciones HLA asociados a los epitopos de células T CD8+ ubicados dentro del *hotspot*, de acuerdo con los resultados de Predivac-3.0. Columna “Secuencia” corresponde a la secuencia aminoacídica de los *hotspots*. No se encontraron *hotspots* para población de Chile Amerindian.

### 3.3.2. Predicción de epítomos y *hotspots* de células T CD4+

La figura 21 presenta en esquema circular el mapeo de epítomos de células T CD4+ con información sobre posición, densidad y clústeres de epítomos de células T CD4+ predichos por Predivac-3.0 para la población de Chile, Chile Mixed y Chile Amerindian. Al igual que con los epítomos de células T CD8+, se incluyeron clústeres de epítomos de células T CD4+ cuya cobertura poblacional es mayor a 30% debido a la baja cobertura observada en la población de Chile Amerindian, a modo de tener una perspectiva más completa de la situación con respecto a esa población. De la misma forma que con los epítomos de células T CD8+, ningún clúster de epítomos de células T CD4+ superó el criterio de tener una cobertura poblacional mayor al 90%, indicando ausencia absoluta a lo largo del proteoma de SARS-CoV-2 de *hotspots* para la población de Chile Amerindian. Chile Mixed posee un perfil muy similar a la población de Chile, en lo que respecta a la posición de epítomos y densidad de epítomos de células T CD4+. Esto se infiere a través del anillo de la figura 21 que aborda la posición de los epítomos y la cantidad de restricciones HLA asociadas a esos epítomos (anillo 1), donde se observa que epítomos ubicados a la misma altura para Chile y Chile Mixed poseen una cantidad de restricciones HLA muy similar. Lo mismo se observa en el anillo que aborda la densidad de los epítomos de células T CD8+. Es en los *hotspots* de epítomos donde se observa mayor diferencia entre Chile y Chile Mixed.



**Figura 21** Visualización circular de mapeo de epítomos de células T CD4+ en el proteoma de referencia de SARS-CoV-2 (Uniprot ID: UP000464024) para Chile y grupos étnicos “Mixed” y “Amerindian”, realizado con Circos. **Anillo 1**, posición de epítomos CD4+ para Chile puntos negros), grupo étnico “Mixed” (puntos amarillos con bordes rojos) y grupo étnico “Amerindian” respectivamente (puntos celestes), **anillo 2**, corresponden a densidad de epítomos CD4+ para Chile (línea negra), grupo étnico “Mixed” (línea amarilla) y grupo étnico “Amerindian” (línea celeste), **anillo 3**, clústeres de epítomos de células T CD4+ de alta cobertura poblacional para Chile (cuadrados negros), grupo étnico “Mixed” (cuadrados amarillos de bordes rojos) y grupo étnico “Amerindian” (cuadrados celestes).

En la tabla 6 se desglosa la información de los *hotspots* de epítomos de células T CD4+ presentes en proteínas estructurales y codificadas por ORFs de SARS-CoV-2, donde se muestra la secuencia, posición y cobertura de los *hotspots*, la cantidad de restricciones HLA asociado a los epítomos de células T CD4+ dentro del *hotspot* predichos por Predivac-3.0. También se presenta la cantidad de epítomos de células T CD4+ de Predivac-3.0 y EpiTex ubicados dentro del *hotspot* para las poblaciones de Chile y Chile Mixed, ya que Chile Amerindian no registra *hotspots* a lo largo de las proteínas codificadas por ORF de SARS-CoV-2. Al igual que lo observado en 19 y 20, la información con respecto a epítomos de células T CD4+ es considerablemente menor que la información de epítomos de células T CD8+. 9 es la cantidad total de *hotspots* de epítomos CD4+ para las proteínas de SARS-CoV-2. La población de Chile presenta 1 *hotspot* para ORF1a, N, ORF10, ORF3b, ORF3c y ORF7b, mientras que para Chile Mixed habría 1 *hotspot* para ORF10, ORF3b y ORF7a. Ninguno de estos *hotspots* tiene contiene epítomos de células T CD4+ de EpiTex.

Proteína	Población	Secuencia	Posición residuos		Cobertura		N° de epítomos		
			Inicio	Fin	Poblacional	Alelos HLA	Predivac	IEDB	Total IEDB
1a	Chile	KTYERHLSLHFVNLDNLRANNTKGSL	2515	2536	90,92%	25	4	0	0
N	Chile	SPRWYFYLTGPEAGLPYGANKDGIWVATEGALNTPKDH	105	145	92,35%	22	9	0	0
ORF10	Chile	INVFAFPFTIYSLLLCRMNSRNYIAQ	4	29	98,04%	29	9	0	0
	Chile Mixed	MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVN	0	34	96,23%	28	9	0	0
ORF3b	Chile	PTIFFAGILVTTIVYLTIV	3	22	98,04%	29	6	0	0
	Chile Mixed	PTIFFAGILVTTIVYLTIV	3	22	96,23%	28	6	0	0
ORF3c	Chile	ILFALLQRYRYKPHS	6	19	92,44%	25	4	0	0
ORF7a	Chile Mixed	KIILFLALITLATCELYHYQECVRGTTVLLKE	2	33	93,14%	25	6	0	0
ORF7b	Chile	VLIIMLIWFVSELEQDHNETCHA	21	43	92,97%	26	5	0	0

**Tabla 6** Desglose de clústeres de epítomos de células T CD4+ inmunodominantes de cobertura poblacional superior a 90% (*hotspots* de células T CD8+) predichos por Predivac-3.0 en proteínas estructurales y codificadas de ORF de SARS-CoV-2 para la población de Chile y Chile Mixed. Ningún epítomo de célula T CD4+ de EpiTex se encontró dentro de los *hotspots* (columna IEDB) en su interior. Columna "Alelos HLA" corresponde a la cantidad de restricciones HLA asociados a los epítomos de células T CD4+ ubicados dentro del *hotspot*, de acuerdo con los resultados de Predivac-3.0. Columna "Secuencia" corresponde a la secuencia aminoacídica de los *hotspots*. No se encontraron *hotspots* para población de Chile Amerindian.

## 4. DISCUSIÓN

### 4.1. Variaciones entre predicciones de epítomos de células T CD4+ y CD8+

Montar una respuesta acelerada a la reinfección es distintivo de la memoria inmunológica y la base de la vacunación. Esta habilidad está mediada por la presencia de células T de memoria específicas de antígeno, las que son capaces de responder más rápido y en una mayor magnitud que linfocitos T vírgenes. La respuesta primaria de linfocitos T vírgenes a inmunógenos no inflamatorios, requiere de la respuesta celular CD4+ (Shedlock & Shen, 2003). En adición, las células T CD8+ son capaces de proveer protección contra infecciones secundarias. Por lo tanto, la inducción combinada de células T CD8+ con anticuerpos pueden proveer una inmunidad protectora óptima (Schmidt & Varga, 2018). Es fundamental considerar los epítomos de células T CD4+ y CD8+ para el desarrollo racional de vacunas que gatillen respuestas protectoras robustas. Sin embargo, de la diferencia observada en los registros para los epítomos de células T CD4+ y CD8+, se infiere que los esfuerzos de investigación se canalizan principalmente en la búsqueda de epítomos de células CD8+. Ejemplo de esto, de las 1.549 restricciones HLA asociados a epítomos de células T de EpiTex, 1.469 corresponderían a HLA clase I (repartido en 35 alelos) mientras que para HLA clase II solo se distinguen 102 restricciones (repartido en 10 alelos). En las figuras 18 y 19 junto con las figuras 20 y 21, se observó que la diferencia de información para epítomos de células T CD4+ y CD8+ también está presente en los resultados de herramientas inmunoinformáticas predictivas (Predivac-3.0, en este caso). Esto último es razonable si se considera que la cantidad de alelos descubiertos para HLA clase I es considerablemente mayor que los descubiertos para HLA clase II, con 24.703 alelos hasta la fecha para HLA clase I y 9.719 alelos para HLA clase II, siendo información de las que se alimentan programas como Predivac-3.0, por tanto, es más probable encontrar información de epítomos de células T CD8+ que de epítomos de células T CD4+.

La respuesta primaria de linfocitos T CD8+ a los agentes infecciosos a menudo es independiente de la respuesta CD4+ (Shedlock & Shen, 2003), esto puede ser factor en la diferencia en la cantidad de epítomos de células T CD4+ y CD8 observada de EpiTex, habiendo mayores descubrimientos de epítomos de células T CD8+ que CD4+. En SARS-CoV-2 aún existe mucho espacio disponible para la investigación y descubrimiento de epítomos de células T CD4+, ya que solo se registran epítomos para las proteínas estructurales S, N y M. Si se considera esto junto con la densidad de epítomos de células T CD8+, es posible que la mayoría de los experimentos realizados se hayan enfocado en las proteínas estructurales de SARS-CoV-2.

## 4.2. Predivac-3.0 y componente étnico en Chile

En Chile los grupos étnicos que poseen información de frecuencias de HLA son los grupos étnicos amerindios repartidos en poblaciones de Isla de Pascua, Huilliches y Mapuches; caucasoides en poblaciones de Santiago; hispánicos en poblaciones de Santiago; y mixtos repartidos en poblaciones de Santiago (Thorsby et al., 2009) (Rey et al., 2013) (Single et al., 2020). Cada grupo étnico presente en Chile tiene asociados alelos HLA característicos (tabla 7).

Population	Ethnicity	Study	HLA Loci
Chile Easter Island	Amerindian	HLA	A B C DRB1
Chile Huilliche	Amerindian	HLA	DRB1
Chile Huilliche NA-DHS_14 (G)	Amerindian	HLA	A B C DRB1
Chile Mapuche	Amerindian	HLA	A B DQB1 DRB1
Chile Santiago	Caucasoid	HLA	A B DRB1
Chile Santiago Mixed	Mixed	HLA	A B C DQB1 DRB1
Chile Santiago pop 2	Hispanic	HLA	DQA1 DQB1

**Tabla 7** Grupos étnicos registrados en poblaciones de Chile con loci HLA específico de la etnia. Información presente en “Allele Frequency Net Database” (AFND) (<http://www.allelefreuencies.net/>).

Predivac-3.0 implementa una aproximación pan-específica basado en el concepto de residuos determinantes de especificidad (SDR), utilizando datos de unión de alta afinidad asumiendo que es el sesgo positivo hacia las características peptídicas subyacentes que se correlacionan con la promiscuidad y la inmunodominancia. Este método asume que las células T con la especificidad requerida, estarán presente en el repertorio de células T en la base de datos que utiliza el programa. Que los grupos étnicos que distinga en Chile sean solamente Amerindian y Mixed y que el perfil de inmunidad de la población chilena sea similar al perfil de la población Mixed puede deberse a la ausencia de información sobre Chile en el programa. En la tabla 8 se muestra el resumen del número de alelos para los cuales AFND posee información de distribución de frecuencias en poblaciones asociadas a diferentes regiones geográficas, donde el registro de Sudamérica junto con Centroamérica es comparativamente inferior a Norteamérica, Europa, y levemente superior el Sureste asiático. Este podría ser un motivo de la falta de información para la población de Chile.

De acuerdo con AFND, la población de Chile Mixed contiene información muestral de obtenida en Santiago de Chile, mientras que los grupos étnicos amerindios se reparten en poblaciones de Isla de Pascua, Talca, Huilliche y Mapuche (tabla 9). Por lo tanto, surge la pregunta si la densidad demográfica es un factor que afecta a los resultados de Predivac-3.0, explicando así la relación que hay entre la población de Chile con Chile Mixed.

REGIÓN	HLA
Europa	281
América del Norte	205
América Central y del Sur	199
Sureste de Asia	194
Noreste de Asia	98
Asia Occidental	88
África subsahariana	66
Oceania	64
Asia del Sur	62
África del Norte	29
Australia	8
Asia Central	0
<b>Total</b>	<b>1294</b>

**Tabla 8** Alelos HLA por continentes de acuerdo con AFND (<http://www.allelefreqs.net/pop6001a.asp>).

**A)**

 Chile (1)			
Population	Region	Study	Sample Size
Chile Santiago Mixed	South and Central America	HLA ■	70

**B)**

 Chile (5)			
Population	Region	Study	Sample Size
Chile Easter Island	South and Central America	HLA ■	21
Chile Huilliche	South and Central America	HLA ■	40
Chile Huilliche NA-DHS_14 (G)	South and Central America	HLA ■	20
Chile Mapuche	South and Central America	HLA ■	66
Chile Talca KIR	South and Central America	KIR ■	90

**Tabla 9** Poblaciones con componente étnico presentes en Chile de acuerdo con AFND. A) Poblaciones cuyo componente étnico pertenece al grupo étnico Mixed. B) Poblaciones cuyo componente étnico abarcan al grupo étnico Amerindian.

La falta de información implica potenciales sesgos en la matriz de puntaje que implementa Predivac-3.0 para asignar un puntaje relacionado con la probabilidad de unión a proteínas HLA. Con respecto a la ausencia de *hotspots* de la población Chile Amerindian, es posible que la diferencia visibilizada en la tabla 4 con respecto a la frecuencia de alelos HLA entre los grupos Chile Easter Island y Chile Mapuche, indique que el perfil de alelos HLA tenga una variación significativa entre poblaciones en Chile asociados con el grupo étnico Amerindian. Aun así, cabe preguntarse sobre si se requiere más información sobre este grupo étnico en Predivac-3.0 para determinar *hotspots* de este grupo étnico en Chile, o si este grupo posee un perfil de inmunidad que los hace más vulnerables frente a la amenaza de SARS-CoV-2. De ser esto último, los resultados obtenidos en esta investigación se obtuvieron con el proteoma de referencia obtenido de Uniprot, por lo que aún quedaría por investigar resultados a partir de otras fuentes de variantes de SARS-CoV-2.

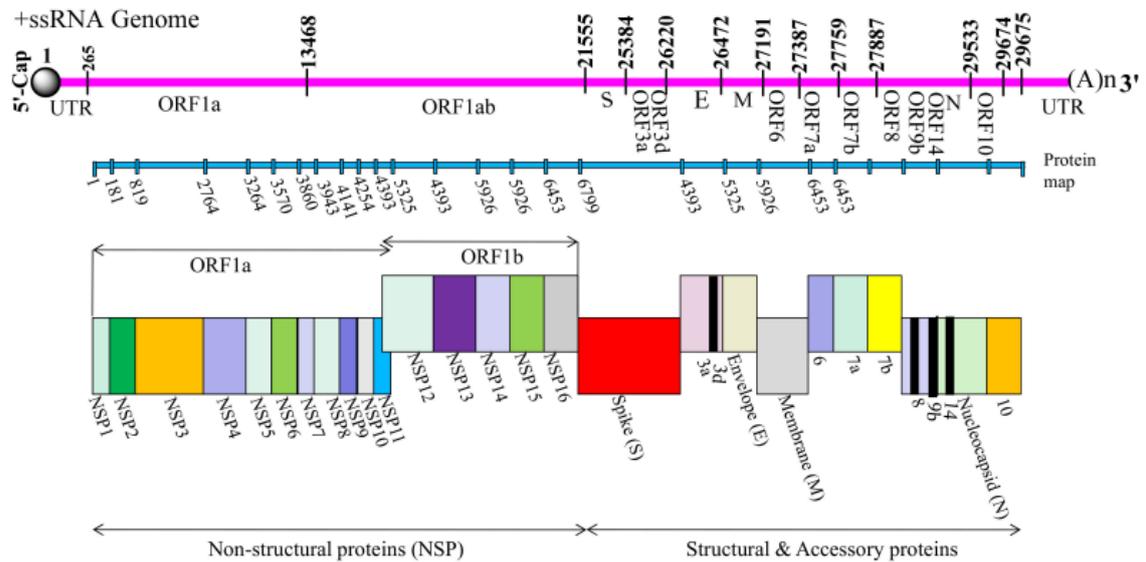
### **4.3. Hotspots**

Es común referirse a clústeres de epítomos de células T como regiones en el proteoma de un patógeno con una alta concentración de epítomos de células T. El concepto de *hotspot* de epítomos incorpora un criterio poblacional relacionado con la cobertura de la población. Oyarzún et al., 2021 define a los *hotspots* como clústeres de epítomos de células T CD4+ y CD8+ asociados con una cobertura poblacional superior al 80%. En esta investigación, este criterio de cobertura poblacional fue modificado para aquellos clústeres de epítomos cuya cobertura fuera superior al 90% de la población objetivo. De todas formas, se adjuntan los clústeres de epítomos de células T CD4+ y CD8+ identificados para las poblaciones de Chile, Chile Mixed, Chile Amerindian.

#### **4.3.1. Marcos de lectura abierto (ORF)**

El genoma de SARS-CoV-2 presenta los ORF1a y ORF1ab que codifican dieciséis proteínas no estructurales (NSP1-NSP16) Con excepción de ORF1ab y ORF1a, todas las ORF forman parte de proteínas estructurales. Las proteínas accesorias codificadas de ORF3a, ORF3d, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF14 y ORF10, son proteínas codificadas por genes estructurales, que, junto con las proteínas estructurales y no estructurales, son las responsables de los aspectos funcionales y estructurales de SARS-CoV-2 (figura 22). Cabe destacar que ORF14 está ausente

en el proteoma de referencia de SARS-CoV-2 presente en Uniprot, utilizado en esta investigación.



**Figura 22** Esquema de arreglo genómico de SARS-CoV-2 extraído de Yadav et al., 2021

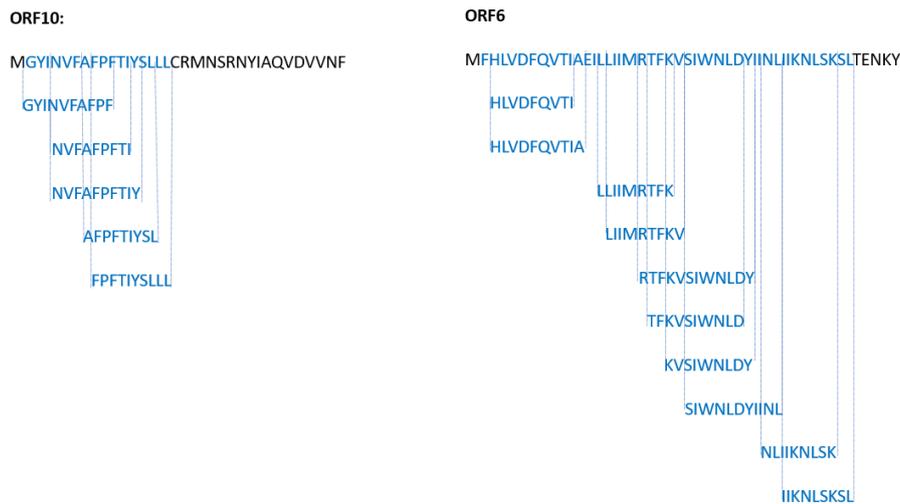
La tabla 10 contiene todos los *hotspots* identificados que cumplieron con el criterio de poseer en el interior de su secuencia más de 5 epítomos de células T CD8+ presentes en la base de datos EpiTex. Todas las proteínas estructurales de SARS-CoV-2 poseen *hotspots* que cumplen con este criterio, aunque son los *hotspots* detectados para ORF, específicamente ORF10 y ORF6, los que llaman la atención ya que contienen todos los epítomos de células T de EpiTex en el interior de sus *hotspots*. La proteína ORF10 es una supuesta proteína viral de 38 aminoácidos con una región  $\alpha$ -hélice (Gordon et al., 2020), altamente ordenada, hidrofóbica y térmicamente estable, que contiene al menos una región transmembrana (Schuster, 2021); interactúa con múltiples proteínas humanas, pero no es esencial para la infección y replicación viral (Hassan et al., 2022).

El *hotspot* descubierto en la región del ORF10 (tabla 6) tiene el potencial de cubrir en un 100% la población en Chile, pero adicionalmente tiene un alto nivel de validación al encontrarse presente los 5 epítomos predichos en la base de datos EpiTex (figura 23), tanto para la población de Chile como para la población de Chile Mixed. ORF10 junto con ORF3b, son expresados exclusivamente por SARS-CoV-2, y tienen una homología detectable limitada con las proteínas de SARS-CoV (Gordon et al., 2020), por lo que considerar este *hotspot* en el diseño racional de una vacuna basada en epítomos, implicaría una estrategia para desarrollar inmunidad contra SARS-CoV-2, pero no aseguraría protección para otros coronavirus.

La proteína codificada por ORF6 posee una función conservada en la mejora de la replicación y crecimiento viral, y se caracteriza por poseer una  $\alpha$ -hélice que se incrusta en la membrana en su extremo N-terminal (Lee et al., 2021). Presenta *hotspots* para la población de Chile y el grupo étnico Mixed que superan el 97% de la población. El *hotspot* para la población de Chile posee la totalidad de los epítomos descubiertos experimentalmente según la base de datos de EpiTex. También posee un alto nivel de validación al encontrarse presente los 10 epítomos predichos en la base de datos EpiTex (figura 23). Cabe destacar que ORF6 también presenta un *hotspot* para Chile Mixed, que, a pesar de no poseer todos los epítomos de EpiTex, si comparte posición con el *hotspot* para la población de Chile.

Proteína	Población	Secuencia	Posición residuos		Cobertura Poblacional	Alelos HLA	N° de epítomos	
			Inicio	Fin			Predivac	IEDB
ORF1ab	Chile	ITFELDERIDKVLNEKCSAYVELGTEVNEFACVADAVIKT	841	882	99,54%	50	10	6
		LRVEAEFYHTTDFPSFLGRYMSALNHTK	1627	1654	94,49%	31	7	8
		KMFDAYVNTFSSFNVPMEKLTAVATAEAEAKNVSLDNDV	2589	2629	99,97%	59	11	10
		LYLDAYNMMSAGFSWVYKQFDYTNLWNTFTRLQSLNVAFNV	6418	6461	98,93%	46	11	11
		ITFELDERIDKVLNEKCSAYVELGTEVNEFACVV	841	875	95,48%	42	8	6
		KMFDAYVNTFSSFNVPMEKLTAVATAEAEAKNVSLDNDV	2589	2629	98,51%	55	11	10
		GTLMIERVSLAIDAYPLTKHPNQEYADVHFLYLQYIRKL	5244	5283	97,53%	52	11	8
		ITFELDERIDKVLNEKCSAYVELGTEVNEFACVV	841	875	95,48%	42	8	6
		KMFDAYVNTFSSFNVPMEKLTAVATAEAEAKNVSLDNDV	2589	2629	98,51%	55	11	10
		GTLMIERVSLAIDAYPLTKHPNQEYADVHFLYLQYIRKL	5244	5283	97,53%	52	11	8
ORF7a	Chile	VKHVYQLRARSVSPKIFIRQEEVQELYSPI	71	100	98,56%	44	6	7
	Chile Mixed	VKHVYQLRARSVSPKIFIRQEEVQELYSPI	71	100	93,74%	42	6	7
ORF8	Chile Mixed	GSKSPIQIDIGNYTVSCLPFTINCQEPKGLSLVRCSEYDFLEVHDRVVLL	66	118	98,88%	57	13	8
ORF10	Chile	MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNF	0	35	100,00%	68	10	5
	Chile Mixed	MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNF	0	37	99,15%	64	10	5
ORF6	Chile	MFHLDFQVTAIEILLIRTFKVSINWLDYIINLIKNSKLTENKY	0	49	99,98%	61	12	10
	Chile Mixed	LIIMRTFKVSINWLDYIINLIKNSKLTENKYS	16	49	97,98%	47	9	7
S	Chile	KGYHLMSPQSAHPGVVFLHVTY	1045	1067	96,81%	35	5	8
	Chile Mixed	HSTQDLFPFFSNVTWFHAIHVSNGTKRFDNVPVLPFNDGVYFA	49	93	97,66%	44	9	11
	Chile Mixed	AYVYGLQPTFLKYNENGTITDAVDCALDPLSETK	264	307	92,03%	42	10	7
N	Chile	AIKLDKDPNFKDQVILLNKHIDAYKTFPPT	336	366	97,82%	43	9	6
	Chile Mixed	GMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPT	316	366	98,57%	56	13	13
M	Chile	LTRPLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATS	129	173	99,84%	54	12	10
	Chile Mixed	ESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLG	135	181	92,84%	50	12	13
E	Chile Mixed	LLVTAILTALRCAYCCINNVSLVKPFSYVSRVKNLNS	27	68	97,63%	47	12	7

**Tabla 10** Selección de *hotspots* predichos por Predivac-3.0 para la población chilena, considerando características de densidad (clústeres), cobertura poblacional y coincidencia con epítomos descubiertos experimentalmente (EpiTex). En verde se destacan los *hotspots* que abarcan la totalidad de epítomos putativos en EpiTex.



**Figura 23** Hotspots de epítomos de células T CD8+ de ORF10 y ORF6 de SARS-CoV-2 para la población de Chile. Secuencias más largas corresponden a los hotspots, donde en azul se destaca la zona que abarcan los epítomos de células T CD8+ de EpiTex. Debajo de los hotspots se muestran los respectivos epítomos de células T CD8+ ubicados en el interior de los hotspots.

#### 4.3.2. Variabilidad genética de proteínas de SARS-CoV-2

La variabilidad genética es un factor importante para considerar frente a la selección de los hotspots a utilizar para el diseño racional de una vacuna basada en epítomos. GISAID (<https://www.gisaid.org/>) es una iniciativa que promueve el intercambio rápido de datos de todos los virus de influenza y coronavirus causantes de COVID-19. Incluye secuencias genéticas y datos clínicos y epidemiológicos relacionados con virus humanos, información de especies específicas y su localización asociadas con virus aviáres y otros animales. Un análisis realizado en 329.942 secuencias de SARS-CoV-2 presente en esta base de datos, estudió la variabilidad genética de proteínas de SARS-CoV-2 (Tabla 11).

Conservation of viral genes.			
Viral gene	Highly conservative, %	Moderately conservative, %	Low conservative, %
ORF1ab	NA	NA	NA
S	3.15	81.02	10.91
ORF3a	52.78	43.22	0.72
E	94.66	4.62	0.1
M	62.4	36.36	0.42
ORF6	96.15	3.13	0.25
ORF7a	83.43	7.12	0.62
ORF7b	85.48	5.47	0.4
ORF8	64.71	33.09	0.38
N	19.62	76.35	0.72
ORF10	73.87	23.62	0.16

**Tabla 11** Estudio de variabilidad genética de proteínas de SARS-CoV-2 mediante el análisis de 329.942 secuencias presentes en GISAID extraído de Zelenova et al., 2021.

De las principales proteínas estructurales, S y N poseen una baja conservación de su secuencia, seguido de M y E, este último con un alto porcentaje de conservación (Zelenova et al., 2021). Esto último es interesante puesto que E registra un *hotspot* para la población de Chile Mixed con una cobertura de 97,63%, lo que lo hace un candidato interesante junto con ORF6 y ORF10 para su uso en una vacuna basada en epítomos para SARS-CoV-2 con resultados prometedores para la población de Chile. Con respecto a las proteínas S y N, queda por averiguar sobre si las regiones de la secuencia que presentan una considerable variabilidad genética corresponden a zonas donde se encuentran ubicados los *hotspots* detectados.

Aunque la cultura, comportamiento y diferencias socioeconómicas entre grupos étnicos puede influenciar sobre la dispersión y reproducción del virus, existe otro factor a considerar consistente en los alelos de HLA, los que pueden conservarse entre individuos del mismo grupo y ser significativamente distintos entre grupos étnicos (Cao et al., 2001). Este problema se complica aún más por el hecho de que los grupos minoritarios raciales y étnicos generalmente están subrepresentados en los ensayos clínicos, limitando la capacidad de probar la eficacia y seguridad de las nuevas intervenciones clínicas en diversas poblaciones, conduciendo a la falta de datos de células T en etnias en las que se propaga una epidemia viral. Por lo tanto, una consideración cuidadosa de los antecedentes genéticos de HLA es primordial para garantizar la efectividad, sostenibilidad y cobertura étnicamente imparcial de la población durante el desarrollo de una vacuna, especialmente considerando las variaciones en las respuestas de las células T en múltiples etnias (Oyarzún et al., 2021).

Aunque la evidencia sugiere que la respuesta inmunitaria adaptativa provocada por infección por SARS-CoV-2 está presente y podrían proteger contra la reinfección, la experiencia con coronavirus estacionales y la experiencia actual con SARS-CoV-2 sugiere que la inmunidad a la infección natural podría disminuir con el tiempo e incluso de reportarse casos de reinfección (Kim et al., 2020). Es fundamental que las investigaciones se centren en comprender las correlaciones entre la infección y la inmunidad humoral y celular inducida por vacunas, definiendo objetivos detallados de respuestas de inmunidad humoral y celular a nivel de epítomo (Poland et al., 2020).

## 5. CONCLUSIÓN

Esta investigación permitió realizar un análisis exploratorio del proteoma del virus SARS-CoV-2 con el propósito de identificar epítomos de células T que tengan potencialmente un mayor interés para la población chilena. Para esto, se utilizó herramientas inmunoinformáticas para identificar epítomos y *hotspots* de células T, los cuales fueron validados mediante información internacional. En base a los resultados obtenidos de esta investigación, es posible concluir lo siguiente:

- Los alelos HLA presentes en la base de datos EpiTex corresponden mayoritariamente a alelos prevalentes a nivel global. Teniendo esto en consideración, se considera que en la mayoría de los experimentos enfocados en descubrir nuevos epítomos puede existir un sesgo hacia alelos HLA de mayor prevalencia. De ser este el caso, las regiones geográficas que posean estos alelos HLA con una baja frecuencia podrían verse menos favorecidos en el proceso de desarrollo de vacunas.
- La diferencia de información presente para epítomos de células T CD8+ y CD4+ indica que la mayoría de los esfuerzos experimentales están enfocados en encontrar epítomos para células T CD8+. La característica de la respuesta inmune mediada por células T CD8+ es considerada crítica frente a infecciones virales, además de que puede ser independiente de la respuesta mediada por células T CD4+, siendo motivo que explicaría el mayor interés que generan en la comunidad científica. SARS-CoV-2 posee un campo amplio de investigación para el descubrimiento de epítomos de células T CD4+.
- La ausencia de *hotspots* para la población de Chile Amerindian puede deberse a diferencias significativas entre las frecuencias de alelos HLA de poblaciones que comparten este componente étnico y falta de información en Predivac-3.0, pero no puede descartarse que, de acuerdo con esta investigación, esta población efectivamente posea factores de susceptibilidad a infección de SARS-CoV-2. Se requiere mayor investigación con respecto a variantes de este virus en la población de Chile.
- Las poblaciones de Chile y Chile Mixed comparten similitud en resultados obtenidos para posición, densidad y *hotspots* de epítomos de células T. Dado que la población Chile Mixed proviene de muestras realizadas en Santiago, cabe la posibilidad de que el factor demográfico sea un motivo de esta situación. Sin embargo, aún queda mucho campo que descubrir con respecto a la investigación de alelos HLA para la población de Chile.

- Los ORF10, ORF6 y E revelan ser candidatos de alto potencial para ser seleccionados en posterior investigación experimental para el desarrollo de una vacuna especialmente diseñada contra SARS-CoV-2 para la población de Chile. Al respecto, su alta conservación de secuencia en comparación a las proteínas estructurales S, M y N, da cuenta de una menor tasa de mutación y posiblemente menor capacidad de escape por parte del virus a una respuesta inmunitaria celular. La proteína ORF10 al ser exclusiva de SARS-CoV-2, puede ser un candidato ventajoso para la protección contra este virus.

## 6. BIBLIOGRAFÍA

Abbas, A. K., Lichtman, A. H., & Pillai, S. (2022). *Inmunología celular y molecular*. Elsevier Health Sciences.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., ... & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, *46*(W1), W537-W544.

Afrough, B., Dowall, S., & Hewson, R. (2019). Emerging viruses and current strategies for vaccine intervention. *Clinical & Experimental Immunology*, *196*(2), 157-166.

Aguilera, X., Mundt, A. P., Araos, R., & Weitzel, T. (2021). The story behind Chile's rapid rollout of COVID-19 vaccination. *Travel Medicine and Infectious Disease*.

Alarcon, J. B., Waine, G. W., & McManus, D. P. (1999). DNA vaccines: technology and application as anti-parasite and anti-microbial agents. *Advances in parasitology*, *42*, 343-410.

Alicia, sanchez-mazas. (2020). A review of HLA allele and SNP associations with highly prevalent infectious diseases in human populations. *Swiss Medical Weekly*, *150*. <https://doi.org/10.4414/smw.2020.20214>

Amanat, F., & Krammer, F. (2020). SARS-CoV-2 Vaccines: Status Report. *Immunity*, *52*(4), 583-589. <https://doi.org/10.1016/j.immuni.2020.03.007>

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, *26*(4), 450-452.

Artaud, C., Kara, L., & Launay, O. (2019). Vaccine Development: From Preclinical Studies to Phase 1/2 Clinical Trials. *Malaria Control and Elimination*, *165–176*. doi:10.1007/978-1-4939-9550-9\_12

Bart, K. J., Foulds, J., & Patriarca, P. (1996). Global eradication of poliomyelitis: benefit-cost analysis. *Bulletin of the World Health Organization*, *74*(1), 35.

Bobbala, S., & Hook, S. (2016). Is there an optimal formulation and delivery strategy for subunit vaccines?. *Pharmaceutical research*, *33*(9), 2078-20

- Bonaldo, M. C., Sequeira, P. C., & Galler, R. (2014). The yellow fever 17D virus as a platform for new live attenuated vaccines. *Human vaccines & immunotherapeutics*, *10*(5), 1256-1265
- Bonham, V. L., Green, E. D., & Pérez-Stable, E. J. (2018). Examining how race, ethnicity, and ancestry data are used in biomedical research. *Jama*, *320*(15), 1533-1534.
- Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC bioinformatics*, *7*(1), 1-5.
- Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., & Fernández-Viña, M. A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human immunology*, *62*(9), 1009-1030.
- Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., & Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, *9*(1), 221-236. <https://doi.org/10.1080/22221751.2020.1719902>
- Covián, C., Retamal-Díaz, A., Bueno, S. M., & Kalergis, A. M. (2020). Could BCG Vaccination Induce Protective Trained Immunity for SARS-CoV-2? *Frontiers in Immunology*, *11*. <https://doi.org/10.3389/fimmu.2020.00970>
- Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, *17*(3), 181-192. <https://doi.org/10.1038/s41579-018-0118-9>
- DeMetsD, Friedman L, Furberg C (2010). *Fundamentals of Clinical Trials* (4th ed.). Springer. ISBN 978-1-4419-1585-6
- Dendrou, C. A., Petersen, J., Rossjohn, J., & Fugger, L. (2018). HLA variation and disease. *Nature Reviews Immunology*, *18*(5), 325-339. <https://doi.org/10.1038/nri.2017.143>
- Doherty, M., Buchy, P., Standaert, B., Giaquinto, C., & Prado-Cohrs, D. (2016). Vaccine impact: Benefits for human health. *Vaccine*, *34*(52), 6707-6714. <https://doi.org/10.1016/j.vaccine.2016.10.025>

Draper, S. J., Sack, B. K., King, C. R., Nielsen, C. M., Rayner, J. C., Higgins, M. K., Long, C. A., & Seder, R. A. (2018). Malaria Vaccines: Recent Advances and New Horizons. *Cell Host & Microbe*, 24(1), 43-56. <https://doi.org/10.1016/j.chom.2018.06.008>

Evans, M. (2008). Recent advances in immunoinformatics: Application of in silico tools to drug development. *Current opinion in drug discovery & development*, 11, 233-241.

Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., & Ladnyi, I. D. (1988). *Smallpox and its eradication* (Vol. 6). Geneva: World Health Organization.

Fischer, W., Perkins, S., Theiler, J., Bhattacharya, T., Yusim, K., Funkhouser, R., ... & Hahn, B. H. (2007). Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nature medicine*, 13(1), 100-106.

Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., & Hurley, J. H. (2021). Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National Academy of Sciences*, 118(2).

Folegatti, P. M., Ewer, K. J., Aley, P. K., Angus, B., Becker, S., Belij-Rammerstorfer, S., Bellamy, D., Bibi, S., Bittaye, M., Clutterbuck, E. A., Dold, C., Faust, S. N., Finn, A., Flaxman, A. L., Hallis, B., Heath, P., Jenkin, D., Lazarus, R., Makinson, R., ... Yau, Y. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: A preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*, 0(0). [https://doi.org/10.1016/S0140-6736\(20\)31604-4](https://doi.org/10.1016/S0140-6736(20)31604-4)

Frahm, N., Korber, B. T., Adams, C. M., Szinger, J. J., Draenert, R., Addo, M. M., Feeney, M. E., Yusim, K., Sango, K., Brown, N. V., SenGupta, D., Piechocka-Trocha, A., Simonis, T., Marincola, F. M., Wurcel, A. G., Stone, D. R., Russell, C. J., Adolf, P., Cohen, D., ... Brander, C. (2004). Consistent Cytotoxic-T-Lymphocyte Targeting of Immunodominant Regions in Human Immunodeficiency Virus across Multiple Ethnicities. *Journal of Virology*, 78(5), 2187-2200. <https://doi.org/10.1128/JVI.78.5.2187-2200.2004>

García-Sastre, A., & Mena, I. (2013). Novel vaccine strategies against emerging viruses. *Current Opinion in Virology*, 3(2), 210-216. <https://doi.org/10.1016/j.coviro.2013.02.001>

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., ... & Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 583(7816), 459-468

Griffin, D. E. (2018). Measles vaccine. *Viral immunology*, 31(2), 86-95.

Gubler, D. J. (2011). Dengue, Urbanization and Globalization: The Unholy Trinity of the 21<sup>st</sup> Century. *Tropical Medicine and Health*, 39(4SUPPLEMENT), S3-S11. <https://doi.org/10.2149/tmh.2011-S05>

Halliday, J. E. B., Hampson, K., Hanley, N., Lembo, T., Sharp, J. P., Haydon, D. T., & Cleaveland, S. (2017). Driving improvements in emerging disease surveillance through locally relevant capacity strengthening. *Science*, 357(6347), 146-148. <https://doi.org/10.1126/science.aam8332>

Hassan, S. S., Lundstrom, K., Serrano-Aroca, Á., Adadi, P., Aljabali, A. A., Redwan, E. M., ... & Uversky, V. N. (2022). Emergence of unique SARS-CoV-2 ORF10 variants and their impact on protein structure and function. *International Journal of Biological Macromolecules*, 194, 128-143.

Hertz, T., Ahmed, H., Friedrich, D. P., Casimiro, D. R., Self, S. G., Corey, L., McElrath, M. J., Buchbinder, S., Horton, H., Frahm, N., Robertson, M. N., Graham, B. S., & Gilbert, P. (2013). HIV-1 Vaccine-Induced T-Cell Responses Cluster in Epitope Hotspots that Differ from Those Induced in Natural Infection with HIV-1. *PLoS Pathogens*, 9(6). <https://doi.org/10.1371/journal.ppat.1003404>

Hsu, D. C., & O'Connell, R. J. (2017). Progress in HIV vaccine development. *Human Vaccines & Immunotherapeutics*, 13(5), 1018-1030. <https://doi.org/10.1080/21645515.2016.1276138>

James, C. A., Xu, Y., Aguilar, M. S., Jing, L., Layton, E. D., Gilleron, M., ... & Seshadri, C. (2022). CD4 and CD8 co-receptors modulate functional avidity of CD1b-restricted T cells. *Nature Communications*, 13(1), 78.

Jiang, S., He, Y., & Liu, S. (2005). SARS Vaccine Development. *Emerging Infectious Diseases*, 11(7), 1016-1020. <https://doi.org/10.3201/eid1107.050219>

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990-993. <https://doi.org/10.1038/nature06536>

Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific reports*, *10*(1), 1-24.

Kaslow, D. C., & Biernaux, S. (2015). RTS, S: toward a first landmark on the Malaria Vaccine Technology Roadmap. *Vaccine*, *33*(52), 7425-7432.

Kim, J. H., Marks, F., & Clemens, J. D. (2021). Looking beyond COVID-19 vaccine phase 3 trials. *Nature medicine*, *27*(2), 205-211.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., & Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *British medical bulletin*, *58*(1), 19-42

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, *19*(9), 1639-1645.

Laverack, G., & Manoncourt, E. (2016). Key experiences of community engagement and social mobilization in the Ebola response. *Global health promotion*, *23*(1), 79-82.

Lee, C. H., & Koohy, H. (2020). In silico identification of vaccine targets for 2019-nCoV. *F1000Research*, *9*.

Lee, J. G., Huang, W., Lee, H., van de Leemput, J., Kane, M. A., & Han, Z. (2021). Characterization of SARS-CoV-2 proteins reveals Orf6 pathogenicity, subcellular localization, host interactions and attenuation by Selinexor. *Cell & bioscience*, *11*(1), 1-12.

Li, H., Liu, S.-M., Yu, X.-H., Tang, S.-L., & Tang, C.-K. (2020). Coronavirus disease 2019 (COVID-19): Current status and future perspectives. *International Journal of Antimicrobial Agents*, *55*(5), 105951. <https://doi.org/10.1016/j.ijantimicag.2020.105951>

Li, L., & Petrovsky, N. (2016). Molecular mechanisms for enhanced DNA vaccine immunogenicity. *Expert review of vaccines*, *15*(3), 313-329.

- Liu, G., Li, D., Li, Z., Qiu, S., Li, W., Chao, C. C., ... & Li, B. (2017). PSSMHCpan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity. *Giga Science*, 6(5), gix017.
- Lubanga, B., Chemtai, A., & Kwaro, D. (2016). The RTS, S/AS malaria vaccine candidate: A status review.
- Ma, E., & Cheng, G. (2022). Host immunity and vaccine development against Dengue virus. *Infectious Medicine*.
- Mackenzie, J. S., Childs, J. E., Field, H. E., Wang, L. F., & Breed, A. C. (2016). The role of bats as reservoir hosts of emerging neuroviruses. In *Neurotropic Viral Infections* (pp. 403-454). Springer, Cham.
- María, R.-A. R., Arturo, C.-V. J., Alicia, J.-A., Paulina, M.-L. G., & Gerardo, A.-O. (2017). The Impact of Bioinformatics on Vaccine Design and Development. *Vaccines*. <https://doi.org/10.5772/intechopen.69273>
- Mathieu, E., Ritchie, H., Ortiz-Ospina, E. et al. A global database of COVID-19 vaccinations. *Nat Hum Behav* (2021)
- McKinney, W. (2010, June). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, No. 1, pp. 51-56).
- Medicine, I. of. (2003). *Microbial Threats to Health: Emergence, Detection, and Response*. <https://doi.org/10.17226/10636>
- Minor, P. D. (2015). Live attenuated vaccines: Historical successes and current challenges. *Virology*, 479, 379-392.
- Minsal: (<https://www.minsal.cl/nuevo-coronavirus-2019-ncov/casos-confirmados-en-chile-covid-19/>)
- Morgat, A., Lombardot, T., Coudert, E., Axelsen, K., Neto, T. B., Gehant, S., ... & UniProt Consortium. (2020). Enzyme annotation in UniProtKB using Rhea. *Bioinformatics*, 36(6), 1896-1901.
- Nadeem, A. Y., Shehzad, A., Islam, S. U., Al-Suhaimi, E. A., & Lee, Y. S. (2022). Mosquirix™ RTS, S/AS01 vaccine development, immunogenicity, and efficacy. *Vaccines*, 10(5), 713.

Nahid, P., Horne, D. J., Jarlsberg, L. G., Reiner, A. P., Osmond, D., Hopewell, P. C., & Bibbins-Domingo, K. (2011). Racial differences in tuberculosis infection in United States communities: the coronary artery risk development in young adults study. *Clinical infectious diseases*, 53(3), 291-294.

Nascimento, I. P., & Leite, L. (2012). Recombinant vaccines and the development of new vaccine strategies. *Brazilian journal of medical and biological research*, 45, 1102-1111.

National Institutes of Health (US); Biological Sciences Curriculum Study. NIH Curriculum Supplement Series [Internet]. Bethesda (MD): National Institutes of Health (US); 2007. Understanding Emerging and Re-emerging Infectious Diseases. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK20370/>

Ndhlovu, Z. M., Piechocka-Trocha, A., Vine, S., McMullen, A., Koofhethile, K. C., Goulder, P. J., ... & Walker, B. D. (2011). Mosaic HIV-1 Gag antigens can be processed and presented to human HIV-specific CD8+ T cells. *The Journal of Immunology*, 186(12), 6914-6924.

Nielsen, C. M., Vekemans, J., Lievens, M., Kester, K. E., Regules, J. A., & Ockenhouse, C. F. (2018). RTS,S malaria vaccine efficacy and immunogenicity during Plasmodium falciparum challenge is associated with HLA genotype. *Vaccine*, 36(12), 1637-1642. <https://doi.org/10.1016/j.vaccine.2018.01.069>

Nielsen, M., Lundegaard, C., & Lund, O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC bioinformatics*, 8(1), 1-12.

Nowak, M. A., May, R. M., Phillips, R. E., Rowland-Jones, S., Lalloo, D. G., McAdam, S., ... & McMichael, A. J. (1995). Antigenic oscillations and shifting immunodominance in HIV-1 infections. *Nature*, 375(6532), 606-611.

Oliver, S. E., Gargano, J. W., Marin, M., Wallace, M., Curran, K. G., Chamberland, M., ... & Dooling, K. (2020). The advisory committee on immunization practices' interim recommendation for use of Pfizer-BioNTech COVID-19 vaccine—United States, December 2020. *Morbidity and Mortality Weekly Report*, 69(50), 1922.

Oyarzún, P., & Kobe, B. (2016). Recombinant and epitope-based vaccines on the road to the market and implications for vaccine design and production. *Human vaccines & immunotherapeutics*, 12(3), 763-767.

Oyarzún, P., Ellis, J. J., Bodén, M., & Kobe, B. (2013). PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. *BMC bioinformatics*, 14(1), 1-11.

Oyarzún, P., Ellis, J. J., Gonzalez-Galarza, F. F., Jones, A. R., Middleton, D., Boden, M., & Kobe, B. (2015). A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: application to emerging infectious diseases. *Vaccine*, 33(10), 1267-1273.

Oyarzún, P., Kashyap, M., Fica, V., Salas-Burgos, A., Gonzalez-Galarza, F. F., McCabe, A., ... & Kobe, B. (2021). A proteome-wide immunoinformatics tool to accelerate T-cell epitope discovery and vaccine design in the context of emerging infectious diseases: an ethnicity-oriented approach. *Frontiers in immunology*, 12.

Pan, D., Sze, S., Minhas, J. S., Bangash, M. N., Pareek, N., Divall, P., ... & Hanif, W. (2020). The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. *EClinicalMedicine*, 100404.

Parham P., Janeway C. The immune system. 4th edition. New York, NY: Garland Science, Taylor & Francis Group; 2015

Peters, B., Nielsen, M., & Sette, A. (2020). T cell epitope predictions. *Annual Review of Immunology*, 38, 123-145.

Peters, B., Sidney, J., Bourne, P., Bui, H. H., Buus, S., Doh, G., ... & Sette, A. (2005). The design and implementation of the immune epitope database and analysis resource. *Immunogenetics*, 57(5), 326-336.

Plotkin, S. A., Orenstein, W., & Offit, P. A. (2012). *Vaccines E-Book*. Elsevier Health Sciences.

Poland, G. A., Ovsyannikova, I. G., & Kennedy, R. B. (2020). SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *The Lancet*.

Potocnakova, L., Bhide, M., & Pulzova, L. B. (2016). An introduction to B-cell epitope mapping and in silico epitope prediction. *Journal of immunology research*, 2016.

Prescott MJ, Lidster K (2017) Improving quality of science through better animal welfare: the NC3Rs strategy. *Lab Anim* 46:152–156

Rajčáni, J., Moško, T., & Režuchová, I. (2005). Current developments in viral DNA vaccines: shall they solve the unsolved?. *Reviews in medical virology*, 15(5), 303-325.

Ramaiah, A., & Arumugaswami, V. (2020). Insights into Cross-species Evolution of Novel Human Coronavirus 2019-nCoV and Defining Immune Determinants for Vaccine Development. *BioRxiv*, 2020.01.29.925867. <https://doi.org/10.1101/2020.01.29.925867>

Rauch, S., Jasny, E., Schmidt, K. E., & Petsch, B. (2018). New vaccine technologies to combat outbreak situations. *Frontiers in immunology*, 9, 1963.

Reche, P. A., & Reinherz, E. L. (2005). PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Research*, 33(suppl\_2), W138-W142.

Redoni, M., Yacoub, S., Rivino, L., Giacobbe, D. R., Luzzati, R., & Di Bella, S. (2020). Dengue: Status of current and under-development vaccines. *Reviews in Medical Virology*, e2101.

Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). 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*, 48(W1), W449-W454.

Rezaei, T., Khalili, S., Baradaran, B., Mosafer, J., Rezaei, S., Mokhtarzadeh, A., & de la Guardia, M. (2019). Recent advances on HIV DNA vaccines development: Stepwise improvements to clinical trials. *Journal of Controlled Release*, 316, 116-137.

Rezza, G., & Weaver, S. C. (2019). Chikungunya as a paradigm for emerging viral diseases: Evaluating disease impact and hurdles to vaccine development. *PLoS Neglected Tropical Diseases*, 13(1). <https://doi.org/10.1371/journal.pntd.0006919>

Rich, R. R., Fleisher, T. A., Shearer, W. T., Schroeder, H. W., Weyand, C. M., & Frew, A. J. (2019). *Inmunología clínica: principios y práctica*. Elsevier.

Robinson, J. M. (2016). Chapter 5 - Vaccine Production: Main Steps and Considerations. En B. R. Bloom & P.-H. Lambert (Eds.), *The Vaccine Book (Second Edition)* (pp. 77-96). Academic Press. <https://doi.org/10.1016/B978-0-12-802174-3.00005-9>

Robinson, J., Barker, D. J., Georgiou, X., Cooper, M. A., Flicek, P., & Marsh, S. G. (2020). Ipd-*imgt/hla* database. *Nucleic acids research*, 48(D1), D948-D955.

Sanchez-Mazas, A. (2020). HLA studies in the context of coronavirus outbreaks. *Swiss Medical Weekly*, 150(1516). <https://doi.org/10.4414/smw.2020.20248>

Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research*, 2017, 2680160. <https://doi.org/10.1155/2017/2680160>

Sandbrink, J. B., & Koblenz, G. D. (2022). Biosecurity risks associated with vaccine platform technologies. *Vaccine*, 40(17), 2514-2523.

Sautto, G. A., Kirchenbaum, G. A., Diotti, R. A., Criscuolo, E., & Ferrara, F. (2019). Next generation vaccines for infectious diseases

Schlottau, K., Rissmann, M., Graaf, A., Schön, J., Sehl, J., Wylezich, C., ... & Beer, M. (2020). SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. *The Lancet Microbe*, 1(5), e218-e225.

Schmidt, M. E., & Varga, S. M. (2018). The CD8 T cell response to respiratory virus infections. *Frontiers in immunology*, 9, 678.

Schuster, N. A. (2021). Characterization and structural prediction of the putative ORF10 protein in SARS-CoV-2. *bioRxiv*, 2020-10

Serdobova, I., & Kieny, M.-P. (2006). Assembling a Global Vaccine Development Pipeline for Infectious Diseases in the Developing World. *American Journal of Public Health*, 96(9), 1554-1559. <https://doi.org/10.2105/AJPH.2005.074583>

Shedlock, D. J., & Shen, H. (2003). Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science*, *300*(5617), 337-339.

Shiina, T., Hosomichi, K., Inoko, H., & Kulski, J. K. (2009). The HLA genomic loci map: Expression, interaction, diversity and disease. *Journal of Human Genetics*, *54*(1), 15-39. <https://doi.org/10.1038/jhg.2008.5>

Stead, W. W., Senner, J. W., Reddick, W. T., & Lofgren, J. P. (1990). Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *New England Journal of Medicine*, *322*(7), 422-427.

Tandrup Schmidt, S., Foged, C., Smith Korsholm, K., Rades, T., & Christensen, D. (2016). Liposome-based adjuvants for subunit vaccines: formulation strategies for subunit antigens and immunostimulators. *Pharmaceutics*, *8*(1), 7.

Tanriover, M. D., Doğanay, H. L., Akova, M., Güner, H. R., Azap, A., Akhan, S., ... & Aksu, K. (2021). Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim results of a double-blind, randomised, placebo-controlled, phase 3 trial in Turkey. *The Lancet*, *398*(10296), 213-222.

Toussaint, N. C., & Kohlbacher, O. (2009). OptiTope—a web server for the selection of an optimal set of peptides for epitope-based vaccines. *Nucleic acids research*, *37*(suppl\_2), W617-W622.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*. 2019 Jan 8;47(D1):D339-D343. doi: 10.1093/nar/gky1006. PMID: 30357391; PMCID: PMC6324067.

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>

*WHO Coronavirus Disease (COVID-19) Dashboard*. (s. f.). Recuperado 10 de Septiembre de 2020, de:

[https://covid19.who.int/?gclid=CjwKCAjwxLH3BRApEiwAqX9arZlpHFE2fLKdNA2jwFFJQik8qTDtgqp0woxYguEq0VafKf1jE-OVRoCGwCQAvD\\_BwE](https://covid19.who.int/?gclid=CjwKCAjwxLH3BRApEiwAqX9arZlpHFE2fLKdNA2jwFFJQik8qTDtgqp0woxYguEq0VafKf1jE-OVRoCGwCQAvD_BwE)

WHO Ebola Response Team. (2016). After Ebola in West Africa—unpredictable risks, preventable epidemics. *New England Journal of Medicine*, 375(6), 587-596.

Wilder-Smith, A., Chiew, C. J., & Lee, V. J. (2020). Can we contain the COVID-19 outbreak with the same measures as for SARS? *The Lancet Infectious Diseases*, 20(5), e102-e107. [https://doi.org/10.1016/S1473-3099\(20\)30129-8](https://doi.org/10.1016/S1473-3099(20)30129-8)

Wilder-Smith, A., Teleman, M. D., Heng, B. H., Earnest, A., Ling, A. E., & Leo, Y. S. (2005). Asymptomatic SARS Coronavirus Infection among Healthcare Workers, Singapore. *Emerging Infectious Diseases*, 11(7), 1142-1145. <https://doi.org/10.3201/eid1107.041165>

Wilken, L., & Rimmelzwaan, G. F. (2020). Adaptive immunity to dengue virus: slippery slope or solid ground for rational vaccine design?. *Pathogens*, 9(6), 470.

Wilson, M. E., & Chen, L. H. (2020). Travellers give wings to novel coronavirus (2019-nCoV). *Journal of Travel Medicine*, 27(2). <https://doi.org/10.1093/jtm/taaa015>

Wong, G., & Qiu, X. (2018). Funding vaccines for emerging infectious diseases. *Human vaccines & immunotherapeutics*, 14(7), 1760-1762.

Woolhouse, M. E. J., Dye, C., Taylor, L. H., Latham, S. M., & woolhouse, M. E. J. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1411), 983-989. <https://doi.org/10.1098/rstb.2001.0888>

World Health Organization (2003). Guidelines on nonclinical evaluation of vaccines. World Health Organization, Geneva

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263. <https://doi.org/10.1126/science.abb2507>

Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., ... & Handu, S. (2021). Role of structural and non-structural proteins and therapeutic targets of SARS-CoV-2 for COVID-19. *Cells*, 10(4), 821.

Zelenova, M., Ivanova, A., Semyonov, S., & Gankin, Y. (2021). Analysis of 329,942 SARS-CoV-2 records retrieved from GISAID database. *Computers in biology and medicine*, *139*, 104981.

Zepeda, G., Tapia, L., & Ortiz, P. (2020). Infección por SARS-CoV-2 y enfermedad por coronavirus-2019 en pediatría. *Revista chilena de enfermedades respiratorias*, *36*(2), 122-132.

Zhang, H., Lund, O., & Nielsen, M. (2009). The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics*, *25*(10), 1293-1299.

Zhang, H., Penninger, J. M., Li, Y., Zhong, N., & Slutsky, A. S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: Molecular mechanisms and potential therapeutic target. *Intensive Care Medicine*, *46*(4), 586-590. <https://doi.org/10.1007/s00134-020-05985-9>

Zhang, L., Udaka, K., Mamitsuka, H., & Zhu, S. (2012). Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Briefings in bioinformatics*, *13*(3), 350-364.

Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current biology*, *30*(7), 1346-1351.

Zhao, H., Harris, R. J., Ellis, J., & Pebody, R. G. (2015). Ethnicity, deprivation and mortality due to 2009 pandemic influenza A (H1N1) in England during the 2009/2010 pandemic and the first post-pandemic season. *Epidemiology & Infection*, *143*(16), 3375-3383.

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., & Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *BioRxiv*, 2020.01.26.919985. <https://doi.org/10.1101/2020.01.26.919985>

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270-273. <https://doi.org/10.1038/s41586-020-2012-7>

Zhou, Z., Huang, C., Zhou, Z., Huang, Z., Su, L., Kang, S., ... & Chen, S. (2021). Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience*, 24(3), 102187.

Zhu, J., & Paul, W. E. (2008). CD4 T cells: fates, functions, and faults. *Blood, The Journal of the American Society of Hematology*, 112(5), 1557-1569.