



UNIVERSITAT
ROVIRA I VIRGILI

DEFINICIÓN DEL REPERTORIO INMUNOLÓGICO “NORMAL”

Lorena Hernández Servian

TRABAJO DE FINAL DE GRADO DE BIOTECNOLOGÍA



Tutora académica: Ana Fernández Bravo, Grado en Biología, Ciencias Médicas Básicas, ana.fernandez@urv.cat

En cooperación con: *Institut de Génétique Humaine* (IGH)

Supervisora: Nika ABDOLLAHI, Grado en Biología, Bases Moleculares de Patologías Humanas, nika.abdollahi@igh.cnrs.fr

Junio de 2022



1. AGRADECIMIENTOS

En primer lugar, quería agradecer a Nika ABDOLLAHI, mi tutora de prácticas, por ayudarme en todo, solucionarme cualquier duda e involucrarse tanto en el trabajo. También me gustaría agradecerle los conocimientos que he adquirido en bioinformática gracias a ella y la paciencia que ha tenido cuando el programa daba errores.

En segundo lugar, me gustaría agradecer a Sofia KOSSIDA, directora de IMGT, por aceptarme en las prácticas, darme una buena acogida.

Asimismo, me gustaría agradecer a mi tutora del trabajo, Ana Fernández Bravo, por su paciencia, por ayudarme a organizarlo todo y estar disponible para cualquier duda.

Por último, me gustaría darle las gracias a mis padres y familiares por su cariño y el apoyo que he recibido en todo momento. Gracias a ellos he podido realizar mi sueño de realizar mis estudios en biotecnología y finalizarlos en Francia.



ÍNDICE

1. AGRADECIMIENTOS.....	2
ÍNDICE	3
2. DATOS DEL CENTRO	5
3. GUÍA PREVENCIÓN DE PLAGIO	5
4. RESUMEN.....	6
5. ABSTRACT.....	6
6. ABREVIATURAS	7
7. INTRODUCCIÓN.....	8
7.1 Sistema inmunitario.....	8
7.1.2 Células mieloides.....	9
7.1.3 Células linfoides.....	9
7.2 Tejidos y órganos del sistema inmunitario	10
7.2.1 Órganos linfoides primarios	10
Órganos linfoides secundarios.....	11
7.3 Inmunidad innata.....	12
7.3.1 “Toll-like receptor” (TLR).....	12
7.3.2 Activación de PRR y modificación de la expresión genética	12
7.3.3 Citotoxicidad de linfocitos NK	12
7.4 Antígenos.....	13
7.4.1 Características.....	13
7.4.2 Moléculas presentadoras de antígenos	14
7.5 Reconocimiento de antígenos.....	15
7.5.1 Antígenos timo-dependientes	15
7.5.2 Antígenos timo-independientes	15
7.6 Inmunidad adaptativa.....	16
7.6.1 Sinapsis inmunológica	16
7.6.2 Activación y diferenciación de linfocitos T	16
7.6.3 Activación y diferenciación de linfocitos B	17
7.7 Diversidad del reconocimiento antigénico	18
7.7.1 Reordenamiento de los genes codificante de los dominios variables de las inmunoglobulinas (Ig)	18
7.7.2 Diversidad de los dominios variables.....	19
7.7.3 Diversidad de los TCR.....	19



8. INTERÉS, HIPÓTESIS Y OBJETIVO	20
9. METODOLOGÍA	21
9.1 Bioinformática para el análisis comparativo de repertorios	21
9.1.1 “ImmuneREF”	21
9.1.2 “IMGT/StatClonotype”	24
9.2 Datos de secuenciación de alto rendimiento	26
9.2.1 Artículo sobre el estudio del COVID	26
9.2.2 Método de secuenciación	27
9.2.3 Buscar y descargar datos	29
10. RESULTADOS Y DISCUSIÓN	30
10.1 Resultados repertorio H3	30
10.2 Resultados repertorio H4	33
10.3 Resultados repertorio M5	35
10.4 Resultados repertorio S26	37
10.5 Comparación H3 con H4	40
10.6 Comparación H3 con M5	41
10.7 Comparación H3 con S26	43
10.8 Comparación de todos los repertorios	44
11. CONCLUSIONES	46
12. AUTOEVALUACIÓN	46
13. BIBLIOGRAFÍA	47
ANEXOS	50



2. DATOS DEL CENTRO

Este trabajo está basado en las prácticas que he realizado en “ImMunoGeneTics” (IMGT).

IMGT® es el sistema de información internacional ImMunoGeneTics®, <http://www.imgt.org/>, a la vanguardia de la inmunoinformática y la inmunogenética con más de 30 años de experiencia. IMGT® proporciona a la comunidad científica bases de datos y herramientas relacionadas con las respuestas inmunitarias adaptativas basadas en la ontología IMGT. IMGT® tiene tres ejes de investigación y desarrollo (I+D). El primer eje incluye la comprensión de las respuestas inmunitarias adaptativas, descifrando la identificación y caracterización de los genes de inmunoglobulina (Ig) y receptor de células T (TCR) en vertebrados con mandíbula. El punto de partida para los otros dos ejes, son el análisis y exploración del repertorio de expresión de Ig y TCR en base a la comparación con el catálogo de referencia de IMGT en condiciones normales y patológicas (eje II) y el análisis de cambios y funciones de aminoácidos, de la ingeniería de anticuerpos y TCR (eje III). IMGT es parte del departamento de " Bases Moleculares de las Enfermedades Humanas " del Instituto de Genética Humana (IGH), una unidad de investigación conjunta del “Centre national de la recherche scientifique” (CNRS) - Universidad de Montpellier (UM)

3. GUÍA PREVENCIÓN DE PLAGIO

Yo, Lorena Hernández Servian, con DNI 25365567V, soy conocedor de la guía de prevención del plagio en la URV Prevención, detección y tratamiento del plagio en la docencia: guía para estudiantes (aprobada en julio 2017) (<http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formaciocompetencies-nuclears/plagi/>) y afirmo que este TFG no constituyen ninguna de las conductas consideradas como a plagio por la URV.

Tarragona, 7 de junio de 2022



4. RESUMEN

Actualmente, pensamos que las enfermedades producen cambios en nuestros repertorios inmunitarios. Para poder responder a esta pregunta, hemos utilizado dos herramientas “immuneREF” y “IMGT/StatClonotype”, que nos permitirán distinguir los estados de salud de los estados patológicos analizando una serie de características.

En este trabajo analizaremos 6 repertorios inmunológicos. 2 de ellos son no patológicos (H3 y H4) y los demás, son patológicos, provienen de pacientes con COVID.

He analizado y comparado los resultados entre los repertorios y los resultados obtenidos son los esperados en comparación con el artículo de base (Weber, 2022).

Podemos ver que hay diferencias entre los repertorios enfermos y normales, pero no son significativas. Esto se debe seguramente a que los datos proporcionados por el artículo (Kuri-Cervantes et al., 2020), no han tenido en cuenta el sexo ni la edad de los pacientes.

Palabras clave: repertorio inmunitario, “immuneREF”, “IMGT/Stat-clonotype”, patológico, no patológico.

5. ABSTRACT

We currently believe that diseases cause changes in our immune repertoires. In order to answer this question, we have used two tools “immuneREF” and “IMGT/StatClonotype”, which will allow us to distinguish between healthy and pathological states by analysing a series of characteristics.

In this work we will analyse 6 immunological repertoires. Two of them are non-pathological (H3 and H4) and the others are pathological, coming from patients with COVID.



I have analysed and compared the results between the repertoires and the results obtained are as expected compared to the base article (Weber, 2022).

We can see that there are differences between the diseased and normal repertoires, but they are not significant. This is probably due to the fact that the data provided by the article (Kuri-Cervantes et al., 2020) did not take into account the sex and age of the patients.

Keywords : immune repertoire, immuneREF, IMGT/Stat-clonotype, pathological, non-pathological.

6. ABREVIATURAS

ABREVIATURAS	SIGNIFICADO
AA	Aminoácidos
ADN	Ácido desoxiribonucleico
AR	"Artritis reumatoide"
ARN	Ácido ribonucleico
BCR	Receptor de célula B
CDR(1,2,3)	Regiones hipervariables
CMH	Complejo mayor de histocompatibilidad
CNRS	"Centre national de la recherche scientifique"
CPA	Célula presentadora de antígenos
dNTP	desoxirribonucleótidos
H	Cadena "heavy" (pesada)
H3	Repertorio "healthy 3"
H4	Repertorio "healthy 4"
HLA	"human leucocyte antigen"
HMS	Hipermutaciones somáticas
I + D	Investigación y desarrollo
Ig	Inmunoglobulinas
IgE	Inmunoglobulina E
IgG	Inmunoglobulina G
IGH	"Institute de Génétique Humaine"
IL-2	Interleuquina 2
IMGT	"ImMunoGeneTics"
INy	Interferón gama
L	Cadena ligera
LES	"Lupus eritematoso sistémico"
L _T	Linfocito T
M5	Repertorio patológico moderado 5
M6	Repertorio patológico moderado 6
NF-κβ	"Nuclear factor-kappa beta "
NK	Linfocitos granulares
NKT	Linfocitos T "natural killer"
PALS	"Periarteriolar lymphoid sheath"
PAMPs	"Pathogen associated molecular pattern"
PBMC	Células mononucleares de la sangre periférica
PRR	"Pattern-recognition receptor"
RE	Retículo endoplasmático
RFcE1	receptores para la inmunoglobulina E
S24	Repertorio patológico severo 24
S26	Repertorio patológico severo 26
SD	Desviación estándar
Tc	Linfocitos T citotóxicos
TCR	Receptor de las células T
TGFβ	"Transforming growth factor"
Th	Linfocitos T "helpers"
TLR	"Toll-like receptor"
Treg	Linfocitos T reguladores
UM	"Université de Montpellier"

Tabla 1: Abreviaciones

7. INTRODUCCIÓN

7.1 Sistema inmunitario

La función principal del sistema inmunológico es mantener la integridad biológica del organismo, para ello debe:

- Reconocer y tolerar los auto-antígenos.
- Reconocer y eliminar los exo-antígenos.

Existen dos tipos de inmunidades; la innata y la adaptativa.

La inmunidad innata es la primera línea de defensa del organismo. En cambio, la adaptativa se basa en el reconocimiento de antígenos y la creación de memoria inmunitaria.

Para entender el sistema inmunitario, vamos a ver las células que participan en este sistema.

7.1.1 Células del sistema inmunitario

Una célula hematopoyética es una célula inmadura que tiene la capacidad de transformarse en cualquier célula sanguínea.

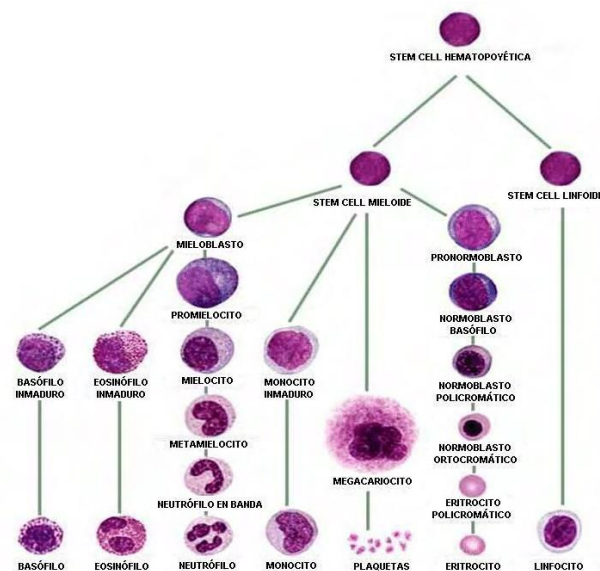


Figura 1: Sacada de un artículo (Limpe, 2007). Podemos ver como la célula hematopoyética se puede diferenciar en dos líneas celulares, mieloides o linfoides. Después según el camino que haya elegido, se diferenciará en un tipo de célula.



7.1.2 Células mieloides

En la primera línea de diferenciación, encontramos por un lado las células mieloides que, según el tipo de infección e intermediarios, se diferenciará en un tipo específico de célula.

Nos vamos a centrar en los basófilos y mastocitos juegan un papel importante en la hipersensibilidad ya que, tienen en la membrana, receptores para la inmunoglobulina E (RFcE). De esta manera, cuando una inmunoglobulina E (IgE) los activan, estas células liberarán sus mediadores químicos (entre ellos la histamina, responsable de la respuesta inflamatoria).

Las células dendríticas, también son importantes puesto que, son células presentadoras de antígenos para los linfocitos T (L_T).

7.1.3 Células linfoides

Las células linfoides es la otra línea celular en la que se puede diferenciar una célula hematopoyética.

Dentro de las células linfoides, hay una división entre linfocitos lisos (B y T) y linfocitos granulares (NK).

7.1.3.1 *Linfocitos NK*

Los linfocitos granulares expresan el marcador CD16, receptor específico de IgG de manera que los reconoce y destruye la célula recubierta por estas inmunoglobulinas.

7.1.3.2 *Linfocitos T*

Los linfocitos T se diferencian en el timo. Estos tienen un receptor membranal formado por un heterodímero de glicoproteínas llamado "T-cell receptor" (TCR); el TCR está asociado a unas proteínas del correceptor CD3.

Hay 3 tipos de L_T: linfocitos T_{αβ}, linfocitos T_{γδ} y los linfocitos T "natural killer" (NKT).

- Los linfocitos T_{αβ} están diferenciados en dos subpoblaciones: los que tienen el marcador CD4 y los CD8.

- Los linfocitos T CD4⁺ constituyen los "T helpers" (Th) y los T reguladores (Treg). La función principal de estos linfocitos es modular la actividad de las demás células inmunitarias.



- Los linfocitos T CD8⁺ son los T citotóxicos (T_c), como indica su nombre, lisan las células.
- Los linfocitos T_{γδ} reconocen Ag por su TCR y las células en estado de estrés por el receptor NKG2D.
- Los linfocitos NKT tienen un TCR de tipo αβ, por lo tanto, reconoce glicolípidos presentados por la molécula CD1D. A su vez, tienen el marcador NK1.1. Su función es ayudar a la diferenciación de monocitos en células dendríticas durante la respuesta de inflamación.

7.1.3.3 *Linfocitos B*

El linfocito B inmaduro, presentan IgM e IgD en su membrana; estas pueden unirse a antígenos específicos; y junto con las proteínas del correceptor CD79, forman el “B-cell receptor” (BCR). Una vez los linfocitos B han estado en contacto con el antígeno y este es endocitado, la célula B se convierte en una célula presentadora de antígenos (CPA).

7.2 Tejidos y órganos del sistema inmunitario

Para comprender el sistema inmunitario, primero necesitamos saber dónde y cómo se desarrollan los linfocitos. Los diferentes linfocitos se desarrollan y maduran en los órganos linfáticos primarios y luego acaban la maduración en los órganos linfoides secundarios.

7.2.1 Órganos linfoides primarios

Los órganos linfoides primarios están compuestos por el timo y la médula ósea. Aquí se producirá el desarrollo de los pre-linfocitos.

7.2.1.1 *Timo*

En este órgano se produce la maduración de los linfocitos T. El timo está constituido por dos lóbulos; el córtex (dónde se produce la selección positiva) y la médula (selección negativa).

Los precursores de los linfocitos T se llaman timocitos. Según el tipo de receptor que tenga cada timocito, madurarán de una forma distinta.

Los timocitos con TCR_{γδ}, se dirigen directamente a la médula para acabar la maduración.



En cambio, los que tienen el receptor $TCR_{\alpha\beta}$, se someten a una etapa de selección positiva. Esta selección consiste en la interacción con el complejo mayor de histocompatibilidad (CMH-I o CMH-II). En el caso de $L_T CD4$, los TCR reconocerán a CMH-II y los $L_T CD8$, los CMH-I (la función principal de este complejo es presentar péptidos de los antígenos a los linfocitos). De esta manera, los linfocitos que no han reconocido los CMH, mueren por apoptosis excepto los que son capaces de reconocer las moléculas CD1D; expresadas por $L_T CD4$ y $CD8$; por lo tanto, se diferenciarán en linfocitos NKT.

Los linfocitos que han sobrevivido se dirigen al lóbulo de la médula. Las células dendríticas y los macrófagos expresan epítomos (parte que reconocerá el parátipo de un anticuerpo y se unirán) de autoantígenos asociados a CMH-I o CMH-II; si los TCR presentan una gran afinidad por CMH, mueren por apoptosis, si no los reconocen, se convierten en linfocitos T efectores y si tienen una afinidad por CMH moderada, se convierten en linfocitos T reguladores (Treg) (Akbar et al., 2014). Los NKT también pasan por esta selección negativa.

Una vez han pasado las dos selecciones, los linfocitos son dirigidos a los órganos linfáticos secundarios.

7.2.1.2 Médula ósea

Es dónde se diferencian los linfocitos B y dónde se produce la selección negativa de estos. Hay dos mecanismos implicados en la selección de linfocitos; la deleción clonal y la anergia.

La deleción clonal empieza consiste en la eliminación de los linfocitos B autorreactivos.

La anergia consiste en que los linfocitos B disminuyen la expresión de su IgM porque hay una alta concentración de antígenos; de tal manera que el linfocito no es capaz de interaccionar con los linfocitos Th , produciéndose un estado de anergia.

Órganos linfoides secundarios

Aquí se produce la diferenciación de los linfocitos. Los órganos linfoides secundarios son el bazo y los ganglios linfáticos.



7.2.1.3 *Bazo*

El bazo está compuesto por la pulpa blanca y la roja. En la pulpa blanca, encontramos la vaina linfática periarterial (PALS) dónde se inician las activaciones de los linfocitos B y T.

7.2.1.4 *Ganglios linfáticos*

Los ganglios, están constituidos por el córtex, el paracórtex y la médula. Dentro de los ganglios encontramos una gran concentración de linfocitos que serán transportados a través del sistema linfático ayudando a que cada linfocito encuentre un antígeno específico y empezar la respuesta inmunitaria adaptativa.

Una vez ya conocemos cómo y dónde se forman los linfocitos, podemos introducir la inmunidad innata y adaptativa.

7.3 Inmunidad innata

Es una defensa muy rápida contra las infecciones. Al ser rápida, se produce un reconocimiento no específico gracias al reconocimiento de “pathogen associated molecular pattern” (PAMPs). Los PAMPs son motivos comunes en muchos microorganismos. Estos PAMPs van a ser reconocidos por los “pattern-recognition receptor” (PRR). Concretamente hablamos de los “toll-like receptors” (TLR) ya que son los que están implicados en las infecciones y las células NK.

7.3.1 “Toll-like receptor” (TLR)

Los TLR son glicoproteínas transmembranales. Hay diferentes tipos y cada uno reconoce una gama diferente de PAMPs. A continuación, veremos cómo funciona el reconocimiento entre un TLR.

7.3.2 Activación de PRR y modificación de la expresión genética

Cuando un TLR reconoce un PAMP, se produce una activación de la región citoplasmática por fosforilación, produciendo una cascada de señalizaciones que conllevará a la liberación de nuclear factor-kappa beta (NF- κ B). NF- κ B entra en el núcleo activando la expresión de muchos genes que participan en la respuesta inmunitaria.

7.3.3 Citotoxicidad de linfocitos NK

Las células NK también presentan receptores para los PRR produciendo señales de activación o inhibición.



Antes de introducir la inmunidad adaptativa, voy a definir y explicar algunos conceptos clave como los antígenos y su reconocimiento.

7.4 Antígenos

Hay dos tipos de antígenos, los timo-dependientes, son los que activan los componentes celulares y humorales; y los timo-independientes, que dan lugar a una respuesta con anticuerpos.

7.4.1 Características

Las características principales de los antígenos son su inmunogenicidad, su epítipo y su especificidad con el parátipo de los anticuerpos.

7.4.1.1 *Inmunogenicidad de los antígenos*

Hay antígenos que son capaces de inducir una respuesta adaptativa, estos antígenos son considerados inmunógenos. Sin embargo, también existen antígenos, suelen ser de un tamaño pequeño, que carece de inmunogenicidad, estos son denominados haptenos.

La inmunogenicidad depende de diferentes parámetros:

- Cuanta más heterogeneidad, mayor es la inmunogenicidad.
- Cuanto mayor es la solubilidad, menor es la inmunogenicidad, pero, puede aumentar añadiendo adyuvantes que inducen su precipitación.
- Las concentraciones muy altas o bajas de antígenos producen una parálisis inmunitaria.

7.4.1.2 *Epítipo*

El epítipo es la región del antígeno reconocida por los anticuerpos de una Ig o por el sitio de reconocimiento de los TLR. La función principal del epítipo es determinar la especificidad antigénica.

Existen dos tipos diferentes de epítipos:

- Secuencial: con una secuencia peptídica continua.
- Conformacional: dos secuencias peptídicas diferentes pero cercanas en el espacio.

Existen los epítomos crípticos, para que lleguen a ser accesibles, la proteína debe desnaturalizarse y de esta manera pueden ser reconocidos por los linfocitos.

Ahora que sabemos cómo nuestros anticuerpos reconocen antígenos, necesitamos conocer cómo se produce la presentación de antígenos por parte de las CPA.

7.4.2 Moléculas presentadoras de antígenos

Las moléculas presentadoras de antígenos son los complejos mayores de histocompatibilidad (CMH); en los humanos, hablaremos de “human leucocyte antigen” (HLA) y a continuación veremos los dos tipos.

7.4.2.1 “Human leucocyte antigen – I” (HLA-I)

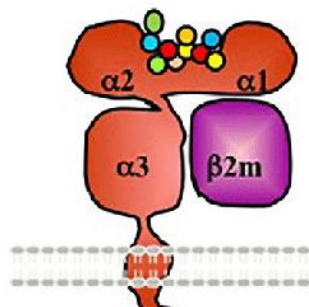


Figura 2: Sacada del artículo (Hickey & Valenzuela, 2016)

La cadena α está compuesta por 3 dominios donde en $\alpha 1$ se encuentra la parte (N terminal); una región transmembranal y una región citoplasmática donde se encuentra (C terminal). En el dominio $\alpha 3$ tiene un sitio de interacción con la molécula CD8 de los linfocitos T citotóxicos.

La función de la $\beta 2$ -microglobulina es estabilizar la cadena α .

Todas las células nucleadas presentan esta molécula salvo las células germinales, del hueso, del cartílago y neuronas. Van a presentar péptidos antigénicos a los linfocitos CD8.

7.4.2.2 “Human leucocyte antigen – II” (HLA-II)

Está compuesto por dos cadenas.

En la cadena α , encontramos los dominios externos $\alpha 1$ (dominio N-terminal) y $\alpha 2$.

En la cadena β , también encontramos los dos dominios $\beta 1$ y $\beta 2$; siendo el primero el dominio N terminal. El dominio $\beta 2$ interacciona con las moléculas CD4.

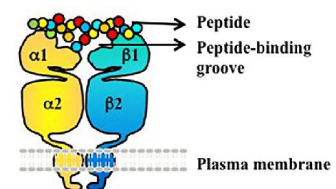


Figura 3: Sacada del artículo “Alloantibody Generation and Effector Function Following Sensitization to HLA (2016)

De esta manera, los dos dominios N terminales, van a formar la estructura dónde se unirá el péptido del antígeno. Como esta estructura está formada por



dos dominios, a diferencia de HLA-I, el péptido puede ser más largo (15 – 20 aminoácidos).

Este tipo de molécula es presentado por las CPA y reconoce linfocitos CD4.

7.5 Reconocimiento de antígenos

El reconocimiento de antígenos se puede dar de dos maneras, dependientes del timo o de forma independiente.

7.5.1 Antígenos timo-dependientes

En este tipo de reconocimiento, entran en juego los linfocitos T. En el caso que el reconocimiento se dé por linfocitos T CD4⁺, el reconocimiento se dará por la molécula HLA-II, que reconoce exo-antígenos; en cambio, los linfocitos T CD8⁺, al reconocer HLA-I, generalmente reconocerá endo-antígenos.

7.5.1.1 *Endo-antígenos peptídicos*

Estos antígenos provienen de la degradación de la proteasoma. Al degradarse, los péptidos liberados serán transportados por el retículo endoplasmático (RE). Dentro del RE, los péptidos se fijan a unas moléculas del CMH-I, formando un complejo que transita por el aparato de Golgi. Dentro del aparato de Golgi, se produce la glicosilación del HLA-I y una vez formado, se dirige a la membrana plasmática.

7.5.1.2 *Exo-antígenos peptídicos*

Estos antígenos provienen de la digestión de una fagolisosoma; los péptidos se dirigen al aparato de Golgi que contienen el complejo CMH-II/CLIP. Como estos péptidos son específicos para CMH-II, se van a unir desplazando CLIP.

Las células que presentan antígenos exógenos son las células dendríticas ya que son las únicas células capaces de presentar los dos CMH (de tipo I y II); el proceso por el cual la célula dendrítica presenta los péptidos de un antígeno exógeno a un CMH-I se llama presentación cruzada.

7.5.2 Antígenos timo-independientes

Algunos antígenos son capaces de activar una subpoblación de linfocitos B (B1) sin necesidad. En esta inmunidad no se van a crear células B de memoria y los anticuerpos producidos son IgM e IgA.

Este tipo de respuesta se produce en el bazo y a nivel de mucosas.



Una vez introducido todos los conceptos anteriores, podemos pasar a la definición de inmunidad adaptativa.

7.6 Inmunidad adaptativa

Esta inmunidad empieza en los tejidos linfáticos secundarios, y ahí se dará la presentación de antígenos a los linfocitos. La característica principal de esta inmunidad es que al final del proceso, se producirán linfocitos de memoria.

Primero voy a introducir el funcionamiento básico de esta inmunidad, es decir la sinapsis inmunológica y la activación de los linfocitos B y T.

7.6.1 Sinapsis inmunológica

La sinapsis inmunológica consiste en tres señales producidas entre una CPA y un linfocito T. En la estructura de una CPA encontramos su CMH junto con una molécula co-estimuladora.

La primera señal consiste en el reconocimiento por medio del TCR a un determinante antigénico asociado a un CMH.

Una segunda señal es necesaria para activar un linfocito T no maduro; la CPA tiene una molécula CD80 o CD86 que se va a liar a la molécula CD28 (receptor de la molécula co-estimuladora) del linfocito T. Esta segunda señal va a producir interleuquina 2 (IL-2) y permitir la expresión del receptor de IL-2 por parte del linfocito.

La última señal se produce con la unión de la IL-2 y su receptor. Con lo que llevará a su posterior activación.

En el caso de los linfocitos T CD4⁺, concretamente los Th, la segunda señal induce a la síntesis de CD40L. Este ligando interacciona con la molécula CD40 de la CPA, y provocará una liberación de citoquinas implicadas en la orientación funcional del linfocito.

7.6.2 Activación y diferenciación de linfocitos T

Una vez separado el linfocito de la CPA, empezará la expansión clonal, es decir, ese linfocito va a replicarse para tener una gran concentración de linfocitos específicos para ese antígeno.

Después, se da a cabo la diferenciación gracias a las citoquinas liberadas. Los linfocitos T CD4⁺, pueden diferenciarse en Th1, Th2 y Th17. Los Th1 están



inducidos por la IL-12; este tipo de linfocitos inducen respuestas celulares porque va a producir que los linfocitos B, produzcan IgG (función de opsonización, marcando antígenos para que sean fagocitados por los macrófagos); los Th1 también activan los linfocitos T citotóxicos y por medio de la producción del interferón gamma (IN γ), aumentan la actividad fagocítica de los macrófagos.

Los Th2, inducidos por IL-4, produce una respuesta humoral. Activan los linfocitos B para que produzcan anticuerpos.

Para acabar con esta familia de linfocitos, tenemos los Th17 producidos por la sinergia entre IL-6 y “transforming growth factor” (TGF β), que están implicados en la respuesta inflamatoria.

Los linfocitos T CD8⁺, se pueden activar de forma dependiente o independiente de linfocitos Th.

De forma independiente, se va a producir la sinapsis inmunológica como he explicado antes.

De forma dependiente de Th. Una vez el Th esté activado por una célula dendrítica, va a activar a esta misma CPA. Esta CPA activada va a exponer a la célula co-estimuladora de su CMH-I; por lo tanto, podrá activar el linfocito T CD8⁺. Posteriormente se dará la expansión clonal.

7.6.3 Activación y diferenciación de linfocitos B

El linfocito B se puede activar de forma dependiente de Th o independiente.

De forma independiente de Th, el linfocito B va a reconocer el antígeno por su BCR de forma directa y empieza la expansión clonal.

De forma dependiente de Th, cuando el linfocito B reconoce al antígeno a través de su BCR, lo va a endocitar, degradar y luego va a exponer los péptidos a través de su CMH-II. En paralelo, otra CPA va a presentar el mismo epítipo a un linfocito Th2. Una vez se produce la activación del linfocito Th2. Ahora este linfocito T, por medio de su TCR, va a reconocer los péptidos antigénicos presentados por el linfocito B. El linfocito Th2, va a producir IL-4 que activará el linfocito B. Por lo tanto, se producirá la expansión clonal y la diferenciación en célula plasmática o célula de memoria.



Como podemos ver, tenemos muchas subfamilias de linfocitos y cada uno de ellos tiene una especificidad concreta para el antígeno. Posteriormente se explica cómo se da esa especificidad.

7.7 Diversidad del reconocimiento antigénico

Para que se de esa especificidad entre epítopo y parátopo, nuestros antígenos tienen que reordenar los genes que codifican para los dominios variables de las inmunoglobulinas. Al igual que en los BCR, en los TCR también se dará un reordenamiento de genes que definirá la especificidad de los linfocitos T.

7.7.1 Reordenamiento de los genes codificante de los dominios variables de las inmunoglobulinas (Ig)

Primero vamos a localizar los genes codificantes para las cadena pesadas y ligeras de las inmunoglobulinas y luego, explicamos el proceso que se da para su reordenamiento.

7.7.1.1 Organización del locus de las Ig

Tanto para las cadenas pesadas (H) y ligeras (L), tenemos 4 tipos de genes; los variables (V), genes de diversidad (D), de unión (J) y constantes (C). Los 3 primeros (V, D y J), pertenecen a la región variable de la Ig y C pertenece a la región constante.

El locus de estos genes se llama IGH (cadena H). Compuesto por IGHV, IGHD, IGHV, IGHC. IGHC; se encuentran en el cromosoma 14. En cambio, para la cadena L, encontramos dos locus diferentes IGK, para la cadena kappa en el cromosoma 2, e IGL, para la cadena lambda y se encuentra en el cromosoma 22.

7.7.1.2 Reordenamiento de V-(D)-J y síntesis de las cadenas pesadas (H) y ligeras (L)

La formación de las cadenas de Ig se da durante la diferenciación de los linfocitos B. Necesitamos reordenar el ADN somático de manera independiente del ag. Por lo tanto, primero se reordenarán los genes de la cadena H y luego, los de la cadena L.

- Cadena H:

Unión de los genes IGHD-IGHJ; se produce la delección de los genes intermediarios y conseguimos una secuencia DJ reorganizada. Después se da la unión de IGHV-D-J; empieza con la transcripción con el gen IGHM, por lo



tanto, tenemos ARN pre-mensajero, IGHV-D-J-Cm (IGHM está compuesto por 4 exones CH1-CH4; se dará un “splicing” con lo que tendremos como resultado el ARNm. (Chi et al., 2020) Luego se dará la traducción, y obtendremos una cadena polipeptídica. El péptido señal es eliminado después de la penetración de la cadena dentro de la cavidad del retículo endoplasmático.

- Cadena L:

Unión de los genes IGKV-IGKJ; habrá una delección de los genes intermediarios con lo cual obtendremos la secuencia VJ reordenada. Posteriormente, se dará la transcripción con el gen IGKC a ARN pre-mensajero, luego se producirá un “splicing” hasta obtener nuestro ARNm (Chi et al., 2020). Finalmente, habrá la traducción y se formará una cadena polipeptídica y al igual que la cadena H, la señal desaparecerá cuando la cadena penetre la cavidad del retículo endoplasmático.

Como podemos ver, al reordenar los genes podemos obtener muchísimas combinaciones diferentes, de manera que nos proporciona una gran diversidad.

7.7.2 Diversidad de los dominios variables

A parte de la reorganización de los genes, la diversidad de los dominios de las inmunoglobulinas viene dado por las hipermutaciones somáticas.

7.7.2.1 *Hipermutaciones somáticas (HMS)*

La HMS consiste en mutaciones que cambian un nucleótido por otro diferente en las regiones hipervariables (CDR1, CDR2 y CDR3) de las Ig. La variabilidad producida por las mutaciones va a generar anticuerpos con más afinidad para el antígeno.

7.7.3 Diversidad de los TCR

Como he mencionado anteriormente, los TCR también sufren una reorganización de los genes que les proporciona una cierta diversidad; pero, además, los TCR sufren adiciones de nucleótidos que también les proporciona una mayor diversidad.

7.7.3.1 *Organización de los genes codificante de los dominios variables de las cadenas de los TCR*

Los genes codificantes de los dominios variables de los TCR se organizan de manera similar que los de Ig. LA diferencia es que tienen un único gen C y que



el dominio variable de una cadena está codificado por V-D-J (β y δ) en cambio la otra cadena, está codificada por V-J (α y γ).

TCRB y TCRG están en el cromosoma 7 y TCRA y TCRD en el cromosoma 14.

7.7.3.2 *Biosíntesis de los dominios variables de las cadenas de los TCR*

Los dominios variables se crean de la misma forma que las Ig seguido de la creación de regiones P y N. Esto producirá una diversidad de las uniones casi ilimitada. Las regiones P se crean al rellenar las zonas de cortes con “nucleótidos P” y las regiones N son producidas al añadir nucleótidos de manera aleatoria. La diversidad de unión también se puede dar por la eliminación de nucleótidos.

8. INTERÉS, HIPÓTESIS Y OBJETIVO

El repertorio inmunitario adaptativo es el conjunto de receptores inmunitarios de anticuerpos y células T de un individuo, y representa tanto el estado inmunitario actual como el anterior del individuo.

Actualmente se piensa que después de estar en contacto con una enfermedad, se producen cambios en los repertorios inmunitarios. Definimos como repertorio inmunitario, las secuencias del ADN de los receptores BCR y TCR.

Numerosos estudios destacan la correlación entre la diversidad de repertorio inmunitario con los estados inmunitarios de un individuo (sano, enfermedad, infección, etc. (Ruschil et al., 2021). Para agruparlos, definiremos repertorio sano y repertorio patológico para poder representar cualquier diferencia que pueda asociarse a una patología. Esto nos va a servir para los posteriores análisis que servirán como guía para el diagnóstico de enfermedades, terapia e indicaciones clínicas. La mayoría de las herramientas disponibles para explorar la diversidad del repertorio están adaptadas a repertorios desequilibrados, es decir, a repertorios con una extensión significativa de uno o más linajes de células inmunitarias. (cáncer, infección grave, etc.).

En los sujetos "sanos", existe una variabilidad fisiológica del repertorio inmunitario que puede estar influida por la edad, el entorno, el grupo étnico, los antecedentes personales, etc. Nuestra hipótesis es que la elección del repertorio normal para estudiar una patología no es trivial, porque dependiendo



del estado del repertorio normal elegido, la desviación de la "normalidad" puede ser más o menos importante.

El objetivo principal de este proyecto es iniciar un estudio sobre el análisis de la diversidad del repertorio inmunitario no patológico (normal) en los seres humanos. El proyecto está organizado en varias etapas:

1. Encontrar las secuencias del repertorio "sano". Estos repertorios pueden describirse como "sanos", "no alterados" o "no patológicos" en la literatura. Es esencial establecer diferentes categorías de datos según la edad, el tejido muestreado, la profundidad de muestreo, el grupo de marcadores de diferenciación, etc.
2. Analizar los repertorios seleccionados con las herramientas proporcionadas por IMGT (IMGT/HighV-QUEST e IMGT/StatClonotype).
3. Caracterizar la colección de repertorios seleccionados con immuneREF (Wever, 2022).

9. METODOLOGÍA

9.1 Bioinformática para el análisis comparativo de repertorios

Para poder ver el impacto de las elecciones de repertorios normales en el análisis de una patología; vamos a comparar diferentes tipos de repertorios (normales y patológicos) mediante dos herramientas de análisis de repertorios: con "immuneREF" (Wever, 2022) e "IMGT/StatClonotype" (Aouinti et al., 2016; Aouinti et al., 2015). Estas herramientas están escritas en el lenguaje R y se dedican a la ciencia de los datos, el análisis estadístico y la representación gráfica. Elegimos "immuneREF" porque es una herramienta de reciente publicación que ofrece un análisis comparativo de directorios a gran escala. "IMGT/StatClonotype" ofrece un análisis profundo y detallado de la comparación de dos directorios.

9.1.1 "ImmuneREF"

"ImmuneREF" es una herramienta que nos permite cuantificar la similitud de diferentes repertorios inmunológicos basándonos en características específicas. Este instrumento establece un diccionario con un conjunto de datos tanto simulados como experimentales; de esta manera, cada conjunto de datos analizado se usará posteriormente como referencia para poderlo



comparar con otros repertorios. A parte de comparar diferentes repertorios, también nos ofrece información individual.

9.1.1.1 Características

Para cada repertorio inmunológico vamos a calcular las siguientes características; la expansión clonal, la frecuencia de las posiciones de los aminoácidos, la arquitectura de la red de similitud de secuencias, la convergencia, la diversidad genética en la línea germinal y la aparición de “k-mers”.

- Perfiles de uniformidad (expansión clonal):

Esta característica se basa en la uniformidad de la distribución de las secuencias en los clones de un repertorio. “ImmuneREF” define un clon como las secuencias con una CDR3 AA idéntica. “ImmuneREF” calcula la Diversidad de Hill (Wever, 2022), una familia de índices de diversidad, para valores alfa de 0 a 10 en pasos de 0,1, donde alfa = 1 se define como uniformidad de Shannon. Cada entrada del perfil varía entre ≈ 0 y 1, y los valores más altos indican una distribución de la frecuencia clonal cada vez más uniforme. “ImmuneREF” determina los perfiles de uniformidad para cada repertorio y evalúa la similitud de uniformidad entre repertorios mediante la correlación de Pearson (Wever, 2022).

- Frecuencia de las posiciones de los aminoácidos:

“ImmuneREF” calcula para cada longitud de CDR3, la frecuencia de aminoácidos en cada posición. Para ello elegimos un rango de entre 8 y 20 aminoácidos para tener resultados interpretables. Esto da como resultado un vector de frecuencia por longitud de CDR3 por directorio que puede ser comparado con otro directorio utilizando el coeficiente de correlación de Pearson (Weber, 2022).

- Arquitectura de la red de similitud de secuencias:

“ImmuneREF” construye una red de similitud de secuencias para cada repertorio. Esta red se ve representada con nodos que nos indican las secuencias de los aminoácidos de las regiones variables de los receptores de antígenos. Los nodos están conectados por bordes de similitud. A diferencia de



muchas otras características, la característica de red combina varias medidas individuales, lo que dificulta el análisis de correlación de Pearson con las cuatro medidas de red estudiadas: la distribución de grados acumulativa, es decir, la fracción de vértices con un grado menor; la media de los nodos ($1 - |MeanHubScoeA - MeanHubScoreB|$), la fracción de clústeres y nodos no conectados y, por último, el porcentaje de secuencias en el mayor componente conectado ($1 - |PerclargestComponentsA - PerclargestComponentB|$) (Wever, 2022).

- Superposición del repertorio (convergencia):

Esta medida se realiza comparando solamente 2 repertorios. Muestra el solapamiento de la secuencia clonal ("ImmuneREF" considera como clon, la similitud del 100% de las secuencias de aminoácidos CDR3).

Superposición $(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)} \times 100$ donde X = repertorio X, Y = repertorio Y (Wever, 2022).

- Diversidad genética en la línea germinal:

La herramienta calcula la frecuencia relativa de los genes de la línea germinal (genes V-(D)-J) entre los clones de cada repertorio. "ImmuneREF" examina las frecuencias de los genes V y J entre los clones de cada individuo. Se determinó el coeficiente de correlación de Pearson para cada uno de los vectores de frecuencia. Los valores de correlación resultantes se combinan en un único valor calculando una media ponderada de estos componentes. El vector de pesos por defecto es $c(V = 1, D = 1, J = 1, VJ = 0)$. Esta medida capta el impacto potencial de un estado patológico (Wever, 2022).

- Aparición de "k-mers" separados:

Para esta característica, "immuneREF" calcula todas las subsecuencias con una longitud k (en nuestro caso $k = 3$), en una longitud máxima de hueco m ($m \leq 3$) de las secuencias de ADN. Los parámetros k y m se eligieron basándose en investigaciones anteriores (Weber et al., 2020), donde se demostró que establecer los parámetros $k = 3, m \leq 3$ conducía a una codificación suficiente para la clasificación de secuencias. La aparición de "k-mers" separados representa información de secuencia de corto y largo alcance codificada en el



repertorio. “ImmuneREF” también correlaciona las distribuciones resultantes entre pares de repertorios utilizando la correlación de Pearson.

9.1.1.2 *Cálculo de las similitudes*

Esta medida se basa en la comparación de dos repertorios. Se calculan las similitudes por cada característica (descritas anteriormente).

9.1.1.3 *Similitud de repertorio*

Una vez calculadas las similitudes por cada par de repertorios, condensamos esos datos en una red con las 6 características.

9.1.1.4 *Similitud del repertorio local*

Es una media para el grado de conexión de un repertorio dentro de una característica. También nos indica el nodo más y menos representativo en cada categoría.

9.1.1.5 *Simulación de los repertorios de receptores inmunitarios adaptativos*

Es necesario simular datos de repertorios para poder asociar datos a características y que nos sirvan de guía más adelante. Para ello, primero los autores del artículo (Wever, 2022) observaron un repertorio real a partir del cual, crearon un modelo viendo características en común.

9.1.2 “IMGT/StatClonotype”

Es un software que compara estadísticamente la diversidad o la expresión del clonotipo IMGT (los aminoácidos) entre dos conjuntos de resultados de IMGT/HighV-QUEST. Evalúa la significación estadística de las diferencias en las proporciones por variable, diversidad y unión de genes y alelos de un determinado locus de un receptor, utilizando siete procedimientos de pruebas múltiples para el ajuste de valores p (Bonferroni, Holm, Hochberg, ŠidákSS, ŠidákSD, Benjamini & Hochberg, Benjamini & Yekutieli): esto permite seleccionar la rigurosidad más relevante para el objetivo de un estudio (Aouinti et al., 2016; Aouinti et al., 2015).

9.1.2.1 *Selección de datos*

Para poderlo utilizar esta herramienta, necesitamos un formato de salida del análisis concreta, proporcionada por “IMGT/HighV-QUEST”. Concretamente el archivo se llama “stats_XXX” (siendo XXX el nombre del análisis hecho por “IMGT/HighV-QUEST”).



9.1.2.2 Selección de los parámetros de visualización

Hemos seleccionado una longitud de las regiones variables de los receptores entre 4 y 45 aminoácidos, de esta manera eliminamos los valores atípicos del procedimiento estadístico.

9.1.2.3 Resultados proporcionados

Una vez analizado los dos repertorios, la aplicación nos proporcionará las comparaciones de los genes y alelos de los dos repertorios. Primero nos muestra las pruebas estadísticas, seguido de diferentes gráficos, luego las longitudes y propiedades de las regiones hipervariables y para finalizar, las asociaciones de los genes VDJ.

- Pruebas estadísticas:

Nos muestra los diferentes clonotipos con las longitudes CDR3 seleccionadas. "IMGT clonotype (AA)" se define como un reordenamiento V-(D)-J único, anclajes CDR3 conservados y una secuencia de unión CDR3 única.

- Gráfico de síntesis

"IMGT/StatClonotype" puede generar la diversidad de clonotipos IMGT por genes. Va a crear un gráfico de síntesis que compara las proporciones normalizadas de diversidad de clonotipos (AA) IMGT de los genes IGHV en dos repertorios.

- Longitudes de CDR

Muestra gráficos de barras para el número de clonotipos en los dos repertorios por la longitud de CDR1, CDR2 o CDR3. Nosotros analizamos solo los CDR3 ya que es la región que sufre más mutaciones.

- Propiedades del CDR

Muestra la distribución de clases de los 20 aminoácidos comunes en una posición CDR específica en los 2 repertorios.

Nos dará los resultados como valores absolutos (número de apariciones de un aminoácido (o clase de aminoácido IMGT) en una posición determinada, para una longitud de CDR dada) (Pommié 2004).



También generará gráficos de variabilidad basados en la “entropía de Shannon”, el “índice de Simpson” y la “variabilidad de Wu-Kabat”.

- Asociaciones del gen VDJ

Muestra mapas de calor interactivos que representan asociaciones de genes VJ, VD o DJ en los dos repertorios. Nosotros vamos a analizar solo las asociaciones VJ ya que la región D es muy pequeña, entonces son datos insignificantes.

9.2 Datos de secuenciación de alto rendimiento

Para empezar a entender y posteriormente analizar los datos experimentales, primero analicé el artículo (Heikema et al., 2020) para ver resultados esperados y poder interpretar mis resultados. Luego explicaré el método de secuenciación del ADN para poder analizarlo y también el método de búsqueda y de descarga de los datos necesarios para el estudio.

9.2.1 Artículo sobre el estudio del COVID

Antes de empezar la parte experimental, se tuvieron que simular los datos para poder comparar posteriormente, los datos experimentales que obtuvieron.

La parte experimental del artículo se hizo para validar la sensibilidad de immuneREF. Este apartado se divide en tres partes; el estudio de la inmunización de ratones mediante los BCRs, el estudio de los TCR β humanos y firmas de citomegalovirus y finalmente, el estudio de los TCR de pacientes recuperados de casos leves de COVID-19.

9.2.1.1 *Simulación de datos*

En el artículo, simularon 200 conjuntos de repertorios inmunitarios con 40 combinaciones de parámetros. Para la simulación, usaron immuneSIM (paquete de R) (Wever, 2022). Cada repertorio contiene 12000 secuencias. Estas secuencias varían siempre en función del receptor, la línea de genes germinales, la especie, la distribución de los genes, entre otros. También añadieron pequeñas secuencias de nucleótidos en posiciones aleatorias de la región CDR3 para modificar el subconjunto de repertorios. También produjeron eliminaciones de secuencias para variar la arquitectura de la red de similitudes.



Con el experimento de pacientes con COVID, pasa igual; el estudio reveló grupos de mayor similitud dentro de los pacientes y las poblaciones celulares. (Wever, 2022).

9.2.1.2 *Estudio de inmunización de ratones de BCRs y estudios de los repertorios de TCR de pacientes recuperados de casos leves de COVID-19*

La muestra la extrajeron de células B de diferentes tejidos; del bazo y de la médula ósea. Hicieron una secuenciación de alto rendimiento de ARN con MiXCR (Heikema et al., 2020).

Los resultados esperados eran ver diferencias en la similitud con los grupos con inmunización con antígenos. Pudieron ver diferencias entre isotipos y poblaciones de células que no habían entrado en contacto con el antígeno (muestras del bazo). También observaron que los grupos de inmunización con antígeno (muestra de la médula ósea) tienen perfiles de similitud con más diferencias en las poblaciones de células plasmáticas (IgG) que con las poblaciones de células sin experiencia con el antígeno.

Con la parte experimental de pacientes con COVID leve, se demostró que hay grupos con mayor similitud dentro de los pacientes y las poblaciones celulares.

9.2.1.3 *Estudio de repertorios de TCR β humanos y firmas de citomegalovirus*

En este estudio evaluaron 1522 repertorios de TCR humanos derivados de células mononucleares de la sangre periférica (PBMC) de diferentes pacientes con estados inmunitarios diversos. Las muestras fueron de pacientes sanos, con lupus eritematoso sistémico (LES) y pacientes con artritis reumatoide (AR).

Se encontraron con una similitud global alta y uniforme en los rasgos individuales.

Al calcular la dispersión, vieron que los grupos de pacientes sanos y AR tienen un rasgo más restringido de puntuaciones de similitud en comparación de LES, que tiene un rango más amplio (Wever, 2022).

Las muestras sufrieron una secuenciación de alto rendimiento basada en ADN.

9.2.2 Método de secuenciación

Los datos utilizados se secuenciaron por síntesis “illumina”. Este tipo de secuenciación se divide en la preparación de una librería, la generación de los



“clusters” y amplificación en puente, secuenciación por síntesis y análisis de datos.

9.2.2.1 *Preparación de una librería*

Han hecho la extracción y purificación del ADN a partir de la muestra. Fraccionamos la muestra con la ayuda de transposasas que cortarán el ADN de manera aleatoria para formar pequeños fragmentos. Posteriormente, se produce la tagmentación, que es la adición de secuencias adaptadoras.

El siguiente paso, es la amplificación cíclica reducida. Añadieron “primers” afines a las secuencias adaptadoras.

De esta manera ya se ha creado la librería y posteriormente, tienen que hacer la generación de “clusters”.

9.2.2.2 *Generación de “clusters”*

Este proceso se realiza dentro de una celda de flujo. La celda está formada por un vidrio con carriles. Cada carril está recubierto por dos oligonucleótidos diferentes. Uno de los oligonucleótidos es complementario a la región 5' de los motivos adicionales y el otro oligonucleótido, es complementario a la región 3'.

La biblioteca preparada anteriormente, se desnatura y la cargan en la celda de flujo. Cada fragmento de la biblioteca se amplificará de forma isotérmica.

Ahora los oligonucleótidos se van a unir a las regiones correspondientes de ADN de nuestra biblioteca de secuenciación monocatenaria. En el momento de unión, actuará la ADN polimerasa. Como resultado obtendrán un ADN de doble cadena que se desnatura; luego lo lavan para eliminar la cadena original.

Ahora se dará la amplificación clonal a través de amplificación de puente.

9.2.2.3 *Amplificación en puente*

La hebra formada anteriormente, se doblará haciendo que la segunda secuencia complementaria se una al segundo tipo de oligonucleótido. Otra vez un ADN polimerasa actuará. Por lo tanto, tenemos una estructura bicatenaria que se desnaturará obteniendo en el proceso, obteniendo dos tipos de hebras la “forward” y la “reverse”. Ahora se lavan las hebras “reverse”. De esta manera quedan las hebras “forward” con el extremo 3' libre y bloqueado.



9.2.2.4 *Secuenciación*

El primer paso es la extensión del “primer” de secuenciación. En las muestras que hemos analizado, las secuenciaron mediante el método “illumina” (Heikema et al., 2020). Este método utiliza desoxirribonucleótidos (dNTP) modificados, con un terminador en la posición 3’ de la desoxirribosa. Otra característica es que los dNTP están marcados con fluorescencia. Por lo tanto, después de cada adición de nucleótido complementario, se emitirá fluorescencia. Después de la detección de luz, el fluoróforo se lava y se añade un segundo dNTP ya que, el grupo terminador 3’ de la desoxirribosa se puede regenerar con un grupo hidroxilo. Este proceso se llama secuenciación por síntesis. Cada nucleótido tiene un color asignado, de esta manera pueden reconocerlo.

Después de terminar este proceso, hace lo mismo con la secuencia “reverse”.

Cuando se finaliza la secuenciación, procedemos al análisis de datos.

9.2.3 Buscar y descargar datos

Los datos utilizados para el análisis con “immuneREF” son “healthy 3” (H3), “healthy 4” (H4), COVID moderado 5 (M5), COVID moderado 6 (M6), COVID severo (S24) y COVID severo (S26). Estos datos están disponibles en “National Institute of Health’s Sequence Read Archive”. El código de acceso para acceder al artículo es: PRJNA630455. Con este código llegamos al artículo (Kuri-Cervantes et al., 2020).

Para la selección de datos, tenemos que fijarnos en que todos tengan las mismas condiciones de edad, sexo y que la muestra provenga del mismo tejido. En nuestro caso, el artículo (Kuri-Cervantes et al., 2020) no lo tuvo en cuenta.

Una vez elegidos los datos, vamos a descargarlos. Para ello hay dos formas, una con la línea de comando y la otra de forma manual.

Para la línea de comando, tenemos que descargar e instalar “conda sra tolos”. Una vez iniciemos el programa y descarguemos los datos, los obtenemos en formato “FASTQ”.

Para descargarlo de forma manual, una vez copiado el código de acceso, le damos a descargar en formato “FASTQ”.



Finalmente pasamos los datos en formato “FASTA” y los podemos analizar en “IMGT/HighV-QUEST”. Aquí seleccionamos el formato en el que queremos el análisis. Para comparar los datos por “immuneREF”, elegiremos el formato “AIRR”. En cambio, para comparar a través de “IMGT/StatClonotype”, necesitamos el formato “CSV”.

A continuación, veremos el repertorio con su identificador:

En el caso del repertorio normal o “healthy” nos encontramos con H3: SRS6898286 y H4: SRS6898236. En el caso de los repertorios normales, no tenemos los datos ni del sexo ni de la edad del donante.

Los repertorios patológicos (con COVID) los dividimos en dos grupos; el grupo con síntomas moderados M5: SRS6898319 y M6: SRS6898312, y el grupo con síntomas severos; S24: SRS6898299 y S26: SRS6898292. En el caso de M5, el donante tenía entre 25 y 30 años, en cambio, el donante de M6, tiene entre 61-65 años. Los datos proporcionados para COVID severo nos indican que S24 tiene entre 51-55 años y S26 entre 71 y 75 años. A pesar de que no se respetó el parámetro de que la edad y el sexo de todos los repertorios sea igual, sí que se respetó que la muestra de origen sea la misma. Todas las muestras fueron extraídas de la sangre.

10. RESULTADOS Y DISCUSIÓN

Primero voy a analizar los resultados individuales de los repertorios normales, y después de uno de los dos repertorios de cada grupo, es decir, un repertorio con COVID normal, y otro con COVID severo puesto que los resultados son muy parecidos, entonces las diferencias no son muy significativas.

10.1 Resultados repertorio H3

Primero analizamos la longitud y propiedades de CDR3 para el repertorio normal H3 y luego las asociaciones de los genes V-J

- Longitud de CDR3

En esta gráfica podemos ver el número de clontipos que encontramos dependiendo de la talla de CDR3 en el repertorio normal H3. Podemos ver una frecuencia de más de 6000 clontipos con una longitud de 15 nucleótidos.

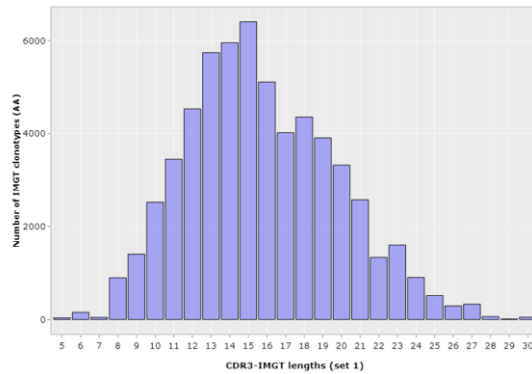


Figura 4: Gráfica de barras proporcionada por “IMGT/StatClonotype”. Muestra el número de clonotipos (AA) a lo largo de la secuencia CDR3 de H3.

- Propiedades de CDR3

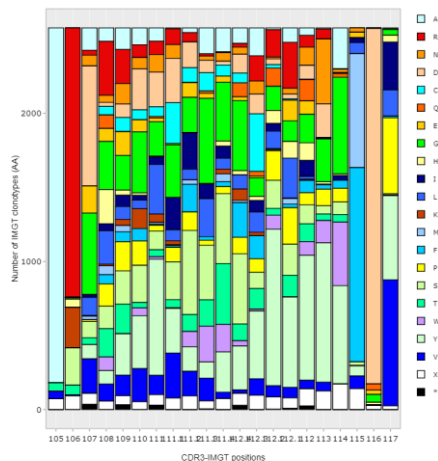


Figura 5: “Heatmap” de la asociación de los genes V con J del repertorio H3. Proporcionado por “IMGT/StatClonotype”.

En la figura 5, podemos observar cuantos aminoácidos hay en las primeras 21 posiciones de la secuencia de CDR3 en H3 y, además, podemos ver el nombre de cada aminoácido. Hemos escogido 21 aminoácidos ya que, si nos fijamos en la longitud de CDR3, a partir de secuencias de 22 AA, tenemos una baja frecuencia de clonotipos AA.

Figura 6: Esta gráfica nos muestra los valores de la entropía de Shanon en las primeras 21 posiciones de CDR3 en H3. La gráfica la he extraído de “IMGT/StatClonotype”.

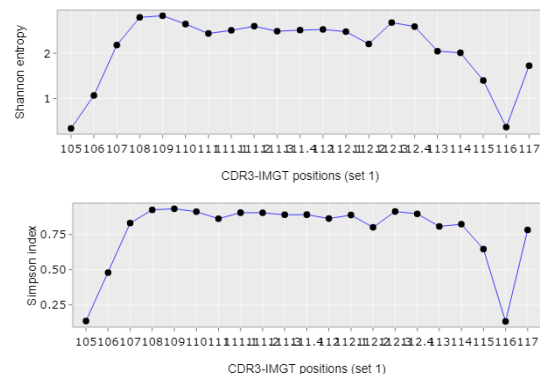


Figura 7: Esta gráfica nos muestra los valores del índice de Simpson en las primeras 21 posiciones de CDR3 en H3. Sacado de “IMGT/StatClonotype”.

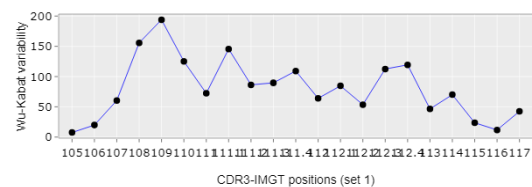


Figura 8: Esta grafica creada por “IMGT/StatClonotype” nos muestra la variabilidad de Wu-Kabat para las primeras 21 posiciones de CDR3 en H3.

En la figura 6, vemos la entropía de Shanon, esta entropía nos indica la menor cantidad de información necesaria para eliminar su incertidumbre. Si nos fijamos, tanto en la posición 105 y 116, el valor de Shanon es muy cercano a 0. Esto se debe a que, en esas posiciones, encontramos muy pocos aminoácidos y uno de ellos en una gran cantidad (Figura 5).

La figura 7, es el índice de Simpson, sirve para medir la diversidad de una comunidad. Si nos fijamos, la gráfica es muy parecida al índice de Shanon ya que esta correlacionada también con la diversidad de aminoácidos, cuantos menos aminoácidos hay en una posición, más pequeño es el valor de Simpson.

La figura 8, es el índice entre 0 y 200 que hemos obtenido al dividir el número de aminoácidos que encontramos en cada posición de la secuencia CDR3 en el repertorio H3 entre la frecuencia del aminoácido más común en esa posición. Por lo tanto, nos dice el grado de variabilidad, podemos ver que en la posición 109, es donde hay una mayor variabilidad de aminoácidos.

- Asociación de los genes V-J

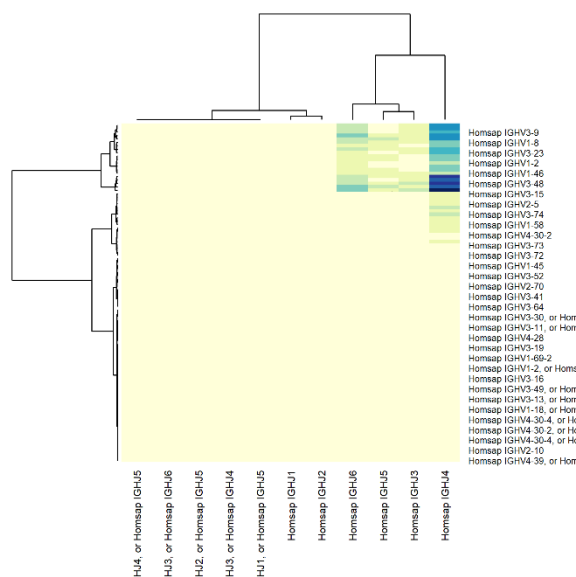


Figura 9: “Heatmap” de la asociación de los genes V con J del repertorio H3. Proporcionado por “IMGT/StatClonotype”.

En la figura 9, vemos representado en forma de “heatmap” la frecuencia de asociaciones entre los genes V y J de las regiones hipervariables. Cuando el color es más oscuro, significa que hay un mayor número de genes recombinados. Podemos observar que la mayoría de los genes V se asocian con IGHJ4. Y de todos los genes V, el que tiene más probabilidades de unirse con IGHJ4 es IGHV3-48.

10.2 Resultados repertorio H4

Ahora vamos a calcular lo mismo para el otro repertorio normal para poder compararlos puesto que queremos definir el repertorio inmunitario no patológico.

- Longitud de CDR3

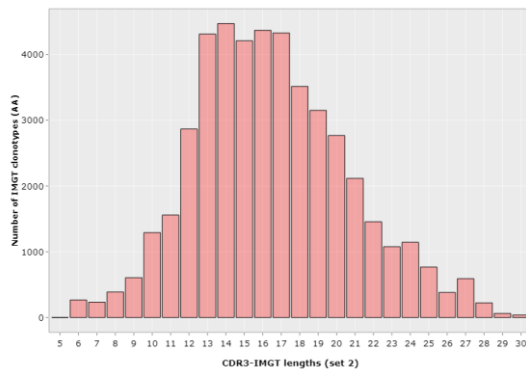


Figura 10: Gráfica de barras proporcionada por "IMGT/StatClonotype". Muestra el número de clonotipos (AA) a lo largo de la secuencia CDR3 de H4.

longitud de entre 12 y 17 aminoácidos. El mayor número de frecuencia en este caso es acerca de los 4000, un valor inferior que en el repertorio H3 que tiene un valor de 6000.

- Propiedades de CDR3

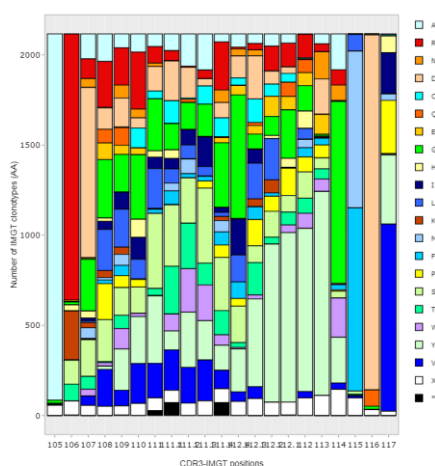


Figura 11: Proporcionado por "IMGT/StatClonotype". Muestra el número de aminoácidos y sus nombres en las primeras 21 posiciones de CDR3 en H4.

En esta gráfica podemos ver el número de clonotipos que encontramos dependiendo de la talla de CDR3 en el repertorio normal H4. Si lo comparamos con la figura 4, podemos observar que tienen una forma parecida, puesto que ambas son repertorios no patológicos. Podemos ver una frecuencia más elevada de clonotipos con una

Figura 12: Esta gráfica nos muestra los valores de la entropía de Shanon en las primeras 21 posiciones de CDR3 en H4. La gráfica la he extraído de "IMGT/StatClonotype"

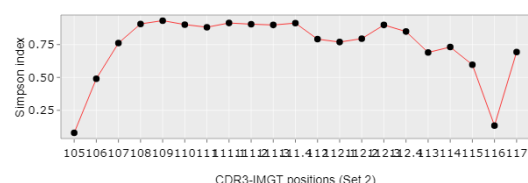
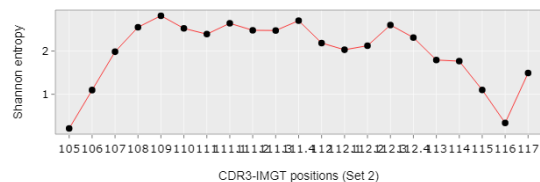


Figura 13: Esta gráfica nos muestra los valores del índice de Simpson en las primeras 21 posiciones de CDR3 en H4. Sacado de "IMGT/StatClonotype"

En la figura 11, podemos observar cuantos aminoácidos hay en las primeras 21 posiciones de la secuencia de CDR3 en H4 y, además, podemos ver el nombre de cada aminoácido. Hemos escogido 21 aminoácidos ya que, si nos fijamos en la longitud de CDR3, a partir de secuencias de 22 AA, tenemos una baja frecuencia de clonotipos AA.

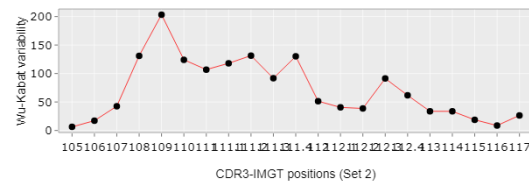


Figura 14: Esta grafica creada por “IMGT/StatClonotype” nos muestra la variabilidad de Wu-Kabat para las primeras 21 posiciones de CDR3 en H4.

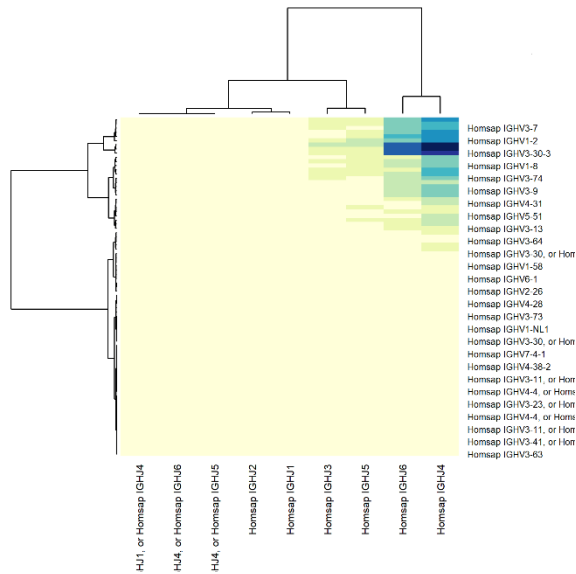
Si comparamos esta figura con la figura 5, podemos ver que la estructura es muy parecida puesto que ambos repertorios son sanos respecto al COVID.

En la figura 12, vemos la entropía de Shanon, esta entropía nos indica la menor cantidad de información necesaria para eliminar su incertidumbre. Si nos fijamos, tanto en la posición 105 y 116, el valor de Shanon es muy cercano a 0. Esto se debe a que, en esas posiciones, encontramos muy pocos aminoácidos y uno de ellos en una gran cantidad (Figura 12).

La figura 13, es el índice de Simpson, sirve para medir la diversidad de una comunidad. Si nos fijamos, la gráfica es muy parecida al índice de Shanon ya que esta correlacionada también con la diversidad de aminoácidos, cuantos menos aminoácidos hay en una posición, más pequeño es el valor de Simpson. La figura 14, es el índice entre 0 y 200 que hemos obtenido al dividir el número de aminoácidos que encontramos en cada posición de la secuencia CDR3 en el repertorio H4 entre la frecuencia del aminoácido más común en esa posición. Por lo tanto, nos dice el grado de variabilidad, podemos ver que en la posición 109, es donde hay una mayor variabilidad de aminoácidos igual que en la secuencia de H3 y la forma de la gráfica, también es muy similar.

Asociación de los genes V-J

En la figura 15, vemos representado en forma de “heatmap” la frecuencia de asociaciones entre los genes V y J de las regiones hipervariables. Cuando el color es más oscuro, significa que hay un mayor número de genes



recombinados. Podemos observar que la mayoría de los genes V se asocia con IGHJ4 e IGHJ6. Y de todos los genes V, el que tiene más probabilidades de unirse con IGHJ4 e IGHJ6 es IGHV3-30-3.

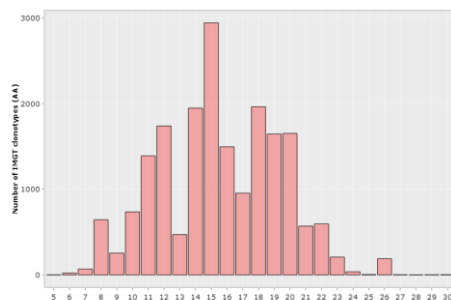
Figura 6: “Heatmap” de la asociación de los genes V con J del repertorio H4. Proporcionado por “IMGT/StatClonotype”.

10.3 Resultados repertorio M5

Ahora vamos a ver los resultados inmunitarios del repertorio patológico M5 para compararlos con el normal y ver si hay diferencias significativas.

- Longitud de CDR3

En esta gráfica podemos ver el número de clontipos que encontramos dependiendo de la talla de CDR3 en el repertorio con COVID normal M5. Si lo comparamos con la figura 4 o 10, podemos observar que tienen una forma muy diferente. En los repertorios normales tenemos una forma de crecimiento continuo y luego, viene el decrecimiento; en cambio aquí la forma no sigue ningún tipo de continuidad. Podemos ver una frecuencia más elevada de clontipos con una longitud de 15 nucleótidos.



También observamos que el mayor número de clontipos AA no llega a 3000, en cambio en el repertorio H3, el valor máximo está cerca de los 6000, es decir el doble.

Figura 7: Gráfica de barras proporcionada por “IMGT/StatClonotype”. Muestra el número de clontipos (AA) a lo largo de la secuencia CDR3 de M5.

- Propiedades de CDR3

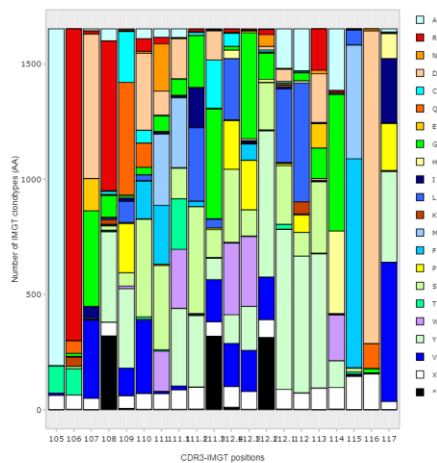


Figura 17: Proporcionado por "IMGT/StatClonotype". Muestra el número de aminoácidos y sus nombres en las primeras 20 posiciones de CDR3 en M5.

En la figura 17, podemos observar cuantos aminoácidos hay en las primeras 20 posiciones de la secuencia de CDR3 en M5 y, además, podemos ver el nombre de cada aminoácido. Si comparamos esta figura con la figura 5 o 12, podemos ver que la estructura difiere bastante puesto que estamos hablando de un repertorio patológico. Solemos encontrar los mismos AA en cada posición, pero con una frecuencia muy diferente.

En la figura 17, vemos la entropía de Shanon, esta entropía nos indica la menor cantidad de información necesaria para eliminar su incertidumbre. Si nos fijamos, tanto en la posición 105 y 116, el valor de Shanon es muy cercano a 0. Esto se debe a que, en esas posiciones, encontramos muy pocos aminoácidos y uno de ellos en una gran cantidad.

La figura 19, es el índice de Simpson, sirve para medir la diversidad de una comunidad. Si nos fijamos, la gráfica es muy parecida al índice de Shanon ya

Figura 8: Esta gráfica nos muestra los valores de la entropía de Shanon en las primeras 20 posiciones de CDR3 en M5. La gráfica la he extraído de "IMGT/StatClonotype"

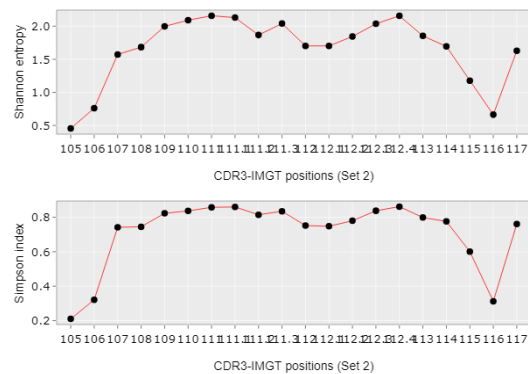


Figura 9: Esta gráfica nos muestra los valores del índice de Simpson en las primeras 20 posiciones de CDR3 en M5. Sacado de "IMGT/StatClonotype".

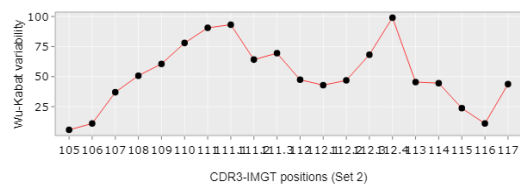


Figura 20: Esta grafica creada por "IMGT/StatClonotype" nos muestra la variabilidad de Wu-Kabat para las primeras 20 posiciones de CDR3 en M5.

que esta correlacionada también con la diversidad de aminoácidos, cuantos menos aminoácidos hay en una posición, más pequeño es el valor de Simpson.

La figura 20 es el índice entre 0 y 100 que hemos obtenido al dividir el número de aminoácidos que encontramos en cada posición de la secuencia CDR3 en el repertorio M5 entre la frecuencia del aminoácido más común en esa posición. Por lo tanto, nos dice el grado de variabilidad, podemos ver que en la posición 112, es donde hay una mayor variabilidad de aminoácidos.

- Asociación de los genes V-J

En la figura 21, vemos representado en forma de “heatmap” la frecuencia de asociaciones entre los genes V y J de las regiones hipervariables. Cuando el color es más oscuro, significa que hay un mayor número de genes recombinados. Podemos observar que la mayoría de los genes V se asocian con IGHJ4. Y de todos los genes V, el que tiene más probabilidades de unirse con IGHJ4 es IGHV3-48. Esta frecuencia de asociación es muy similar a la del repertorio H3, a diferencia, podemos observar que hay más frecuencia de asociación con más genes J en este repertorio que en los repertorios normales.

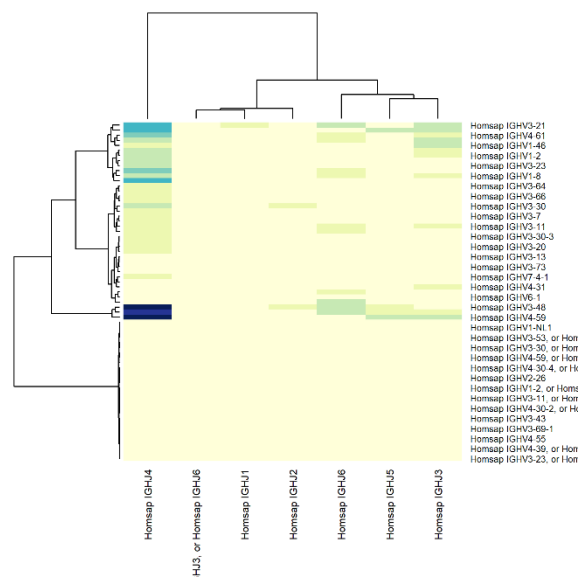


Figura 21: “Heatmap” de la asociación de los genes V con J del repertorio M5. Proporcionado por “IMGT/StatClonotype”.

10.4 Resultados repertorio S26

Por último, dentro de análisis individual, analizamos el repertorio patológico severo S26 para ver las diferencias entre normal y también con el repertorio patológico no severo.

- Longitud de CDR3

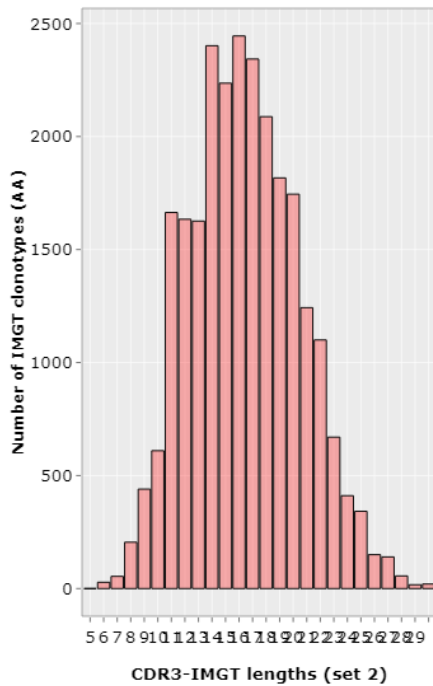


Figura 22: Gráfica de barras proporcionada por “IMGT/StatClonotype”. Muestra el número de clonotipos (AA) a lo largo de la secuencia CDR3 de S26.

En esta gráfica podemos ver el número de clonotipos que encontramos dependiendo de la talla de CDR3 en el repertorio con COVID severo S26. Si lo comparamos con la figura 4 o 10, podemos observar que tienen una forma bastante similar, ya que hay una pendiente creciente continua y luego una decreciente. Podemos ver una frecuencia más elevada entre secuencias de CDR3 compuestas entre 14 y 18 nucleótidos. A partir de secuencias con 20 nucleótidos, los valores son muy pequeños, por esa razón, analizaremos hasta la posición 20.

es el valor máximo de número de clonotipos de AA más pequeño.

También observamos que el mayor número de clonotipos AA no llega a 2500,

- Propiedades de CDR3

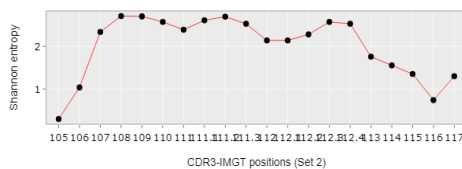


Figura 23: Esta gráfica nos muestra los valores de la entropía de Shanon en las primeras 20 posiciones de CDR3 en S26. La gráfica la he extraído de “IMGT/StatClonotype”.

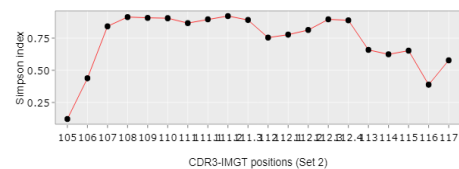


Figura 24: Esta gráfica nos muestra los valores del índice de Simpson en las primeras 20 posiciones de CDR3 en S26. Sacado de “IMGT/StatClonotype”.

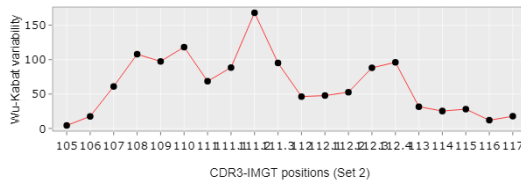


Figura 25: Esta grafica creada por "IMGT/StatClonotype" nos muestra la variabilidad de Wu-Kabat para las primeras 20 posiciones de CDR3 en S26.

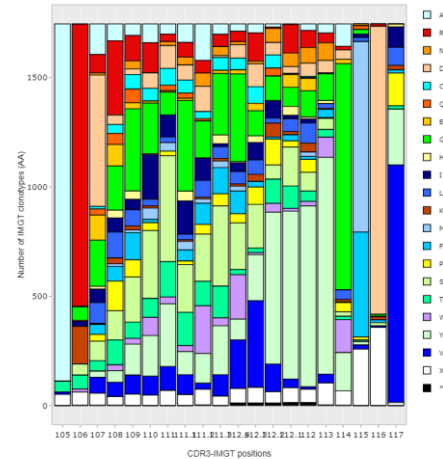


Figura 26: Proporcionado por "IMGT/StatClonotype". Muestra el número de aminoácidos y sus nombres en las primeras 20 posiciones de CDR3 en S26

En la figura 26, podemos observar cuantos aminoácidos hay en las primeras 20 posiciones de la secuencia de CDR3 en S26 y, además, podemos ver el nombre de cada

aminoácido. Si comparamos esta figura con la figura 5 o 12, podemos ver que la estructura difiere bastante puesto que estamos hablando de un repertorio patológico. Solemos encontrar los mismos AA en cada posición, pero con una frecuencia muy diferente. Si lo comparamos con la figura 18, también encontramos diferencias en el número de clonotipos puesto que este repertorio pertenece a un caso grave de COVID.

En la figura 23, vemos la entropía de Shanon, esta entropía nos indica la menor cantidad de información necesaria para eliminar su incertidumbre. Si nos fijamos, tanto en la posición 105 y 116, el valor de Shanon es muy cercano a 0. Esto se debe a que, en esas posiciones, encontramos muy pocos aminoácidos y uno de ellos en una gran cantidad.

La figura 24, es el índice de Simpson, sirve para medir la diversidad de una comunidad. Si nos fijamos, la gráfica es muy parecida al índice de Shanon ya que esta correlacionada también con la diversidad de aminoácidos, cuantos menos aminoácidos hay en una posición, más pequeño es el valor de Simpson.

La figura 25 es el índice entre 0 y 150 que hemos obtenido al dividir el número de aminoácidos que encontramos en cada posición de la secuencia CDR3 en el repertorio S26 entre la frecuencia del aminoácido más común en esa posición. Por lo tanto, nos dice el grado de variabilidad, podemos ver que en la posición 113, es donde hay una mayor variabilidad de aminoácidos.

- Asociación de los genes V-J

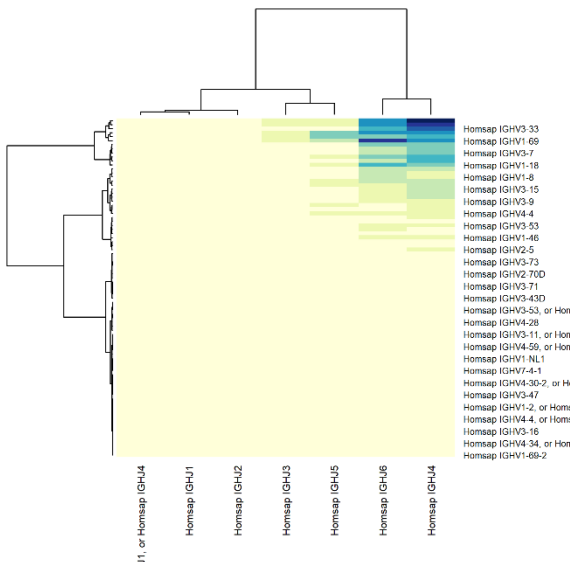


Figura 27: "Heatmap" de la asociación de los genes V con J del repertorio S26. Proporcionado por "IMGT/StatClonotype".

En la figura 27, vemos representado en forma de "heatmap" la frecuencia de asociaciones entre los genes V y J de las regiones hipervariables. Cuando el color es más oscuro, significa que hay un mayor número de genes recombinados. Podemos observar que la mayoría de los genes V se asocian con IGHJ4 e IGHJ6. Y de todos los genes V, el que tiene más probabilidades de unirse con IGHJ4 es IGHV3-33; y el que tiene más probabilidades de asociarse con

IGHJ6 es IGHV1-60. Esta frecuencia de asociación es muy similar a la de los repertorios normales.

10.5 Comparación H3 con H4

Ahora voy a comparar los dos repertorios normales para ver si hay diferencias significativas.

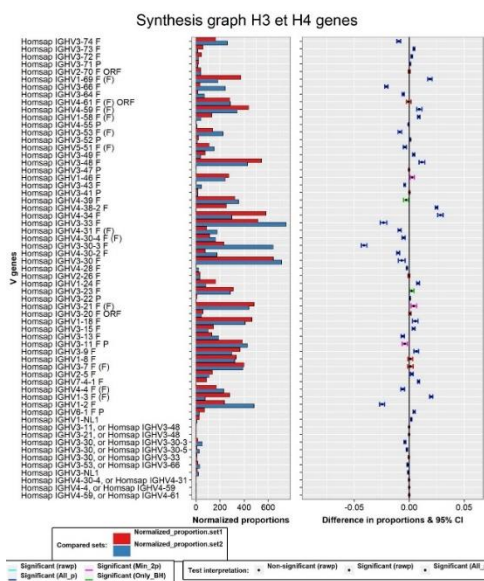


Figura 28: Gráfica comparativa de la frecuencia de todos los genes V entre H3 y H4. A la derecha nos muestra el intervalo de confianza de los genes.

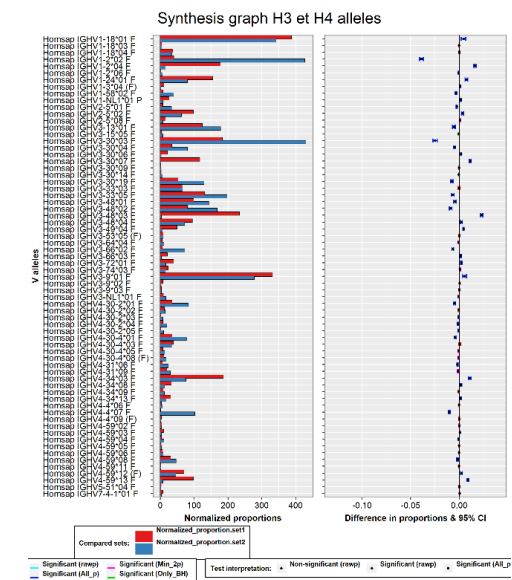


Figura 29: Gráfica comparativa de la frecuencia de todos los alelos V entre H3 y H4. A la derecha nos muestra el intervalo de confianza de los alelos.

En la figura 28 y 29, nos fijaremos en los valores en la gráfica del intervalo de confianza tanto de los genes como de los alelos. Estas gráficas nos indican las diferencias de un gen determinado en dos directorios. Las líneas rojas no son significantes. Tanto en la figura 28 como en la 29, vemos una predominancia de líneas azules, esto nos indica las diferencias significativas.

Aquí nos vamos a fijar en la última gráfica a la derecha que nos dice “evenness”. Aquí observamos la uniformidad respecto a la talla de los clones. Es decir que cuanto más rápido disminuya la curva, significa que más diferentes son los diferentes grupos de clones, respecto a su talla. Podemos ver que los grupos de clones de H3, tienen una talla más parecida entre ellos que en H4. Por eso, vamos a coger como referencia H3.

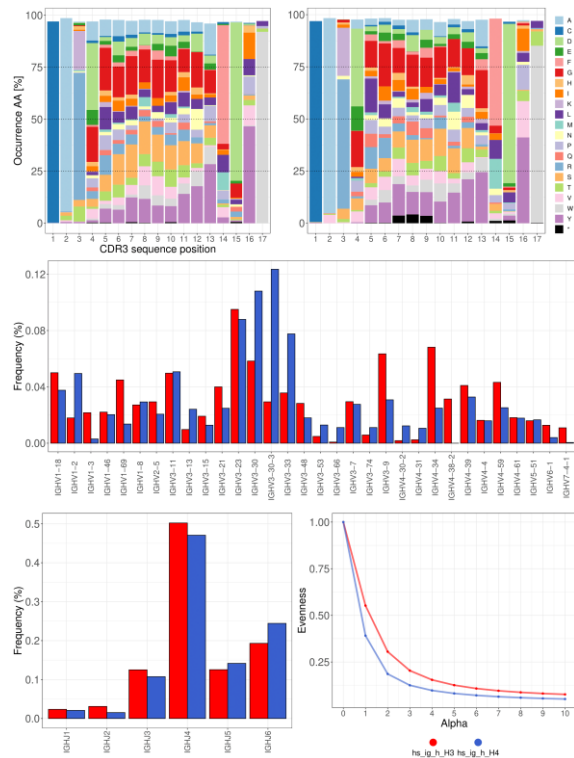


Figura 30: Datos proporcionados por “ImmuneREF”. Las dos tablas de arriba nos muestran la frecuencia y nombre de aminoácido en las primeras 17 posiciones de la secuencia CDR3 para los repertorios H3 y H4. En medio vemos la comparación de genes V y abajo a la izquierda para los genes J. Por último, tenemos la gráfica de uniformidad de clones respecto a la talla.

10.6 Comparación H3 con M5

Ahora comparamos el repertorio

normal que hemos elegido, con un repertorio patológico con síntomas normales de COVID.

En la figura 31 y 32, nos fijaremos en los valores en la gráfica del intervalo de confianza tanto de los genes como de los alelos. Estas gráficas nos indican las diferencias de un gen determinado en dos directorios. Las líneas rojas no son significantes. Tanto en la figura 31 como en la 32, vemos una predominancia de líneas azules, esto nos indica las diferencias significativas. Si comparamos

estas gráficas con la comparación de H3 y H4, podemos encontrar en estas más valores significativos que comparando solo los repertorios normales

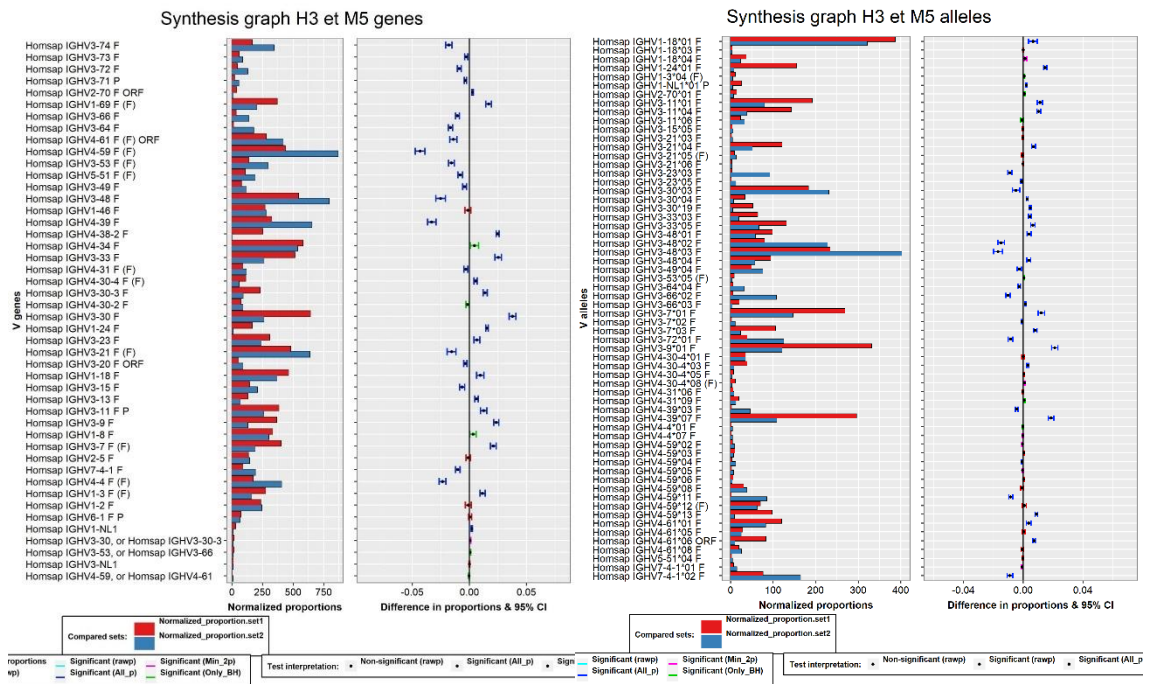


Figura 31: Gráfica comparativa de la frecuencia de todos los genes V entre H3 y M5. A la derecha nos muestra el intervalo de confianza de los genes.

Figura 32: Gráfica comparativa de la frecuencia de todos los alelos V entre H3 y M5. A la derecha nos muestra el intervalo de confianza de los alelos.

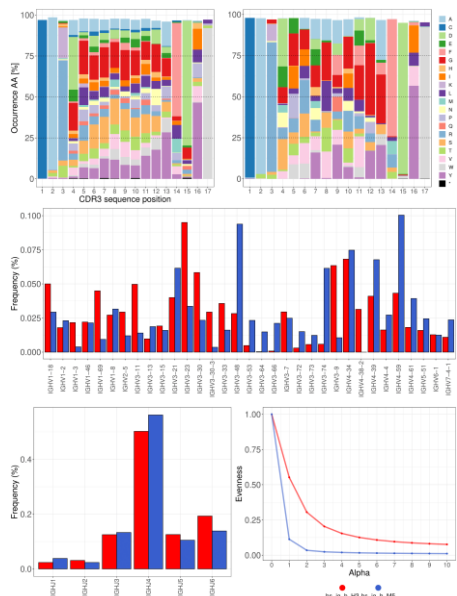


Figura 33: Datos proporcionados por "ImmuneREF". Las dos tablas de arriba nos muestran la frecuencia y nombre de aminoácido en las primeras 17 posiciones de la secuencia CDR3 para los repertorios H3 y M5. En medio vemos la comparación de genes V y abajo a la izquierda para los genes J. Por último, tenemos la gráfica de uniformidad de clones respecto a la talla.

En la figura 33, la parte inferior derecha, vemos que para un valor de alfa = 1, hay una gran diferencia, eso quiere decir que las tallas para los diferentes grupos de clones en M5 es muy diferente que para H3.

10.7 Comparación H3 con S26

Por último, vamos a comparar los análisis del repertorio normal con el repertorio patológico que corresponde a COVID severo.

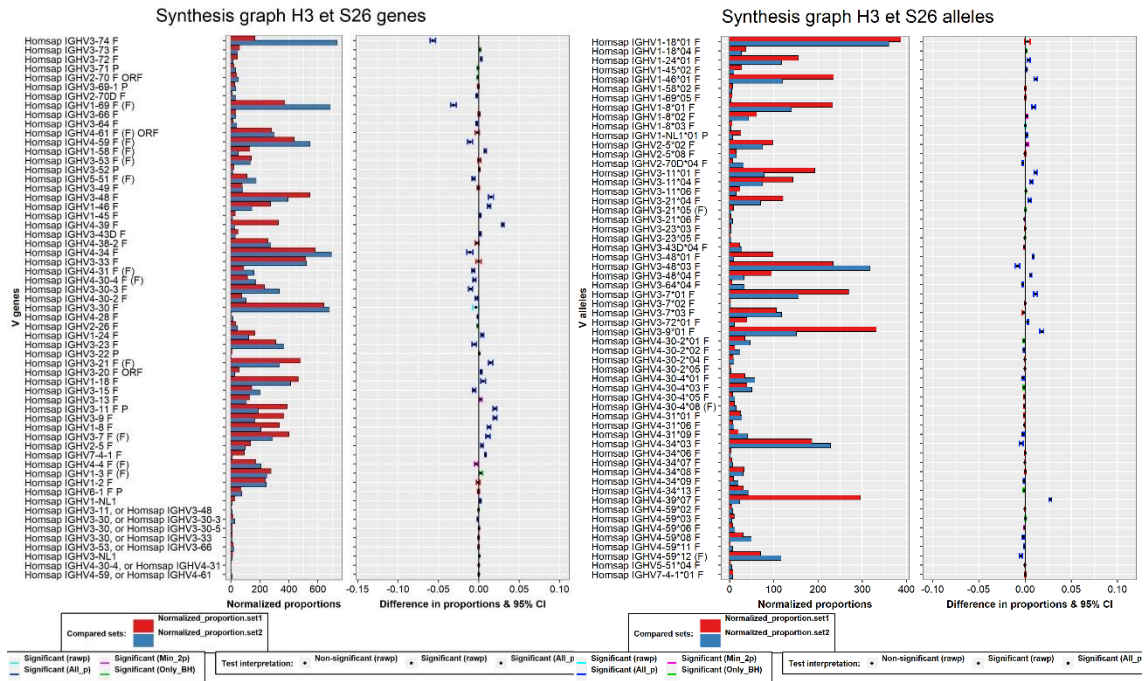


Figura 34: Gráfica comparativa de la frecuencia de todos los genes V entre H3 y S26. A la derecha nos muestra el intervalo de confianza de los genes.

Figura 35: Gráfica comparativa de la frecuencia de todos los alelos V entre H3 y S26. A la derecha nos muestra el intervalo de confianza de los alelos.

En la figura 34 y 35, nos fijaremos en los valores en la gráfica del intervalo de confianza tanto de los genes como de los alelos. Estas gráficas nos indican las diferencias de un gen determinado en dos directorios. Las líneas rojas no son significativas. Tanto en la figura 34 como en la 35, vemos una predominancia de líneas azules, esto nos indica las diferencias significativas. Si comparamos estas gráficas con las del repertorio normal, vemos que hay menos resultados significativos en cambio para COVID normal; hay más resultados significativos que para COVID severo.

En la figura 36, la parte inferior derecha, vemos que para un valor de $\alpha = 2$, hay una gran diferencia, eso quiere decir que las tallas para los diferentes grupos de clones en S26 es muy diferente que para H3.

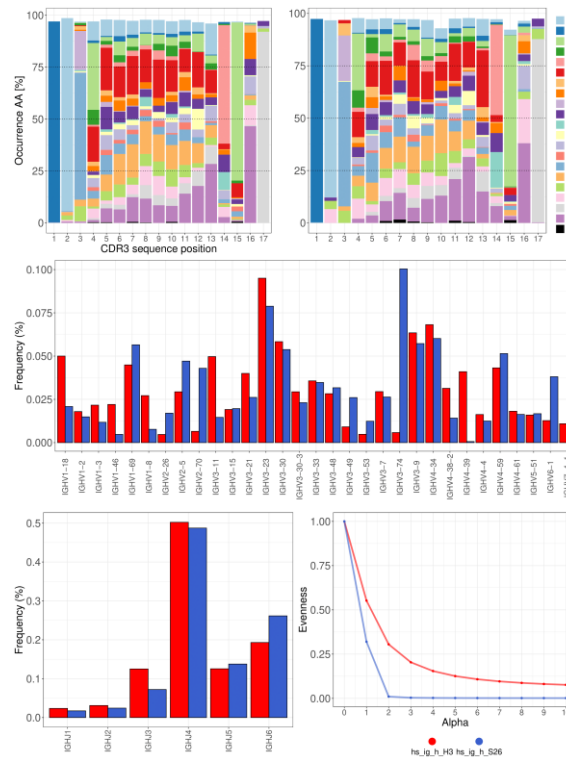


Figura 36: Datos proporcionados por “ImmuneREF”. Las dos tablas de arriba nos muestran la frecuencia y nombre de aminoácido en las primeras 17 posiciones de la secuencia CDR3 para los repertorios H3 y S26. En medio vemos la comparación de genes V y abajo a la izquierda para los genes J. Por último, tenemos la gráfica de uniformidad de clones respecto a la talla.

10.8 Comparación de todos los repertorios

Ahora vamos a hacer una comparación de las 6 características que definen un repertorio inmunitario, escogiendo los repertorios sanos como referencia.

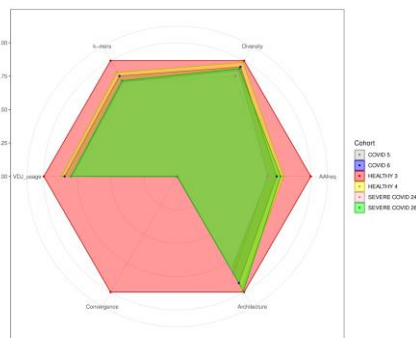


Figura 37: “Radar plot” proporcionado por “ImmuneREF”. Nos asigna un valor de similitud para cada característica y cada repertorio tomando como referencia H3, que le damos un valor de 1.

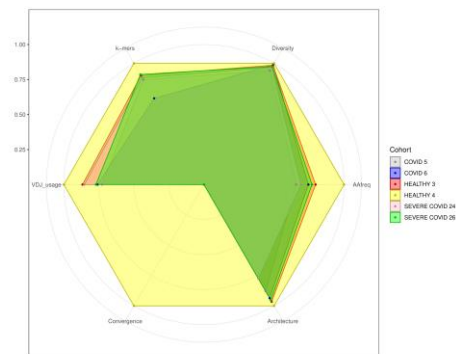


Figura 38: “Radar plot” proporcionado por “ImmuneREF”. Nos asigna un valor de similitud para cada característica y cada repertorio tomando como referencia H4, que le damos un valor de 1.



Hemos creado dos comparaciones de estas características tomando como referencia los directorios H3 y H4. La comparación de la convergencia es siempre 0 porque esta característica se compara sólo para 2 repertorios y aquí tenemos los 6.

En el gráfico de la izquierda, los repertorios "sanos" tienen en casi todas las características, la mayor similitud excepto en la arquitectura cuando el sano 3 es la referencia, ya que el COVID severo (S26), tiene aún más similitud en esta área, pero, por otro lado (S24), es el otro directorio con COVID severo, tiene la arquitectura más diferente, por lo que tiene muchas diferencias con S26. Luego, en todos los demás dominios, los repertorios COVID normales (M5 y M6), tienen casi los mismos valores excepto en arquitectura y diversidad; en diversidad, M5 es el repertorio más diferente en comparación con el H3. Así, en los dominios VDJ y "k-mers", el orden de similitud es el de los sanos, seguido del COVID moderado y después del COVID grave. La frecuencia de las posiciones de los aminoácidos es la característica con más diferencias dado que, H4 está justo por debajo de 0,75; seguido de la COVID severa (S26), después de la COVID normal (M6), la COVID severa (S24) y la última COVID normal (M5).

Ahora el gráfico de la derecha, tomamos como referencia el directorio H4 (saludable 4). En comparación con el otro repertorio normal, tienen los valores más similares, excepto en los k-mers, que el COVID severo (S26) tiene aún más similitud y (S24) está justo detrás de (H3); y en la frecuencia de posiciones de aminoácidos donde (H3) tiene la mayor similitud pero, está justo por debajo de la posición de similitud de 0,75. En los genes VDJ, después del repertorio sano (H3), el COVID severo tiene más similitudes que el COVID normal y el repertorio (M5) está por debajo de 0,75. En la arquitectura, los valores más parecidos a la referencia después de H3, es el COVID severo (S26), seguido del COVID moderado y para terminar con un valor cercano a 0,75, el COVID severo (S24). Y los resultados de la diversidad son los mismos que cuando se utiliza H3 como referencia.



11.CONCLUSIONES

Con el análisis de los resultados podemos determinar que el repertorio inmunológico normal y patológico (para el COVID), no cambia mucho, es decir que no encontramos diferencias significativas que nos demuestren que el repertorio va a cambiar al estar en contacto con una enfermedad.

Con estas métricas no se puede hacer una distinción clara entre los diferentes estados del sistema inmunitario y la patología, pero, sí que observamos pequeñas diferencias entre los repertorios patológicos y los normales. En estos dos meses no pudimos ver exhaustivamente diferentes repertorios normales.

También podemos pensar que algunas de las diferencias vienen determinadas por la edad y el sexo puesto que los datos que hemos obtenido a partir del artículo (Kuri-Cervantes et al., 2020), no lo han tenido en cuenta.

Este proyecto ha establecido una línea de análisis para evaluar el impacto de la elección del repertorio normal para los estudios de patología y también para evaluar la variabilidad entre los repertorios normales.

12.AUTOEVALUACIÓN

En este apartado, voy a autoevaluarme en relación con las prácticas realizadas en IMGT, como en el trabajo de final de grado en biotecnología.

En primer lugar, gracias a estas prácticas, he aumentado mis conocimientos tanto en inmunología como en bioinformática. Aquí he aprendido un nuevo lenguaje de programación (R), y a crear y analizar datos a través de mi programa. También mi estancia en IMGT me ha permitido mejorar mi nivel de francés e inglés. He aprendido a interpretar los resultados y a trabajar con diferentes formatos, para que puedan ser analizados por el programa. También he aprendido a transferir datos de un ordenador a otro gracias a la codificación.

En segundo lugar, gracias a este trabajo de final de grado, he aprendido a relatar correctamente una memoria científica. También a seleccionar los datos y explicarlos. Uno de los puntos más interesantes es la discusión de resultados y conclusiones, desde mi punto de vista, es lo más complejo ya que tienes que



saber analizar los datos correctamente y entender a la perfección la teoría para poder llegar a unas explicaciones acerca de los resultados.

En conclusión, tanto las prácticas como el trabajo me han ayudado a nivel profesional ya que la bioinformática es un campo muy recurrido e interesante. Es imprescindible saber analizar y tratar datos, y esta experiencia me ha ayudado muchísimo. También a la hora de escribir la memoria, he podido enriquecer mis conocimientos sobre inmunología, he aprendido a sintetizar y organizar un trabajo científico. Por finalizar, también estoy satisfecha a nivel personal, por haber compartido tiempo con un gran equipo y con otros estudiantes de prácticas y por llevarme alguna amistad.

13.BIBLIOGRAFÍA

1. Akbar, R., Robert, P. A., Pavlović, M., Jeliakov, J. R., Snapkov, I., Slabodkin, A., Weber, C. R., Scheffer, L., Miho, E., Haff, I. H., Haug, D., Lund-Johansen, F., Safonova, Y., Sandve, G. K., & Greiff, V. (2021). A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell reports*, 34(11), 108856. <https://doi.org/10.1016/j.celrep.2021.108856>
2. Aouinti, S., Giudicelli, V., Duroux, P., Malouche, D., Kossida, S., & Lefranc, M. P. (2016). IMGT/StatClonotype for Pairwise Evaluation and Visualization of NGS IG and TR IMGT Clonotype (AA) Diversity or Expression from IMGT/HighV-QUEST. *Frontiers in immunology*, 7, 339. <https://doi.org/10.3389/fimmu.2016.00339>
3. Aouinti, S., Malouche, D., Giudicelli, V., Kossida, S., & Lefranc, M. P. (2015). IMGT/HighV-QUEST Statistical Significance of IMGT Clonotype (AA) Diversity per Gene for Standardized Comparisons of Next Generation Sequencing Immunoprofiles of Immunoglobulins and T Cell Receptors. *PloS one*, 10(11), e0142353. <https://doi.org/10.1371/journal.pone.0142353>
4. Chi, X., Li, Y., & Qiu, X. (2020). V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology*, 160(3), 233–247. <https://doi.org/10.1111/imm.13176>



5. Francis, J. M., Leistriz-Edwards, D., Dunn, A., Tarr, C., Lehman, J., Dempsey, C., Hamel, A., Rayon, V., Liu, G., Wang, Y., Wille, M., Durkin, M., Hadley, K., Sheena, A., Roscoe, B., Ng, M., Rockwell, G., Manto, M., Gienger, E., Nickerson, J., ... Pregibon, D. C. (2022). Allelic variation in class I HLA determines CD8+ T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2. *Science immunology*, 7(67), eabk3070. <https://doi.org/10.1126/sciimmunol.abk3070>
6. Heikema, A. P., Horst-Kreft, D., Boers, S. A., Jansen, R., Hiltemann, S. D., de Koning, W., Kraaij, R., de Ridder, M., van Houten, C. B., Bont, L. J., Stubbs, A. P., & Hays, J. P. (2020). Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota. *Genes*, 11(9), 1105. <https://doi.org/10.3390/genes11091105>
7. Hickey, M. J., Valenzuela, N. M., & Reed, E. F. (2016). Alloantibody Generation and Effector Function Following Sensitization to Human Leukocyte Antigen. *Frontiers in immunology*, 7, 30. <https://doi.org/10.3389/fimmu.2016.00030>
8. Kuri-Cervantes, L., Pampena, M. B., Meng, W., Rosenfeld, A. M., Ittner, C., Weisman, A. R., Agyekum, R. S., Mathew, D., Baxter, A. E., Vella, L. A., Kuthuru, O., Apostolidis, S. A., Bershaw, L., Dougherty, J., Greenplate, A. R., Pattekar, A., Kim, J., Han, N., Gouma, S., Weirick, M. E., ... Betts, M. R. (2020). Comprehensive mapping of immune perturbations associated with severe COVID-19. *Science immunology*, 5(49), eabd7114. <https://doi.org/10.1126/sciimmunol.abd7114>
9. Li, H., Limenitakis, J. P., Greiff, V., Yilmaz, B., Schären, O., Urbaniak, C., Zünd, M., Lawson, M., Young, I. D., Rupp, S., Heikenwälder, M., McCoy, K. D., Hapfelmeier, S., Ganal-Vonarburg, S. C., & Macpherson, A. J. (2020). Mucosal or systemic microbiota exposures shape the B cell repertoire. *Nature*, 584(7820), 274–278. <https://doi.org/10.1038/s41586-020-2564-6>
10. Llimpe, Y. (2007). *Alteraciones citogenéticas en leucemia mieloide aguda M2 (FAB) en adultos y su correlación con inmunofenotipo y morfología*. CYBERTESIS. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/2447>



11. Mikocziova, I., Greiff, V., & Sollid, L. M. (2021). Immunoglobulin germline gene variation and its impact on human disease. *Genes and immunity*, 22(4), 205–217. <https://doi.org/10.1038/s41435-021-00145-5>
12. Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G., & Lefranc, M. P. (2004). IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *Journal of molecular recognition : JMR*, 17(1), 17–32. <https://doi.org/10.1002/jmr.647>
13. Ruschil, C., Kemmerer, C. L., Beller, L., Gabernet, G., & Kowarik, M. C. (2021). Next Generation Sequencing of Cerebrospinal Fluid B Cell Repertoires in Multiple Sclerosis and Other Neuro-Inflammatory Diseases-A Comprehensive Review. *Diagnostics (Basel, Switzerland)*, 11(10), 1871. <https://doi.org/10.3390/diagnostics11101871>
14. Soto, C., Bombardi, R. G., Kozhevnikov, M., Sinkovits, R. S., Chen, E. C., Branchizio, A., Kose, N., Day, S. B., Pilkinton, M., Gujral, M., Mallal, S., & Crowe, J. E., Jr (2020). High Frequency of Shared Clonotypes in Human T Cell Receptor Repertoires. *Cell reports*, 32(2), 107882. <https://doi.org/10.1016/j.celrep.2020.107882>
15. Weber, C. R. (2022b, enero 1). Reference-based comparison of adaptive immune receptor repertoires. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2022.01.23.476436v1>
16. Weber, C. R., Akbar, R., Yermanos, A., Pavlović, M., Snapkov, I., Sandve, G. K., ... & Greiff, V. (2020). immuneSIM: tunable multi-feature simulation of B- and T-cell receptor repertoires for immunoinformatics benchmarking. *Bioinformatics*, 36(11), 3594-3596.



ANEXOS

*Normativa de Treball Fi de Grau Facultat d'Enologia
Aprovada per Junta de Facultat d'Enologia del dia 30 d'octubre de 2014*

ANNEX 2

FITXA DE SEGUIMENT DEL TUTOR/A del TFG

Nom i Cognoms de l'Alumne/a: Lorena Hernández Servian

Nom i Cognoms del Tutor/a: Ana Fernández Bravo

Data de la entrevista amb l'alumne: 22/04/2022, 03/05/2022

Recomanacions durant el seguiment: Les recomanacions donades a l'alumna han estat les següents: i) Recerca i lectura dels articles rellevants relacionats amb el tema de treball; ii) Seguir un mètode per a descriure els resultats obtinguts fins el moment. iii) Fer petites reunions per *teams* o escriure correus però a comentar els resultat, les possibles interpretacions i els problemes que han pogut sorgir.

Observacions: La alumna durant el treball realitzat fins el moment ha mostrat un gran interès per tindre uns bons resultats, a més de tindre iniciatives al hora de organitzar els resultats en la memòria. Lorena no ha deixat de fer preguntes sobre el treball i ha mantingut un constant contacte amb mi per poder dur a terme correctament l'escrit. Per altra banda, ha rebut correctament qualsevol comentari de la meua part i ha portat a terme aquestes recomanacions en el seu treball. En general, Lorena esta fent un bon treball i te una gran capacitat d'anàlisi.

Observacions Darrera revisió:

En base al treball realitzat, expressi que l'alumna Lorena Hernández Servian ha superat les expectatives que s'havien depositat en ella, a nivell professional.

Signatura del Tutor/a

Signatura del Alumne/a

Tarragona a 25 de maig 2022