



UNIVERSIDAD DE JAÉN

Facultad de Ciencias Experimentales

Trabajo Fin de Grado

Identificación de polimorfismos en genes candidatos del cromosoma X asociados a enfermedades inmunitarias en la población europea

Alumno: López Ureña, Manuel

Junio, 2021



UNIVERSIDAD DE JAÉN



Trabajo Fin de Grado

Identificación de polimorfismos en genes candidatos del cromosoma X asociados a enfermedades inmunitarias en la población europea

Alumno: López Ureña, Manuel

Jaén, Junio, 2021

ÍNDICE

RESUMEN.....	1
ABSTRACT	1
1. INTRODUCCIÓN.....	2
1.1. ¿Qué es un exoma?	2
1.2. Variabilidad genética.....	2
1.3. Enfermedades mendelianas y técnicas NGS.....	4
1.4. Bases de datos y el proyecto 1000Genomes	7
1.5. Enfermedades autoinmunes (AID) y el cromosoma X.....	7
1.6. Justificación del estudio.....	10
2. HIPÓTESIS Y OBJETIVOS	10
3. MATERIAL Y MÉTODOS	11
3.1. Búsqueda de genes del cromosoma X relacionados con el sistema inmune	11
3.2. Recolección de SNPs de cada gen.....	12
3.3. Análisis y reestructuración de SNPs.....	14
3.4. Aplicación de filtros y búsqueda de frecuencias genotípicas.....	14
3.5. Búsqueda de la influencia en la expresión de la proteína y del ARNm.....	15
4. RESULTADOS	16
4.1. SNPs que afectan a la secuencia de la proteína	16
4.1.1. Gen CCDC22.....	17
4.1.2. Gen CD40LG	19
4.1.3. Gen CYSLTR1.....	20
4.1.4. Gen TLR7	22
4.1.5. Gen TLR8	23
4.1.6. Gen XIAP.....	24
4.2. SNPs que afectan en la expresión del ARNm.....	25
4.2.1. Gen CCDC22.....	25
4.2.2. Gen CD40LG	26
4.2.3. Gen CYSLTR1.....	27
4.2.4. Gen TLR7	28
4.2.5. Gen TLR8	29
4.2.6. Gen XIAP.....	30
4.3. Resumen de los resultados	31
5. DISCUSIÓN.....	32
6. CONCLUSIONES.....	34
7. REFERENCIAS	35

RESUMEN

Introducción: Una de las variaciones genéticas más estudiadas son los SNPs, los cuales son usados para realizar estudios GWAS, y así encontrar alguna asociación con distintas enfermedades. El cromosoma X está muy poco estudiado en estos estudios, y es por ello por lo que decidimos analizarlo.

Objetivos: La finalidad de este estudio fue encontrar aquellos genes que porten SNPs relacionados con distintas enfermedades asociadas a infecciones virales.

Material y Métodos: Mediante diferentes páginas web (como Ensembl o SNP Nexus), procedimos a estudiar los SNPs más significativos pertenecientes al sistema inmune y hallados en el cromosoma X.

Resultados: Pudimos encontrar 6 genes candidatos, además de ver una influencia de tanto SNPs intrónicos, como intergénicos de otras regiones en la expresión de los genes candidatos encontrados.

Conclusiones: Se puede destacar la importancia de CD40LG, TLR7 y TLR8 por su participación en rutas clave del sistema inmune, así como su asociación con distintas enfermedades.

ABSTRACT

Introduction: One of the most studied genetic variations are SNPs, which are used to carry out GWAS studies, and thus find an association with different diseases. The X chromosome is very little studied in these studies, and that is why we analyzed it.

Objectives: The purpose of this study was to find those genes that carry SNPs related to different diseases associated with viral infections.

Material and Methods: Through different web pages (such as Ensembl or SNP Nexus), we proceeded to study the most significant SNPs belonging to the immune system and found on the X chromosome.

Results: We were able to find 6 candidate genes, in addition to seeing an influence of both intronic and intergenic SNPs from other regions on the expression of the candidate genes found.

Conclusions: The importance of CD40LG, TLR7 and TLR8 is highlighted due to their participation in key pathways of the immune system, as well as their association with different diseases and intergenic SNPs from other regions.

1. INTRODUCCIÓN

1.1. ¿Qué es un exoma?

La vida como la conocemos se basa fundamentalmente en el genoma, es decir, toda aquella información biológica que define cómo es cada individuo. La mayoría de los genomas están constituidos por ADN (ácido desoxirribonucleico), aunque siempre hay excepciones, como es el caso de algunos virus, cuya información genética se constituye por ARN (ácido ribonucleico).

El genoma humano es un claro ejemplo de genoma animal pluricelular, el cual se divide en dos tipos diferentes: el genoma mitocondrial, constituido por alrededor de 17kb (kilobases), codificando sólo para 37 genes; y el genoma nuclear, constituido por más de 3 billones de nucleótidos agrupados en 24 cromosomas (22 autosomas y 2 sexuales, el X e Y), conteniendo casi 40.000 genes (Brown, 2008).

Tras diferentes métodos de análisis del genoma humano, finalmente se pudo observar que, de esos casi 40.000 genes, 26.000 poseían una seguridad bastante elevada, mientras que los otros 12.000, tenían homología con ratón o con otro tipo de evidencias, pero estas eran más débiles.

Sin embargo, toda esta información genética no se encuentra de forma continua, si no que se divide en regiones codificantes, las que se transcribirán (denominados exones, siendo solo el 2% de todo el genoma), regiones no codificantes intercalando a los exones (llamados intrones, constituyendo el 24% del genoma), y ADN intergénico (el mayor componente del genoma, con un 75%, se corresponde fundamentalmente con fósiles de retrotransposones y otros tipos de ADN repetido) (Craig Venter *et al.*, 2001). Pues bien, al conjunto de los exones que componen nuestro genoma es lo que se denomina exoma (Ng *et al.*, 2009).

1.2. Variabilidad genética

Ahora bien, la información que existe en nuestro genoma puede variar. Esto es lo que se denomina polimorfismos o mutaciones, las cuales originarán nuevas combinaciones durante la meiosis, a las cuales, la selección natural actuará sobre ellas. Estas mutaciones pueden ser directas (si afectan directamente a los fenotipos originales), inversas (si restauran el fenotipo original), o supresoras (si suprimen los efectos de la mutación anterior).

Igualmente, si se produce una mutación, estas pueden causar, o no, alteraciones en la proteína, es decir, según los efectos fenotípicos que causen, podemos clasificar las mutaciones como sinónimas/no sinónimas (cambian/no cambian los aminoácidos de la proteína), de pérdida/ganancia de función, condicionales (si solo se expresa la proteína en determinadas condiciones), o letales (si su presencia causa la muerte prematura).

Cabe destacar que las mutaciones en sí pueden ocasionarse de forma espontánea, que sería la forma natural en las que se producen, o de forma inducida en un laboratorio con agentes físicos o químicos que tengan efectos mutagénicos (Pierce, 2009).

Todas las mutaciones que se pueden producir en nuestro genoma se pueden clasificar en tres grupos:

- Mutaciones cromosómicas: se definen tal cual a las que afectan a cromosomas enteros, pudiéndonos encontrarlos con deleciones o duplicaciones de secciones del cromosoma, inversiones, o translocaciones (parte de un cromosoma se intercambia con otra parte de otro cromosoma) (Contreras, 2009).
- Elementos transponibles: también conocidos como transposones, son segmentos de ADN con capacidad de moverse por el genoma, constituyendo una parte cuantitativamente importante del genoma (entre el 25-50% del genoma humano son transposones). Podemos encontrarlos con transposones de ADN y retrotransposones (Frutos, 1997). Estos tienen variabilidad en la secuencia de inserción dentro del genoma.
- Mutaciones puntuales: son las que se definen como cualquier cambio producido en la secuencia de nucleótidos de ADN. Según el tejido que afectan, pueden ser somáticas o germinales, pero lo más común es clasificarlas en función del cambio que produjeron en el ADN. Es así como encontramos transiciones (cambio de purinas por otra purina, o una pirimidina por otra), transversiones (cambio de una purina por una pirimidina o viceversa), inserciones (se añade uno o varios nucleótidos a la secuencia), deleciones (se elimina uno o varios nucleótidos de la secuencia), o microsatélites (repeticiones de nucleótidos) (Mansego *et al.*, 2007).
- Mutaciones por variación del número de copias. Se producen por recombinación homóloga entre segmentos altamente repetidos que flanquean los genes, dando lugar a deleciones completas o bien multiplicación del número de copias.

Dentro de estas, debemos resaltar los SNPs (*Single Nucleotide Polymorphism*), los cuales constituyen el 93% de las variaciones, y se definen como el cambio en una sola base nucleotídica en el genoma (Sherry *et al.*, 2001) que, aunque no aporte ninguna información en especial acerca del gen en el que está situado, sí puede ser utilizada para relacionarlo con el fenotipo que exprese aquel individuo que porte dicho SNP (Chial, 2008).

La frecuencia en la que aparece en una población un SNP está regida por el MAF (*Minor Allele Frequency*, frecuencia del alelo menos común), el cual debe tener un valor superior al 1% en la población para considerarse como un SNP (si no, se consideraría como mutación puntual). Los SNPs situados en regiones codificantes pueden generar cambios sinónimos o no sinónimos (cambio de aminoácidos), mientras que los situados en regiones no codificantes (tanto aguas abajo, que sería antes de leerse la región codificante, como aguas arriba, si se encuentran después de la región codificante) pueden llegar a alterar el proceso de splicing (post-procesamiento del ARN después de la transcripción), los lugares por donde se unen factores de transcripción, o incluso el ARN entero (si el cambio producido modifica el patrón de lectura, lo que se conoce como “*frameshift*”) (Streisinger *et al.*, 1966). Por último, también se pueden encontrar en las regiones UTR (tanto 3'UTR como 5'UTR), que son zonas del gen que se transcriben, pero no se traducen, cuyas funciones son, entre otras, determinar la vida media del ARNm, y el momento, así como el subcompartimento celular, donde tendrá lugar su traducción (Anthony *et al.*, 2021). En ocasiones, los SNPs se heredan conjuntamente, escapando de la recombinación (por estar muy próximos entre sí), conformando lo que conocemos como haplotipo (Belmont *et al.*, 2005).

1.3. Enfermedades mendelianas y técnicas NGS

Un porcentaje importante de los trastornos mendelianos causados por mutaciones se pueden resolver mediante la secuenciación de exomas, es decir, determinar la composición de todos los exones de un determinado paciente, que, comparándolo con un pariente sano, nos permite diferenciar entre un individuo sano, y un individuo que padece cierta variación, permitiendo identificar qué mutación causó la enfermedad génica (Bamshad *et al.*, 2011).

Así es como entra en acción la secuenciación de exomas, la cual se empezó allá por 2005 con la primera plataforma de secuenciación masiva (Majewski *et al.*, 2011).

Secuenciar esta parte del genoma nos permite ver las diferentes variaciones alélicas y su relación con el fenotipo que estos expresan, ya que la mayoría de mutaciones producidas en exones podrían causar enfermedades mendelianas (Biesecker *et al.*, 2011). Con la incorporación de más conocimientos bioinformáticos y el avance de la tecnología de secuenciación, resulta más cómodo y más sencillo el encontrar este tipo de enfermedades mediante la secuenciación exómica usando los métodos de nueva generación (*Next Generation Sequencing*, NGS). Aun así, es necesario tener tanto pacientes como familiares fenotipados adecuadamente, al igual que su modo de herencia, estado de la enfermedad, y frecuencia poblacional (Gilissen *et al.*, 2012).

Pero, ¿qué es la NGS? También recibe el nombre de secuenciación masiva paralela, y se define como todas aquellas tecnologías que permiten secuenciar el ADN y/o ARN de forma más rápida y económica, con la finalidad de ser capaz de secuenciar un genoma completo en torno a un solo día, realizando secuenciaciones de millones de pequeños fragmentos de ADN a la vez (Behjati y Tarpey, 2013), aunque rara vez se realiza una secuenciación completa, pues como ya hemos visto anteriormente, gran parte del genoma es poco relevante, por lo que se suele optar por secuenciar sólo el exoma. Lo bueno de la NGS, es que no se necesita un conocimientos genómicos previos o sobre genomas, además de que ofrece una resolución de nucleótido por nucleótido, permitiendo detectar relaciones intergénicas, variantes alélicas o polimorfismos nucleotídicos. Igualmente, se necesita menos cantidad de muestra (con cantidades en nanogramos, basta), y su reproducibilidad es mucho mayor que los métodos de primera generación, o métodos de Sanger (Huerta y Burke, 2016).

Las enfermedades monogénicas raras, es decir, las que aparecen por la modificación de un gen en concreto, no superan las 5000, y lo normal es que el gen causante sea desconocido (Hamosh *et al.*, 2000). La secuenciación de exomas entra en acción cuando la enfermedad es causada por una serie de complejos genéticos que interaccionan entre sí, y no se tiene muy claro de dónde se origina dicha enfermedad, realizándose, por tanto, una secuenciación del exoma completo (si se supiera el gen causante, entonces se haría una secuenciación dirigida a la zona donde se halla ese gen). De hecho, enfermedades comunes como la esquizofrenia (Mitchell y Porteous, 2011) o el autismo (Vissers *et al.*, 2010) se ha podido determinar que pueden deberse a una interacción multifactorial de diferentes enfermedades raras monogénicas (McClellan y King, 2010).

Aunque antiguamente se realizaban secuenciaciones de genes candidatos por el método de Sanger, relacionando genes asociados a enfermedades similares (Botstein y Risch, 2003), las tecnologías NGS están evolucionando, mejorando el rendimiento y su asequibilidad (Mardis, 2011), abriendo las puertas de poder secuenciar millones de variantes genómicas (que pueden provenir de familiares), para interpretar los resultados y destacar cuál de estas variantes causan la enfermedad (Ng *et al.*, 2010). Como ya se sabe que la mayoría de las enfermedades mendelianas se producen por trastornos en el exoma (Botstein y Risch, 2003), descartando todo aquel genoma que no forme parte de los exones, por lo que la cantidad de variantes a seguir se reduce drásticamente (alrededor de 25.000 variantes aproximadamente) (Roach *et al.*, 2010), que, si le añadimos la eficiencia de la NGS, podemos estudiar simultáneamente hasta 100 exomas a la vez (Gilissen *et al.*, 2012).

Para identificar los genes causantes de la enfermedad monogénica, existen diferentes estrategias para ello. He aquí algunos ejemplos:

- Estrategia de vinculación: secuenciar varios miembros de la familia afectados, de tal forma de que se pueda ver la variación compartida, además de apoyarse de los datos de los familiares no afectados. Ya se pudo apreciar en artículos pasados (Ng *et al.*, 2010) que, comparando a dos hermanos enfermos (que se espera una similitud en su ADN del 50%), fueron capaces de reducir considerablemente la cantidad de genes candidatos causantes de la enfermedad.
- Estrategia de homocigosidad: se da para casos de trastornos recesivos raros a los que se le añade una posible consanguinidad. Esta estrategia, a diferencia de la anterior, solo identifica genes que se hallan en un gran tramo homocigoto que contengan gran cantidad de información de polimorfismos de un solo nucleótido (SNP), ya que anteriormente se afirmó que el exoma proporciona información suficiente de SNPs para que se pueda realizar un mapeo de homocigosis fiable (Becker *et al.*, 2011).
- Estrategia del “double-hit”: si se sospecha que el trastorno es recesivo, pero no hay consanguinidad, y además, no podemos disponer de familiares, se puede secuenciar el exoma para seleccionar tramos tanto homocigotos y heterocigotos, como ya hicieron estudios anteriormente (Gilissen *et al.*, 2010).

1.4. Bases de datos y el proyecto 1000Genomes

La incorporación de las tecnologías NGS supuso una ampliación en el conocimiento de diferentes genomas, lo que conllevó a que dicha información se empezará a almacenar en base de datos que pudieran estar al alcance de la mayoría. Estas bases de datos se fueron especializando y completando con el paso del tiempo, pudiendo llegar hoy en día a tener bases de datos que nos proporcionen información de referencia a gran escala de diferentes variaciones genéticas halladas en una población en concreto (Soussi *et al.*, 2019), como son las bases de datos públicas de gnomAD o 1000genomes, cuyas motivaciones y ambiciones son las mismas: conocer mejor el genoma humano.

La que utilizaremos en nuestro estudio será 1000genomes, la cual se centra en representar la mayor cantidad de variantes genéticas con frecuencias superiores al 1% en diferentes poblaciones, información que podía ser útil de cara para comprender las variaciones del genoma humano, mejorando así la precisión y sensibilidad a la hora de secuenciar material genético (Birney y Soranzo, 2015), avanzando en el desarrollo del conocimiento y comprensión de las enfermedades biológicas (Sudmant *et al.*, 2015), pues esta base de datos permite usar arrays (Voight *et al.*, 2013), completar información incompleta de diferentes genotipados (Howie *et al.*, 2012), y clasificar y filtrar las variantes genómicas que nos pudiéramos encontrar (Xue *et al.*, 2012).

1.5. Enfermedades autoinmunes (AID) y el cromosoma X

Es sabido que las enfermedades autoinmunes (autoimmune diseases, AID), es decir, reacciones inmunes autodestructivas (Kono y Theofilopoulos, 2017), están altamente relacionadas con una expresión alterada de los genes que interactúan en el desarrollo del sistema inmune, pudiendo afectar a una parte considerable de la población (Invernizzi *et al.*, 2009). Dentro de todos los cromosomas que componen el genoma humano, cabe destacar la importancia de los cromosomas sexuales, pues un importante número de genes implicados en una respuesta inmune están localizados en el cromosoma X (Libert *et al.*, 2010).

Es por ello que la mayor parte de las AIDs presentan una preponderancia femenina destacable comparándolo con los hombres (Eaton *et al.*, 2007), debido a que ellas portan dos cromosomas X, mientras que ellos solo portan uno (Balaton *et al.*, 2018). Y como la dosis genética en hembras sería el doble que la de machos, uno de los cromosomas X femeninos se inactiva en el proceso conocido como “inactivación del cromosoma X (XCI)” (Lyon, 1961), produciéndose de forma progresiva y aleatoria en todos los alelos de dicho cromosoma, y cuyo proceso se regula por genes no codificantes (es decir, que su función la realizan sin necesidad de traducirse) como son los genes Tsix, Xist,... (Augui *et al.*, 2011). Es por ello que varones con síndrome de Klinefelter (tienen una disposición cromosómica sexual de XXY) son propensos a desarrollar enfermedades típicas de mujeres (XX), como sería el lupus eritematoso sistémico (Sawalha *et al.*, 2009).

Pero centrándonos en los varones, que son hemicigotos para el cromosoma X, y por tanto, no hay ningún tipo de inactivación en este cromosoma, supone una gran desventaja para ellos, pues cualquier mutación que se produzca en este cromosoma se va a expresar fenotípicamente (Albert *et al.*, 2018), apreciándose en cómo enfermedades autoinmunes primarias (PID) ligadas al cromosoma X suelen manifestarse en varones, cuando en mujeres solo tienen efectos leves o simplemente son portadoras sin sufrir la enfermedad (Pessach y Notarangelo, 2009). Es así como podemos confirmar que el cromosoma X interviene de manera representativa en la autoinmunidad (si hablamos de mujeres) y en la inmunodeficiencia (si hablamos de hombres) (Bussone y Mouthon, 2008).

Enfermedades autoinmunes tales como la artritis reumatoide (Suzuki *et al.*, 2003), la diabetes mellitus tipo 1 (Nielsen *et al.*, 2003) o el lupus eritematoso sistémico (Prokunina *et al.*, 2002), están relacionadas con cambios alélicos de determinados SNPs de genes del cromosoma X.

Estas enfermedades en mujeres se dan por el escape de ciertos genes a la inactivación del cromosoma X (Souyris *et al.*, 2019), mientras que los hombres, que solo tienen una dosis de cromosoma X, la predisposición a contraer la enfermedad es mayor por presentar hemicigosis.

Si nos fijamos en infecciones de diferentes virus, como el de la influenza, se pudo observar que la infección podía variar dependiendo del sexo, ya que las hormonas sexuales intervenían en las respuestas inmunitarias. En otro tipo de infecciones, como las del VIH, la predisposición a la infección era mayor en varones, quizás por presentar menor dosis génica de aquellos genes que participan en la ruta del interferón (como el TLR7). Comparado con mujeres, estas producían mayor cantidad de interferón (IFN) en respuesta a la infección de este virus (Rechtien y Altfeld, 2019). En general, las mujeres suelen desarrollar respuestas inmunes innatas y adaptativas más fuertes que los hombres (Schurz *et al.*, 2019), y que no parecen verse alteradas aun si la mujer está embarazada (Chen *et al.*, 2020).

Esto podría explicar por qué los hombres infectados por covid-19 sufren una enfermedad más grave y su mortalidad es mayor si lo comparamos con mujeres, aunque también habría que recalcar que algunos estudios demostraron que las mujeres son más propensas a lavarse las manos (Johnson *et al.*, 2003), así como ser más preventivas con las enfermedades (Bertakis *et al.*, 2000), lo que también influye en su probabilidad de infección. Ya hay estudios que han realizado la comparativa entre sexos según el país (Scully *et al.*, 2020), mostrándonos la mayor susceptibilidad masculina a infecciones con este virus con gráficas como la siguiente:

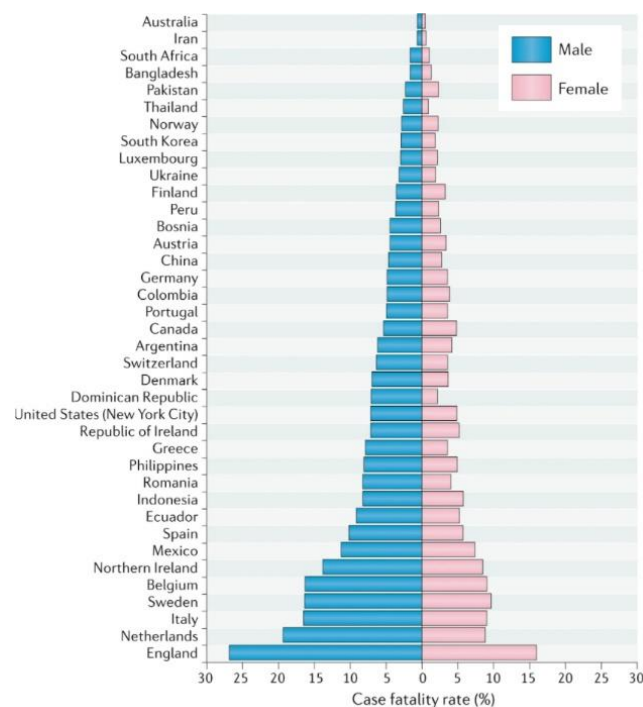


FIGURA 1.1. Análisis comparativo de las tasas de letalidad de covid-19 según el sexo en diferentes países (Scully *et al.*, 2020).

1.6. Justificación del estudio

Los estudios de asociación del genoma completo (GWAS), son actualmente la tecnología más comúnmente utilizada para descubrir la base genética de enfermedades de todo tipo. Esto se realiza mediante el genotipado masivo (usando alrededor de 500.000-1.000.000 SNPs) de cohortes de casos y controles, con el fin de encontrar marcadores genéticos (causales o no) implicados en una mayor susceptibilidad (Manolio, 2010).

Casi todos los estudios GWAS que se han realizado, obviaban al cromosoma X debido a la complicación que este acarrea a la hora de comparar estadísticamente los datos de los hombres con las mujeres (Schurz *et al.*, 2019). Recientemente se han desarrollado técnicas para poder incorporar este cromosoma a las estadísticas, como el software XWAS (Gao *et al.*, 2015) o PLINK (Chang *et al.*, 2015), pero lo cierto es que es difícil encontrar datos relacionados con los cromosomas sexuales en los estudios actuales de covid-19, por ejemplo, donde están encontrando muchos genes candidatos que agravan la enfermedad en cromosomas autosómicos (Ganna, 2021), pero no están teniendo en cuenta la cantidad de genes del sistema inmunológico que se encuentran en el cromosoma X.

Es por ello que decidimos analizar los diferentes genes que nos podemos encontrar en este cromosoma, y así complementar información acerca de la relación de los marcadores que encontraremos con diferentes enfermedades infecciosas.

2. HIPÓTESIS Y OBJETIVOS

Suponemos como hipótesis que SNPs con una frecuencia relativa alta en genes del sistema inmune hallados en el cromosoma X pueden ser responsables de la mayor gravedad de ciertas infecciones en hombres europeos por ser hemicigotos, si los comparamos con mujeres.

Esta hipótesis nos lleva a querer cumplir los siguientes objetivos:

- Analizar los genes del cromosoma X relacionados con el sistema inmune.
- Identificar SNPs potencialmente funcionales presentes en dichos genes.
- Encontrar posibles haplotipos que estén relacionados con enfermedades inmunes, y aumenten la predisposición a tener dichas enfermedades.
- Determinar la influencia de las variaciones en los SNPs en la expresión de la proteína y el ARN mensajero.

3. MATERIAL Y MÉTODOS

3.1. Búsqueda de genes del cromosoma X relacionados con el sistema inmune

Primeramente, se identificaron los genes implicados (Tabla 3.1) en la respuesta inmunitaria del cromosoma X, así como su posición en este, analizando un total de 45 genes (Meester *et al.*, 2020):

TABLA 3.1. Tabla con todos los genes del cromosoma X analizados en este estudio (Meester *et al.*, 2020).

GEN	LOCALIZACIÓN	GEN	LOCALIZACIÓN	GEN	LOCALIZACIÓN
APEX2	Xp11.21	IKBK	Xq28	PIM2	Xp11.23
ARHGAP4	Xq28	IL1RAPL2	Xq22.3	SASH3	Xq26.1
ARHGEF6	Xq26.3	IL13RA1	Xq24	SEPT6	Xq24
CCDC22	Xp11.23	IL13RA2	Xq23	SH2D1A	Xq25
CD40LG	Xq26.3	IL2RG	Xq13.1	SH3KBP1	Xp22.12
CFP	Xp11.23	IRAK1	Xq28	TAB3	Xp21.2
CXCR3	Xq13.1	ITM2A	Xq21.1	TFE3	Xp11.23
CYBB	Xp21.1-p11.4	MPP1	Xq28	TLR7	Xp22.2
CYSLTR1	Xq21.1	MSL3	Xp22.2	TLR8	Xp22.2
DOCK11	Xq24	MSN	Xq12	TMSB4X	Xp22.2
ELF4	Xq28	NKAP	Xq24	TSC22D3	Xq22.3
FOXP3	Xq21.1	NOX1	Xq22.1	VSIG4	Xq12
GAB3	Xq13.1	NXT2	Xq23	WAS	Xp11.23
GPR174	Xq21.1	OTUD5	Xp11.23	ZBTB33	Xq24
IGBP1	Xq13.1	P2RY10	Xq21.1	XIAP	Xq25

3.2. Recolección de SNPs de cada gen

A continuación, se debe seleccionar toda la información que se encuentra en la base de datos de 1000genomes (Auton *et al.*, 2015), y para ello utilizamos la página web de Ensembl (Howe *et al.*, 2021), concretamente la versión del genoma GRCh38.p13. De todas las poblaciones ofrecidas por 1000genomes, nuestro estudio se centró en cuatro bastante similares:

- CEU: residentes de Utah con ancestros del norte y oeste de Europa. Posee 187 muestras.
- GBR: población británica en Inglaterra y Escocia. Posee 107 muestras.
- IBS: población española. Posee 162 muestras.
- TSI: habitantes de la Toscana, Italia. Posee 112 muestras.

Se procede a buscar los genes del cromosoma X que pueden estar relacionados con enfermedades autoinmunes, y seleccionamos las coordenadas de cada gen (Meester *et al.*, 2020).

Posteriormente hacemos uso de las herramientas de “VCF to PED Converter” de Ensembl para extraer los SNPs de estos genes, incorporando las coordenadas del gen de interés, así como las poblaciones que nos interesaban estudiar.

Esta herramienta nos proporciona un archivo, con formato .info, que nos da el código de todos los SNPs, y un archivo comprimido que, al descomprimirlo nos da un fichero en formato .ped, el cual nos proporciona la información del genotipo para cada SNP en la región genómica seleccionada para todos los individuos de la población seleccionada. Tanto el archivo .ped (Fig. 3.1) como el .info (Fig. 3.2) los abriríamos con Microsoft Excel para trabajar con ellos.

[illegible]

FIGURA 3.1. Archivo .ped del gen TLR7 abierto en Excel.

[illegible]

FIGURA 3.2. Archivo .info del gen TLR7 abierto en Excel.

Por otro lado, toda la información aportada en el archivo .ped no discriminaba los datos por sexo, así que optamos por eliminar todas aquellas muestras heterocigotas para así quedarnos con muestras homocigotas, y así pudieran ser más representativas de los varones (ya que estos solo tienen una sola copia del cromosoma X y por tanto son hemicingotos).

3.3. Análisis y reestructuración de SNPs

El análisis se realizó en la página web de SNPnexus (Dayem Ullah *et al.*, 2018), utilizando la opción del genoma GRCh37/hg19 por ser la opción más completa, e incorporamos los datos que se pueden encontrar en el archivo .info, a cuyos códigos se le debe incorporar “dbSNP” delante de cada uno para que SNPnexus pueda reconocer los SNPs como tales.

A continuación la página nos da opción de elegir diferentes algoritmos e información concreta de los datos que le incorporamos, destacando la selección de la región europea, así como las enfermedades detectadas bajo “Genetic Association of Complex Diseases and Disorders (GAD)”. Los datos conseguidos en este análisis podían ser descargados como archivos .txt que podrían ser abiertos con facilidad también en Excel, con la finalidad de agrupar todo en una misma hoja.

Así, conseguimos un Excel con todos los SNPs de todos los genes estudiados, cada uno con su frecuencia en Europa, el cambio producido en la secuencia, y las enfermedades relacionadas.

3.4. Aplicación de filtros y búsqueda de frecuencias genotípicas

Una vez que tuvimos la lista completa de todos los SNPs que pudimos extraer de Ensembl, sumado a la información obtenida de SNPnexus, procedimos a aplicar diferentes filtros para encontrar aquellos SNPs con potenciales efectos funcionales, que determinarían los genes candidatos.

Para ello, filtramos todos los SNPs que estaban asociados a enfermedades de tipo inmune o causadas por infecciones, además de que deben estar presentes en la población en al menos un 1%, ya que valores muy altos de prevalencia indican que esos SNPs se han establecido en la población y por tanto no supondrían una diferencia significativa tenerlos o no.

En la página de Ensembl, buscando el SNP de interés, podemos encontrarnos los datos de tanto las frecuencias alélicas como genotípicas de cada SNP, centrándonos especialmente en estas últimas, ya que la información proporcionada está separada por sexo (aunque la suma de todas las frecuencias genotípicas, masculinas y femeninas, sigue dando 1), además que es más relevante ver cómo se presentan dichos alelos en la población (de forma homocigota o heterocigota en mujeres, o qué alelo predomina más en varones).

3.5. Búsqueda de la influencia en la expresión de la proteína y del ARNm

Los archivos extraídos de SNPnexus ya nos proporcionan la información sobre la influencia en la expresión/función de la proteína, pudiendo ver los propios datos de 1000genomes (si produce un cambio sinónimo o no sinónimo, o si no le afecta como tal), o viendo predicciones creadas por algoritmos como los de PolyPhen (Ramensky *et al.*, 2002), SIFT (Vaser *et al.*, 2016), y ClinVar (Landrum *et al.*, 2018), que predicen si los cambios producidos son benignos o malignos.

Para saber la alteración en la expresión del ARNm, utilizamos la página web de GTEx Portal (Aguet *et al.*, 2020), en donde podemos ver cómo y dónde se expresa el gen así como los SNPs asociados a eQTL (*expression Quantitative Trait Loci*), es decir, que están asociados significativamente a loci que contribuyen en los niveles de expresión del ARNm (Strunz *et al.*, 2018). En esta página podemos ver todos los SNPs asociados al gen que busquemos (visualizándose como puntos rojos), regulando su expresión. De todos los tejidos que podemos elegir, nosotros nos centramos en analizar sangre, pulmón y bazo, por su relación con el sistema inmune o por verse las consecuencias de una infección en estos. Todos estos SNPs esta asociados estadísticamente a un p-valor que confirman su relación con el gen de interés, y que solo se representan si poseen un valor de p inferior a 10^{-4} .

4. RESULTADOS

4.1. SNPs que afectan a la secuencia de la proteína

Buscando una frecuencia europea superior a 0.01 en aquellos SNPs que estaban asociados a enfermedades infecciosas/autoinmunes, y que se hallaban en regiones exónicas (ya sean regiones codificantes, o situadas en 3UTR o 5UTR), pudimos resaltar 44 SNPs repartidos en 12 genes.

El problema de estos SNPs es que solo el 18% de los seleccionados generan cambios no sinónimos en la proteína (Fig. 4.1).

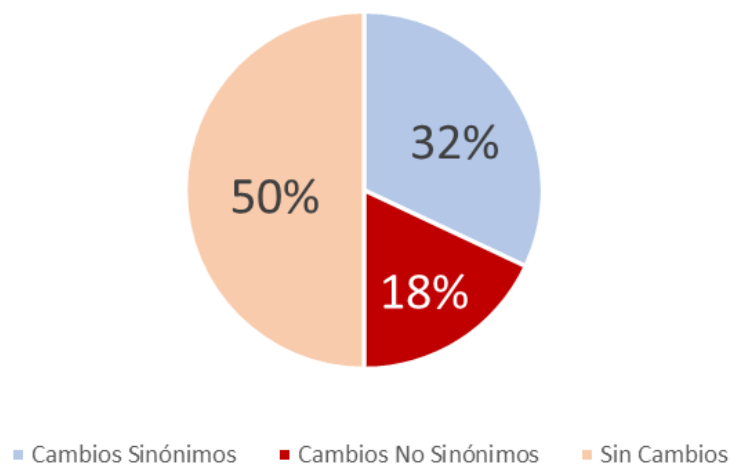


FIGURA 4.1. Representación gráfica de los cambios producidos en la producción de proteína de los SNPs que cumplían las condiciones del primer filtro (algunos SNPs tenían más de un cambio).

Así que se procedió a examinar ese porcentaje de SNPs que producían cambios notables en la secuencia de la proteína, pudiendo tener esta vez 7 SNPs repartidos en 6 genes candidatos, que explicaremos a continuación.

4.1.1. Gen CCDC22

Este gen interviene en la activación de linfocitos B. Según los datos de 1000genomes, podemos encontrar dos SNPs relacionados con un cambio en la proteína:

- rs143790434 (C-T). Produce un cambio no sinónimo en el aminoácido 384 del gen, donde originalmente hay arginina, y esta puede cambiar a serina (R>S), o a cisteína (R384C).

En cuanto a frecuencias genotípicas (Fig. 4.2), podemos separarlas en función del sexo: el 49.5% de las europeas estudiadas por 1000genomes presentaban homocigosis CC, el 0.8% presentaban heterocigosis CT, y no se encontraron homocigotos TT. En cuanto a los varones estudiados, se encontró un 49.6% con C, y un 0.8% con T (solo portan un cromosoma X, y por tanto, solo un alelo).

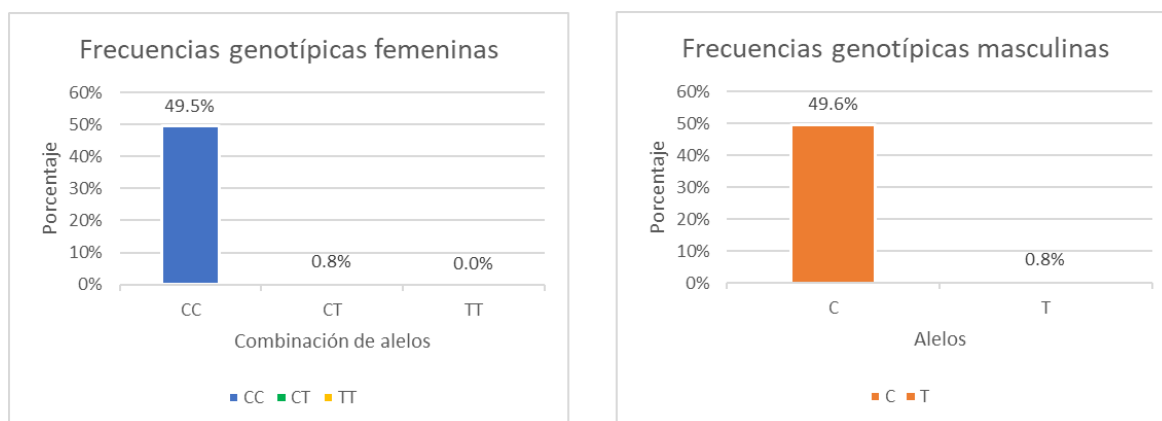


FIGURA 4.2. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs143790434, extraído de Ensembl.

No se han encontrado predicciones por parte de SIFT y PolyPhen, pero ClinVar sugiere que dicho cambio es benigno.

- rs2294020 (G-A). El cambio no sinónimo se produce en el aminoácido 249 del gen, en el que inicialmente se encuentra glutamina, pero que puede cambiar a histidina (Q249H).

Hablando de las frecuencias genotípicas (Fig. 4.3), las femeninas eran del 28% con homocigosis GG, el 12.9% presentaban heterocigosis GA, y el 3.4% presentaban homocigosis AA. En cuanto a las masculinas, se aprecia un 34.8% con G, y un 12.9% con A.

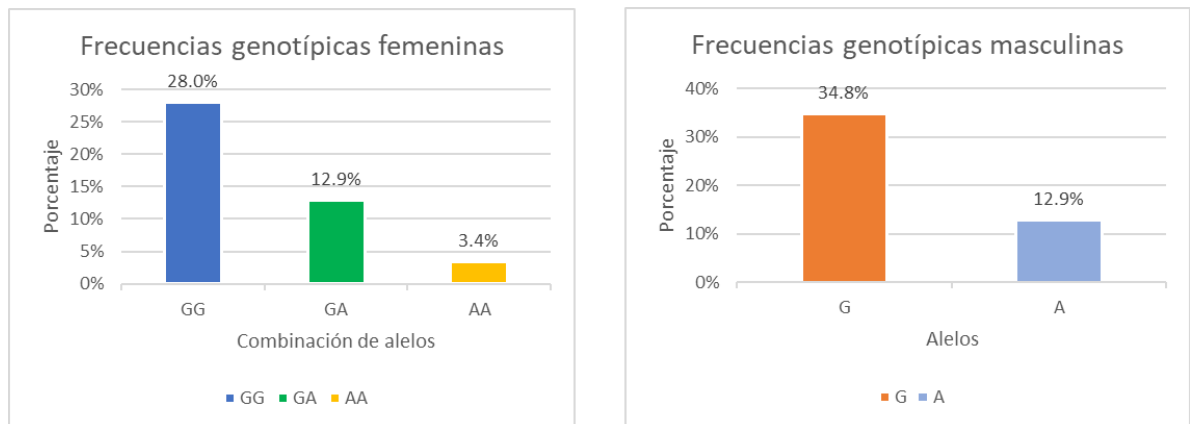


FIGURA 4.3. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs2294020, extraído de Ensembl.

Tampoco se encuentran predicciones de SIFT y PolyPhen, pero ClinVar sugiere que dicho cambio es benigno.

Por otro lado, otros dos SNPs son resaltados por los algoritmos SIFT y PolyPhen:

- rs144632022 (G-A). Posee una puntuación de 0.08 en SIFT, y de 0.666 en PolyPhen, pudiendo producir un cambio en el aminoácido 430, cambiando alanina por treonina (A430T). Por parte de ClinVar, sugiere que es un cambio benigno. Sin embargo, apenas se encuentra el alelo A en la población.
- rs147222955 (G-A). Tiene una puntuación de 0.01 en SIFT, y de 1 en PolyPhen, pudiendo producir un cambio en el aminoácido 546, cambiando ácido aspártico por asparagina (D546N). Por parte de ClinVar, sugiere que existen interpretaciones conflictivas de patogenicidad.

Las frecuencias genotípicas de este alelo son similares al del alelo anterior.

Respecto a la enfermedades asociadas, este gen se relaciona inmunológicamente con la rinitis alérgica (Suttner *et al.*, 2010).

4.1.2. Gen CD40LG

Este gen también participa en la activación de linfocitos B. El proyecto 1000genomes nos remarca un SNP que produce un cambio en la proteína:

- rs148594123 (G-A). El cambio no sinónimo sucede en el aminoácido 198 del gen, donde originalmente hay glicina, la cual puede cambiar a arginina (G198R).

En cuanto a frecuencias genotípicas (Fig. 4.4), las mujeres estudiadas presentan un 49.7% de homocigotos GG, un 2.6% son heterocigotos GA, y no se encontraron homocigotos AA. Por otro lado, el 46.9% de los varones presentan G, y un 0.8% presentan A.

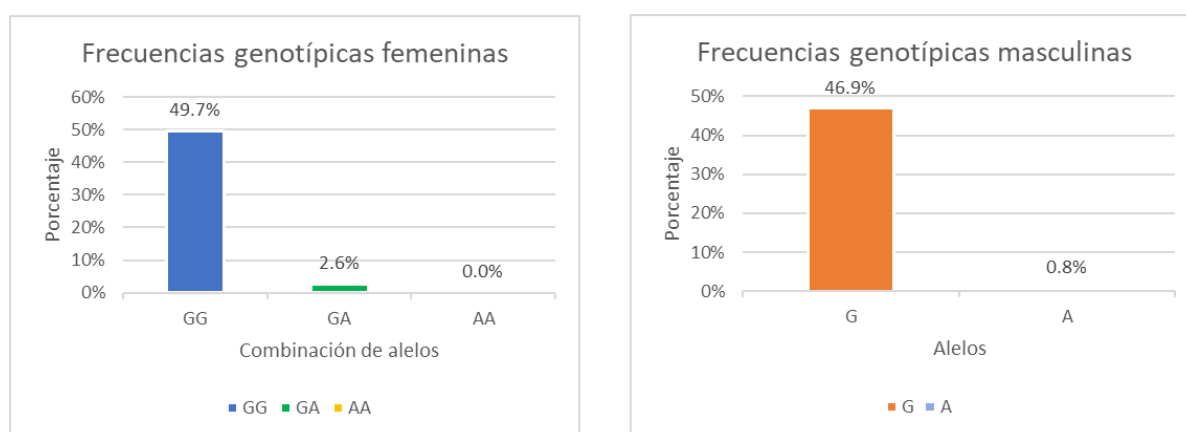


FIGURA 4.4. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs148594123, extraído de Ensembl.

Los datos aportados por los algoritmos de SIFT y PolyPhen proporcionan una puntuación de 0.3 y 0.706 respectivamente, sugiriendo que dicho cambio puede ocasionar daño. Además proponen que este mismo SNP puede ocasionar el mismo cambio en otra zona (concretamente, en el aminoácido 219, G219R), proporcionando una puntuación de 0.31 en SIFT, y 0.138 en PolyPhen, causando esta vez un cambio benigno.

Por otro lado, ClinVar supone un cambio benigno.

Según los algoritmos SIFT y PolyPhen, puede haber otro SNP problemático (rs186774716, C-T), pero que tiene una puntuación asignada muy baja, además de que apenas se encuentra el alelo minoritario en la población.

Si vemos las enfermedades asociadas a este gen, podemos ver una relación con el síndrome hiper-IgM (Garber *et al.*, 1999), el lupus eritematoso sistémico (Citores *et al.*, 2004), la malaria (Sabeti *et al.*, 2002), complicaciones con trasplantes de riñón, tuberculosis (Campbell *et al.*, 2003), la enfermedad de Graves (causa hipertiroidismo por un fallo inmune) (Houston *et al.*, 2004), síndromes de deficiencia inmunológica (Ameratunga y Woon, 2009), reacciones alérgicas cutáneas inducidas por antibióticos (Kim *et al.*, 2009), el pénfigo (enfermedad cutánea relacionada con autoinmunidad) (Malheiros y Petzl-Erler, 2009), bronquiolitis producida por el virus respiratorio sincitial (Siezen *et al.*, 2009), asma inducida por infecciones víricas (Janssen *et al.*, 2007) y artritis reumatoide (Rosado *et al.*, 2008).

4.1.3. Gen *CYSLTR1*

Este gen se encarga de producir receptores para cisteinil leucotrienos en individuos con o sin asma. Los datos de 1000genomes destacan un SNP que causa cambio en la secuencia de la proteína:

- rs320995 (G-A). El cambio no sinónimo sucede en el aminoácido 309 del gen, donde originalmente hay fenilalanina, y puede cambiarse por leucina (F309L). En cuanto a frecuencias genotípicas (Fig. 4.5), el 2.2% de las mujeres presentan homocigosis GG, el 16.5% presentan heterocigosis GA, y el 33.6% son homocigotos AA. Si vemos las frecuencias masculinas, el 9.7% presentan el alelo G, y un 38%, el alelo A.

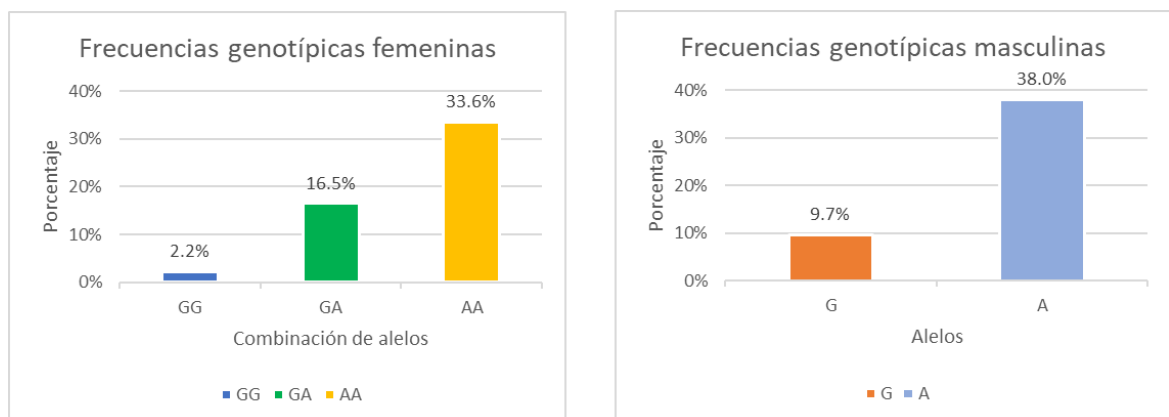


FIGURA 4.5. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs320995, extraído de Ensembl.

Los algoritmos usados no encuentran una predicción para este SNP.

Sin embargo, SIFT y PolyPhen sugieren como problemáticos otros dos SNPs distintos dentro de CYSLTR1:

- rs191425988 (C-T). Tiene una puntuación asignada de 0.04 en SIFT, y 0.999 en PolyPhen, indicando un cambio de valina a metionina en la posición 277 (V277M), indicando en ambos algoritmos que el cambio es dañino.
- rs148732852 (C-G). La puntuación, en este caso, es de 0.38 en SIFT, y 0.926 en PolyPhen, cambiando serina por treonina en la posición 245 (S245T). Aunque SIFT menciona que es tolerable este cambio, PolyPhen dice lo contrario.

En ambos SNPs, la presencia de los dos alelos es casi nula, solo predominando el alelo C en ambos casos.

Este gen se asocia con hiperreactividad bronquial (Hong *et al.*, 2007), bronquiolitis causada por el virus respiratorio sincitial (Schuurhof *et al.*, 2010), hipersensibilidad a medicamentos (Kim *et al.*, 2007) y rinosinusitis (Al-Shemari *et al.*, 2008).

4.1.4. Gen TLR7

Este gen es uno de los más importantes para el sistema inmune, pues participa de forma importante en la respuesta innata. Los datos de 1000genomes destacan un SNP que causa cambio en la proteína:

- rs179008 (A-T). El cambio no sinónimo sucede en el aminoácido 11 del gen, donde puede cambiar glutamina por leucina (Q11L).

En cuanto a frecuencias genotípicas (Fig. 4.6) femeninas, el 30% homocigosis AA, el 19.5%, heterocigosis AT, y el 2.8%, homocigosis TT. Viendo las frecuencias en varones, el 37.6% tienen el alelo A, y el 10.1%, el alelo T.

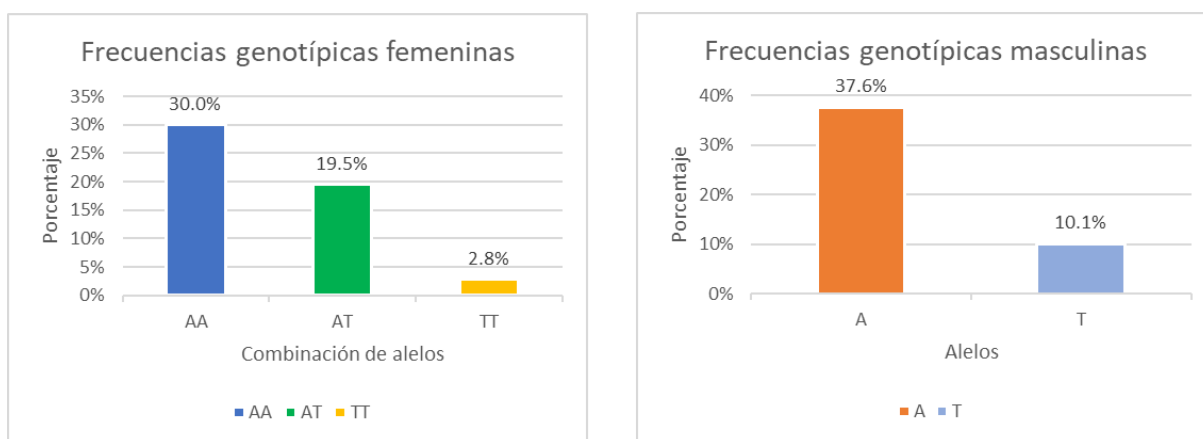


FIGURA 4.6. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs179008, extraído de Ensembl.

Solo el algoritmo SIFT asocia una puntuación de 0.25, prediciendo tolerancia.

Tanto ClinVar, como SIFT y PolyPhen predicen más SNPs problemáticos, pero no se tendrán en cuenta por presentar puntuaciones bajas.

Este gen está asociado sobre todo con infecciones por VIH (Bochud *et al.*, 2007), pero también pueden relacionarse con fibrosis crónica causada por hepatitis C (Schott *et al.*, 2007), lupus eritematoso sistémico (Sanchez *et al.*, 2009), bronquiolitis sinticial (Siezen *et al.*, 2009), causar nacimientos prematuros (Velez *et al.*, 2009), enfermedad celiaca (Dubois *et al.*, 2010), esclerosis múltiple (Enevold *et al.*, 2010), fiebre hemorrágica del Dengue (Silva *et al.*, 2010) y artritis reumatoide (Enevold *et al.*, 2010).

4.1.5. Gen TLR8

Junto al gen anterior, son de los más importantes en el sistema inmunitario por su participación en la respuesta innata. El proyecto 1000genomes nos permitió identificar un SNP que causa cambio en la secuencia de la proteína:

- rs5744080 (C-T). El cambio no sinónimo sucede en el aminoácido 215 del gen, donde puede cambiar histidina por glutamina (H215Q).

En cuanto a frecuencias genotípicas (Fig. 4.7), el 22.3% de los mujeres estudiadas son homocigotas CC, el 22.5% presentan heterocigosis CT, y el 7.6% son homocigotas TT. En cuanto a los varones, el 33.3% presentan el alelo C, y el 16.2%, el alelo T.

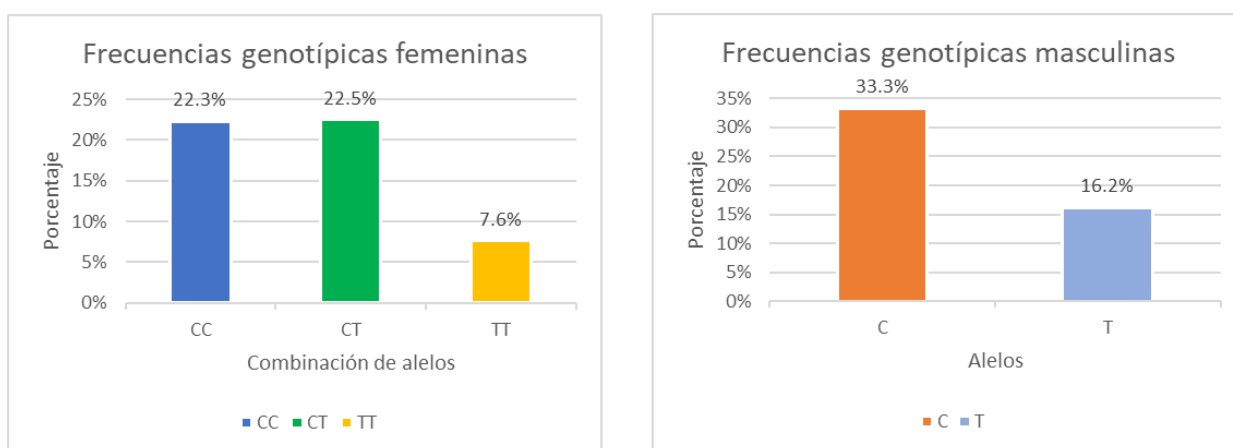


FIGURA 4.7. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs5744080, extraído de Ensembl.

Ninguno de los algoritmos usados detectó un efecto funcional problemático para este SNP.

El algoritmo SIFT predice un SNP (rs3764880, G-A), que puede cambiar metionina por valina en la posición 1 (M1V), asignando una puntuación de 0.32, pero ambos alelos apenas están presentes en la población (solo A).

En cuanto a enfermedades asociadas, este gen se relaciona de igual modo con infecciones de VIH (Bochud *et al.*, 2007), aunque también pudieron relacionarlos con bronquiolitis sinticial (Siezen *et al.*, 2009), causar nacimientos prematuros (Velez *et al.*, 2009), tuberculosis pulmonar (Davila *et al.*, 2008), enfermedad celiaca (Dubois *et al.*, 2010), esclerosis múltiple (Enevold *et al.*, 2010), fiebre hemorrágica del Dengue (Silva *et al.*, 2010), artritis reumatoide (Enevold *et al.*, 2010).

4.1.6. Gen XIAP

El último gen se encarga de generar diferentes proteínas apoptóticas, además de intervenir en la inmunidad a través de la inducción de una respuesta inflamatoria. El proyecto 1000genomes destaca un SNP causante de un cambio no sinónimo en la proteína:

- rs5956583 (A-C). El cambio no sinónimo sucede en el aminoácido 423 del gen, donde puede cambiar glutamina por prolina (Q423P).

En cuanto a frecuencias genotípicas (Fig. 4.8), las mujeres presentan un 19.3% de homocigosis AA, un 24.7% de heterocigosis AC, y un 8.3% de homocigotos CC. En cuanto a los varones, el 43.4% presentan el alelo A, y un 16.5% presenta el alelo A.

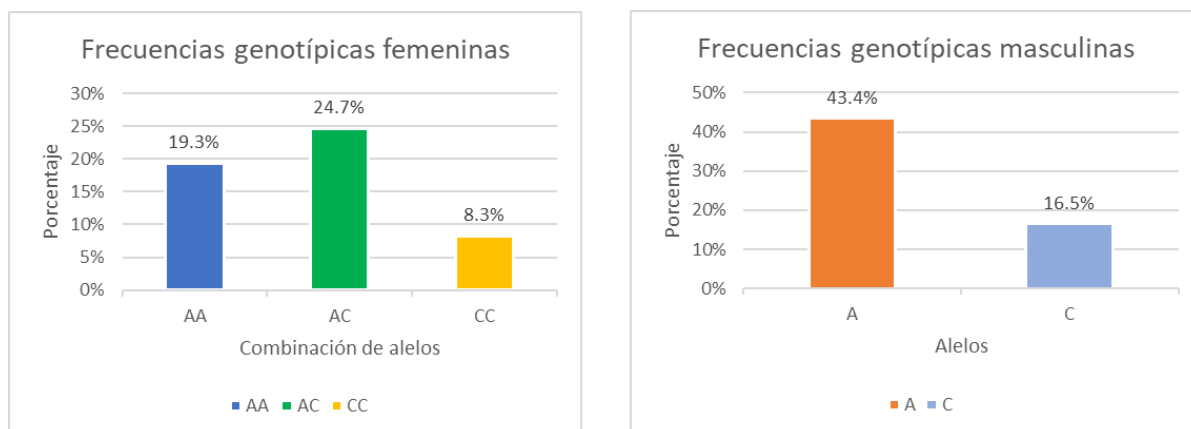


FIGURA 4.8. Representación gráfica de las frecuencias genotípicas europeas femeninas (izquierda) y masculinas (derecha) de rs5956583, extraído de Ensembl.

Todos los algoritmos utilizados asocian dicho cambio a ser benigno, aportando puntuaciones de 0.09 en SIFT, y de 0.001 en PolyPhen.

Otros SNPs son predichos como problemáticos por los algoritmos, pero que no poseen unos valores elevados en sus puntuaciones asignadas, por lo que tampoco se tendrán en cuenta.

Si observamos las enfermedades inmunes que están asociadas a este gen, podemos ver una asociación con síndromes de deficiencia inmunológica (Ameratunga y Woon, 2009), o con desórdenes linfoproliferativos (Salzer *et al.*, 2008)

4.2. SNPs que afectan en la expresión del ARNm

Posteriormente se buscaron los mismos genes en la página web GTEx Portal con el fin de encontrar SNPs que pudieran regular la expresión de sus ARNm. Concretamente miramos los SNPs que se expresaban en sangre, bazo y pulmón, observando y buscando aquellos con un p-valor lo más bajo posible.

4.2.1. Gen CCDC22

Lo primero que podemos encontrar, es que existe una mayor TPM (transcripción por millón de kilobases) en el bazo que en el resto de tejidos (Fig. 4.9). Sin embargo, la mayoría de SNPs que encontramos con p-valores muy inferiores, se encuentran en sangre.

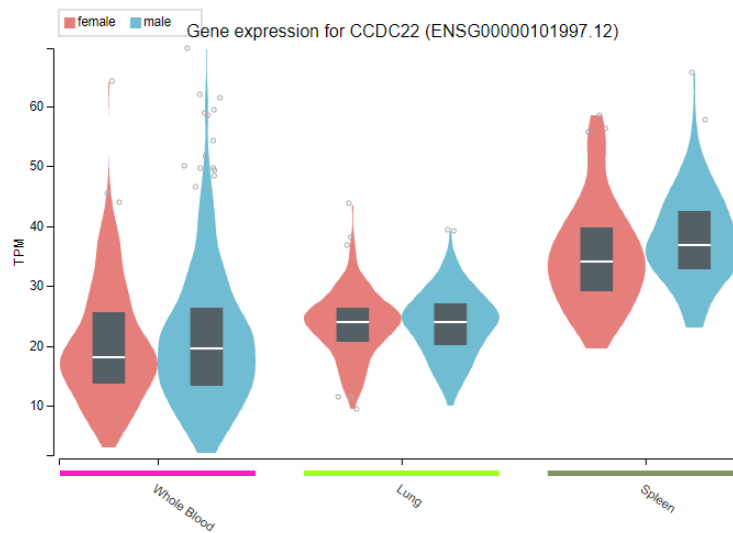


FIGURA 4.9. Representación gráfica de la expresión del gen CCDC22 en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

El SNP de ejemplo que veremos es rs2075866 (G-A), que se encuentra en el gen CACNA1F (situado aguas abajo de CCDC22), posee un p-valor de $4.4 \cdot 10^{-17}$, y cuyos alelos que lo componen son bastante frecuentes en la población europea. Si vemos las frecuencias genotípicas separadas por sexos (Fig. 4.10), vemos que en mujeres, un 4% presentan homocigosis GG, un 20.9% presentan heterocigosis GA, y un 27.4%, homocigosis AA. Viendo a los varones, un 32.2%, el alelo G, y un 14.1%, el alelo A.

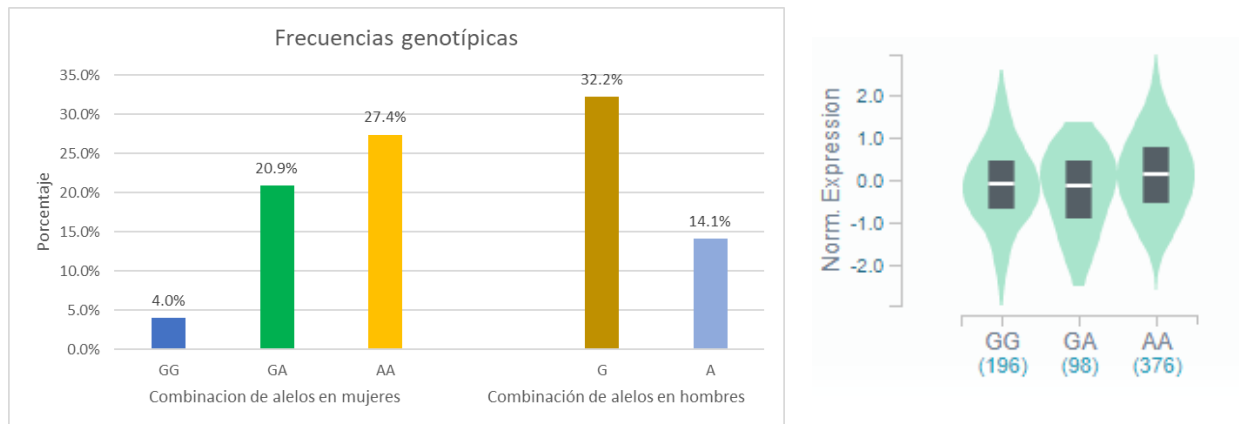


FIGURA 4.10. Representación gráfica de las frecuencias genotípicas europeas (izquierda) de rs2075866, extraído de Ensembl, junto a los niveles de expresión (derecha), extraído de GTEx Portal.

4.2.2. Gen CD40LG

En este caso, la expresión de este gen se da con mucho menor TPM en los tejidos seleccionados (Fig. 4.11). Podemos encontrar algunos SNPs en pulmón, pero solo predomina un alelo frente a otro en la población europea, por lo que no merece la pena mencionarlos.

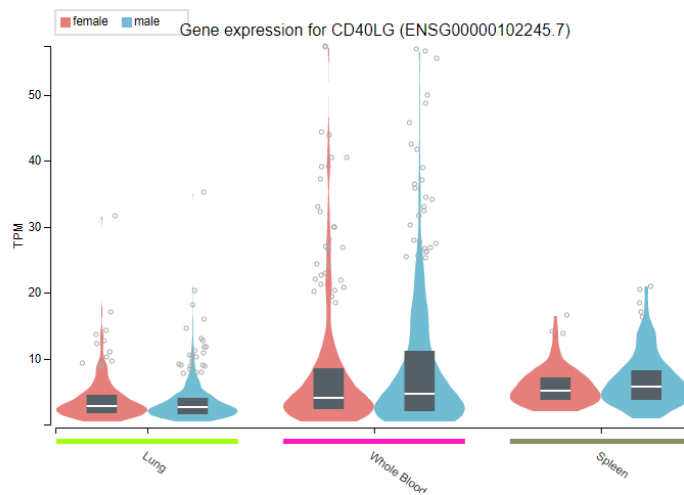


FIGURA 4.11. Representación gráfica de la expresión del gen CD40LG en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

4.2.3. Gen *CYSLTR1*

La expresión de este gen parece tener unos valores mayores de TPM en bazo (Fig. 4.12). Sin embargo, los SNPs más significativos de los tejidos estudiados se encuentran en pulmón.

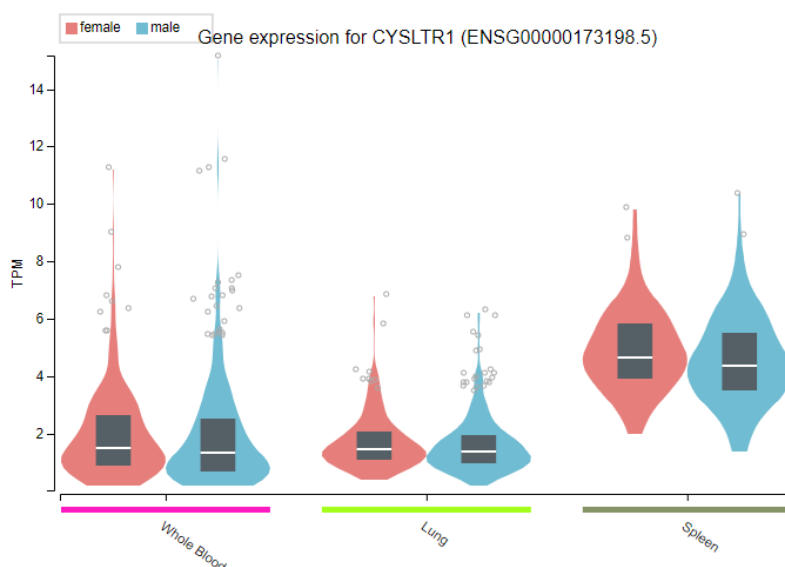


FIGURA 4.12. Representación gráfica de la expresión del gen *CYSLTR1* en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

El SNP de ejemplo que veremos es rs140162383 (C-T), que es una variante intergénica (aguas abajo de *CYSLTR1*) con un p-valor de $1.2 \cdot 10^{-6}$. Las frecuencias genotípicas (Fig. 4.13) que se encuentran en la población europea también se pueden diferenciar por sexos, viendo que las mujeres presentan un 33.6% de homocigosis CC, un 16.3%, heterocigosis CT, y un 2.4%, homocigosis TT, mientras que los varones, el 37.8% tienen el alelo C, y el 9.9%, el alelo T.

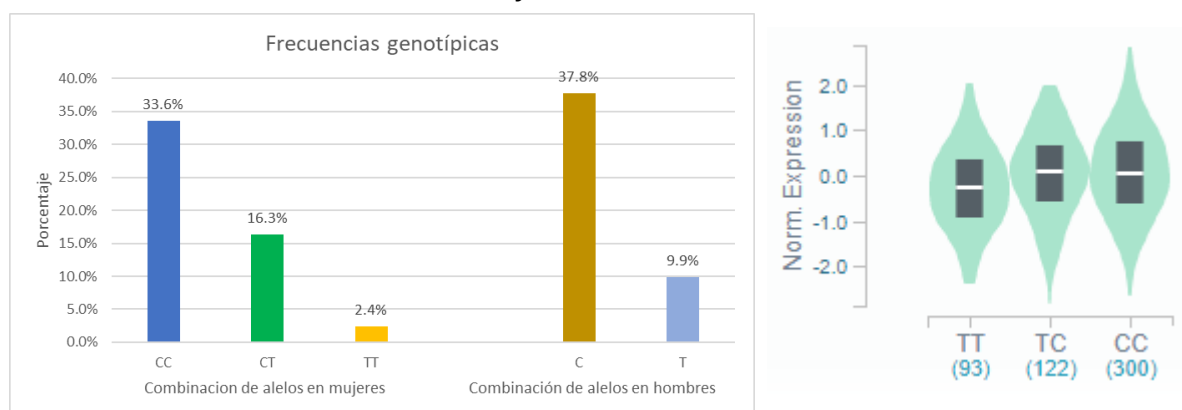


FIGURA 4.13. Representación gráfica de las frecuencias genotípicas europeas (izquierda) de rs140162383, extraído de Ensembl, junto a los niveles de expresión (derecha), extraído de GTEx Portal.

4.2.4. Gen TLR7

Igualmente vemos una mayor TPM en bazo (Fig. 4.14), aunque podemos encontrar SNPs en cualquiera de los tres tejidos con valores de p bastante bajos.

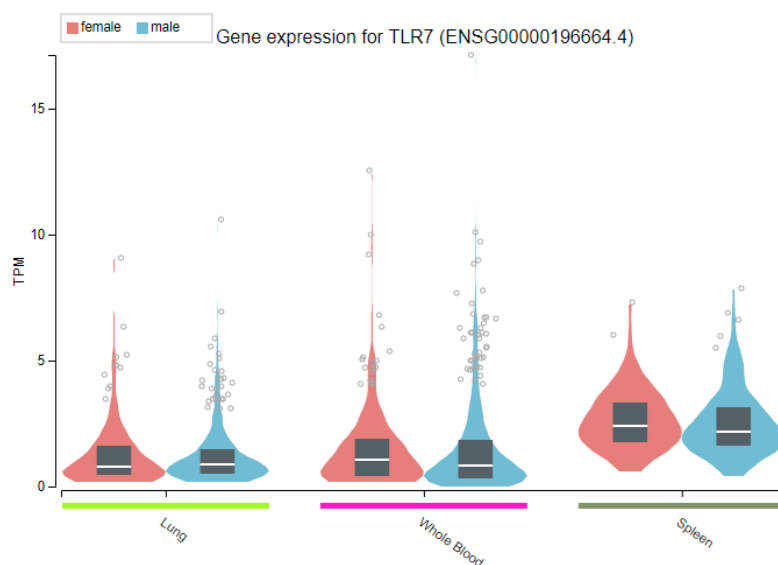


FIGURA 4.14. Representación gráfica de la expresión del gen TLR7 en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

El ejemplo que vemos es rs2897827 (C-T), siendo también una variante intergénica (aguas abajo de TLR7), pero con un p-valor de $4.2 \cdot 10^{-18}$. Las frecuencias genotípicas (Fig. 4.15) que se encuentran en la población europea también se pueden diferenciar por sexos, viendo que las mujeres presentan un 46.5% de homocigosis CC, un 5.6%, heterocigosis CT, y un 0.2%, homocigosis TT, mientras que los varones, el 44.3% tienen el alelo C, y el 3.4%, el alelo T.

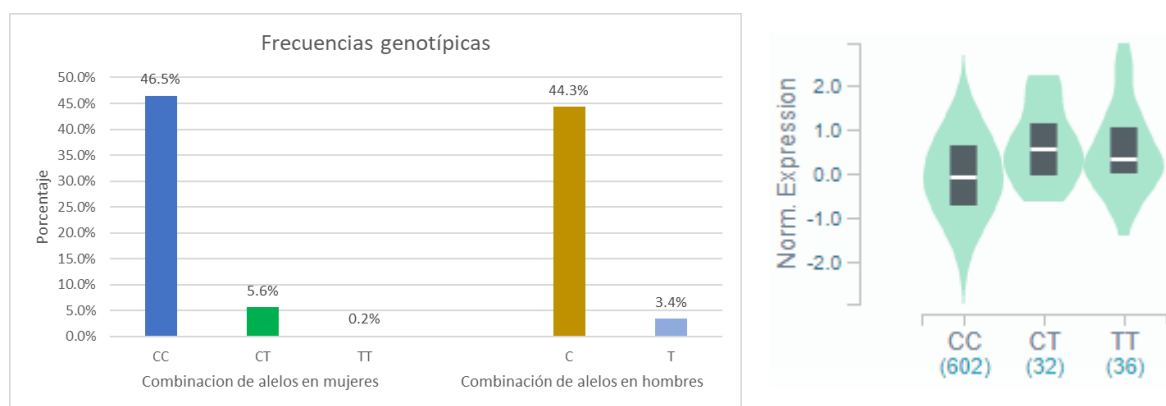


FIGURA 4.15. Representación gráfica de las frecuencias genotípicas europeas (izquierda) de rs2897827, extraído de Ensembl, junto a los niveles de expresión (derecha), extraído de GTEx Portal.

4.2.5. Gen TLR8

La principal expresión de este gen se da en sangre (Fig. 4.16). Observando los demás tejidos, apenas hay TPM en pulmón, y en bazo es bastante bajo.

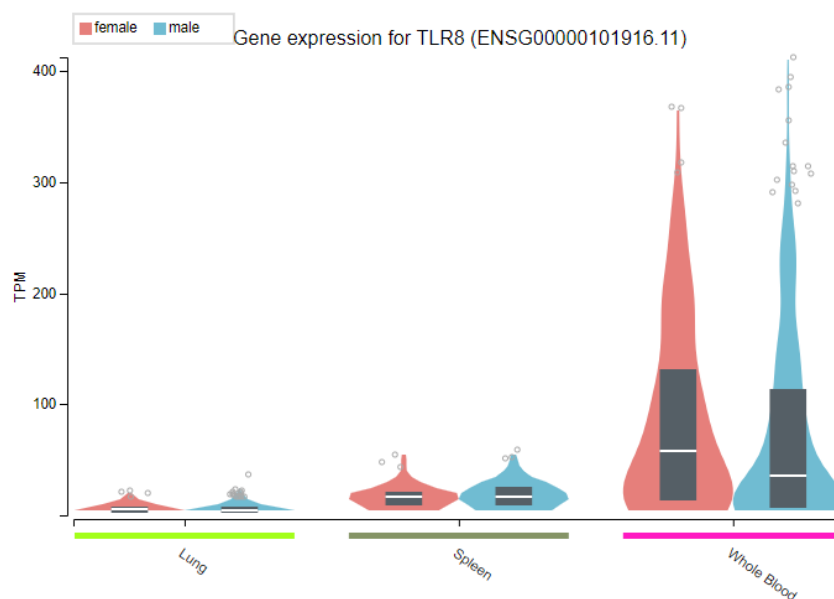


FIGURA 4.16. Representación gráfica de la expresión del gen TLR8 en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

El ejemplo que vemos es rs180726824 (G-A), que también es una variante intergénica (aguas arriba de TLR8), pero con un p-valor de $5.0 \cdot 10^{-15}$. Las frecuencias genotípicas (Fig. 4.17) que se encuentran en la población europea se pueden diferenciar por sexos, viendo que las mujeres presentan un 27.2% de homocigosis GG, un 21.7%, heterocigosis GA, y un 3.4%, homocigosis AA, mientras que los varones, el 32.8% tienen el alelo G, y el 14.9%, el alelo A.

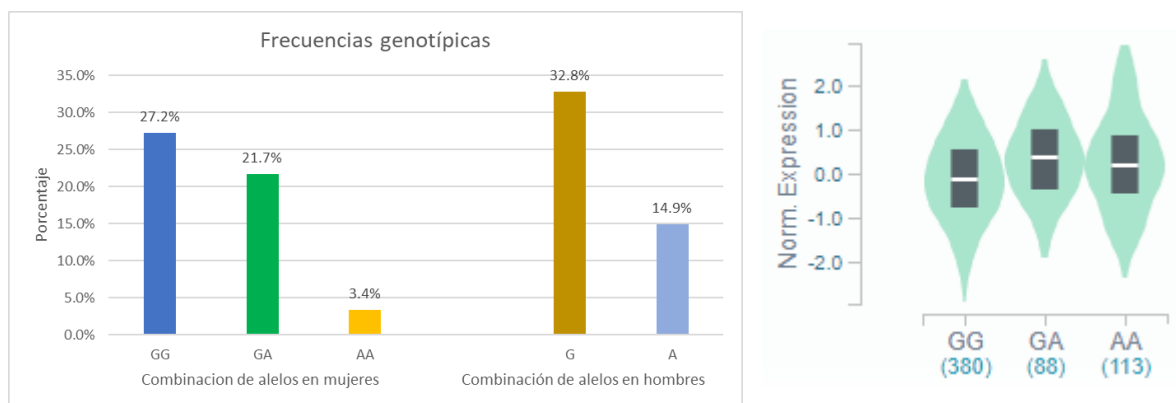


FIGURA 4.17. Representación gráfica de las frecuencias genotípicas europeas (izquierda) de rs180726824, extraído de Ensembl, junto a los niveles de expresión (derecha), extraído de GTEx Portal.

4.2.6. Gen XIAP

El valor TPM mayor se da en pulmón (Fig. 4.18), aunque la mayoría de SNPs con alta significación se encuentra en sangre.

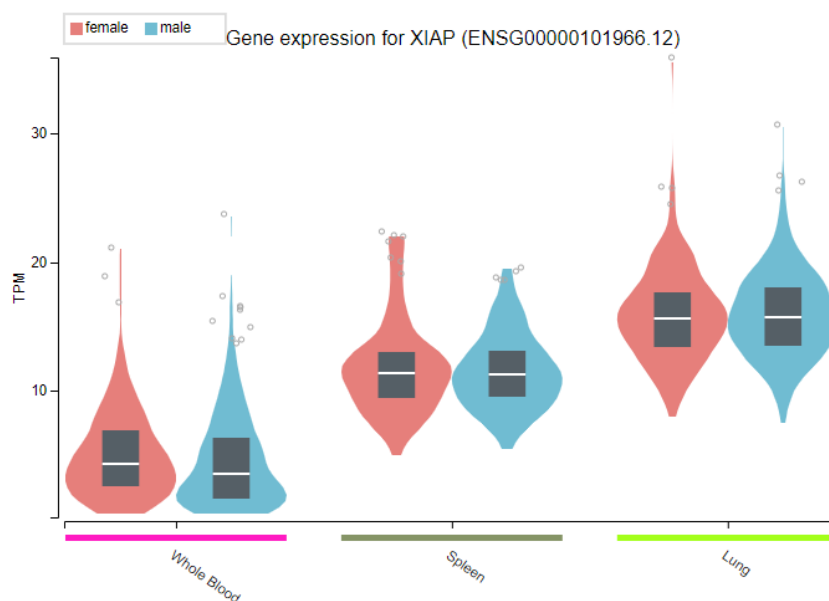


FIGURA 4.18. Representación gráfica de la expresión del gen XIAP en función del tejido donde se exprese, separando los resultados en mujeres (naranja) y varones (azul). Imagen extraída de GTEx Portal.

El ejemplo que vemos es rs7876441 (C-T), una variante intrónica de XIAP, con un p-valor de $9.9 \cdot 10^{-7}$. Las frecuencias genotípicas (Fig. 4.19) que se encuentran en la población europea se pueden diferenciar por sexos, viendo que las mujeres presentan un 8.5% de homocigosis CC, un 24.5%, heterocigosis CT, y un 19.3%, homocigosis TT, mientras que los varones, el 16.5% tienen el alelo C, y el 31.2%, el alelo T.

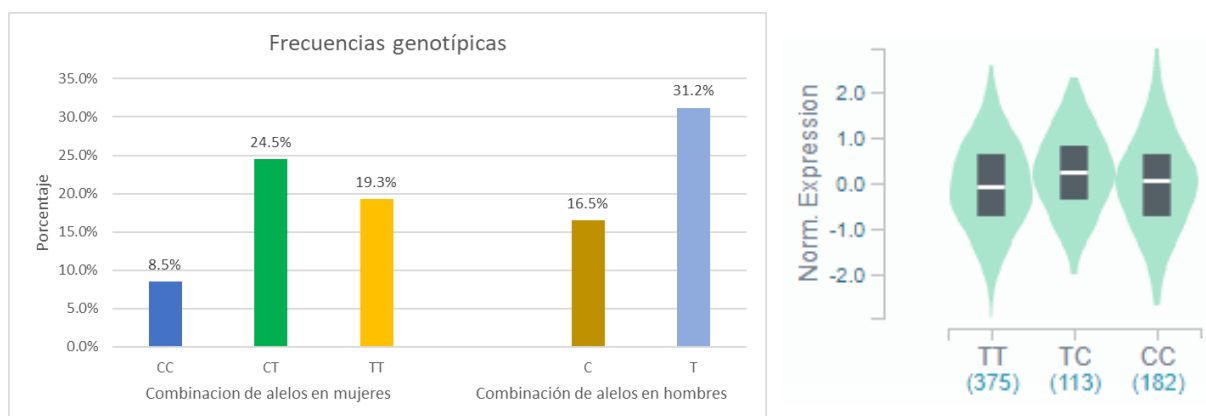


FIGURA 4.19. Representación gráfica de las frecuencias genotípicas europeas (izquierda) de rs7876441, extraído de Ensembl, junto a los niveles de expresión (derecha), extraído de GTEx Portal.

4.3. Resumen de los resultados

Tras desarrollar los resultados, realizamos la siguiente tabla (Tabla 4.1) que resumiera los genes candidatos que llegamos a recalcar, junto a los SNPs potenciales que componen dichos genes:

TABLA 4.1. Tabla que resume todos los SNPs mencionados en el apartado de Resultados.

GENES						
	CCDC22	CD40LG	CYSLTR1	TLR7	TLR8	XIAP
Cambios no Sinónimos						
Según 1000genomes	rs143790434	rs148594123	rs320995	rs179008	rs5744080	rs5956583
	rs2294020					
Según SIFT/PolyPhen	rs144632022	rs186774716	rs191425988		rs3764880	
	rs147222955		rs148732852			
Influencia en la Expresión de ARNm (ejemplos)						
	rs2075866		rs140162383	rs2897827	rs180726824	rs7876441

5. DISCUSIÓN

Antes de aplicar filtros, pudimos identificar más de 7000 SNPs en esos 45 genes que analizamos, pero muchos de ellos no tenían ninguna patología asociada, y la mayoría no tenían una prevalencia en la población del más del 1%. Por otro lado, algunos SNPs que se relacionaban con enfermedades, no eran precisamente enfermedades inmunitarias, por lo que también se acabaron descartando. Finalmente, escogiendo solo los SNPs que influían con un cambio no sinónimo, pudimos llegar a un total de 7 SNPs potenciales en 6 genes candidatos.

De todos esos genes, llaman especialmente la atención los siguientes genes:

- TLR7 y TLR8, por participar en la ruta de señalización de receptores “Toll-Like” (Fig. 5.1), ya que ambos son receptores capaces de detectar ARN monocatenario procedentes de virus, que posean un alto contenido en G y C (Lester y Li, 2014). Un mal funcionamiento de estos receptores ocasionaría una mayor facilidad por los virus para infectar al individuo.
- CD40LG, por participar en la ruta de señalización de la necrosis tumoral Kappa B (Fig. 5.1), siendo los principales receptores para que se puedan activar de forma adecuada los linfocitos B, y que así estos produzcan anticuerpos (Lesley *et al.*, 2006). Una anomalía en este receptor causaría una ineficiente respuesta inmunitaria adaptativa.

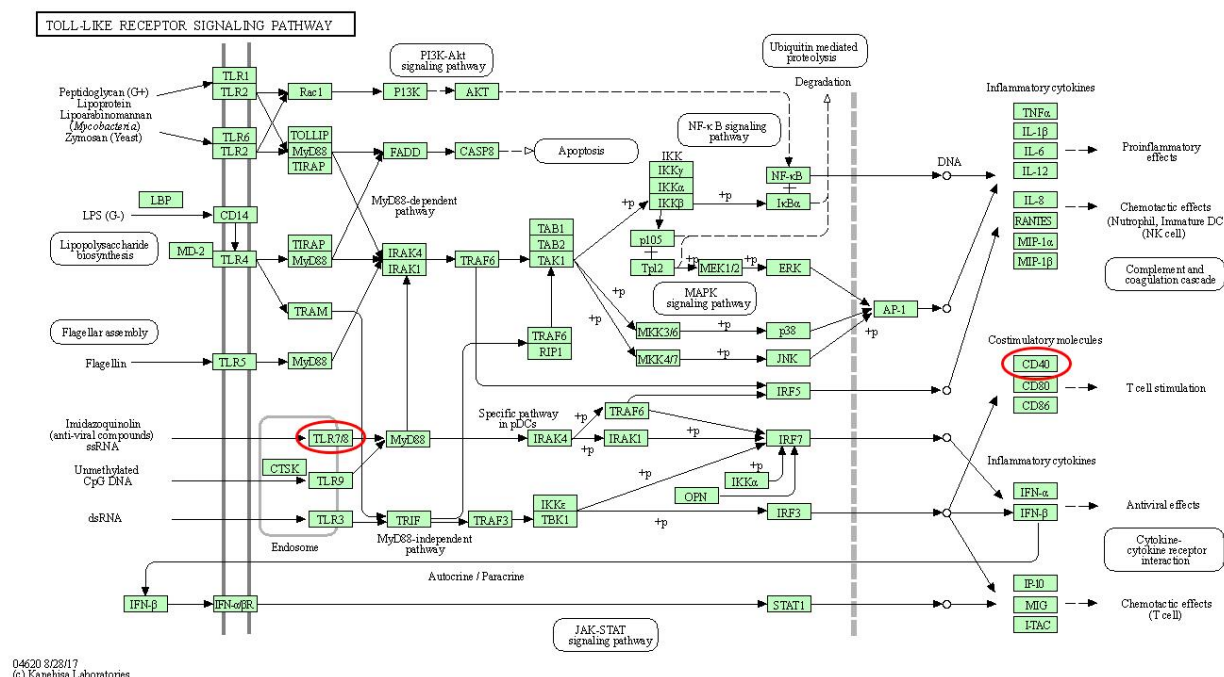


FIGURA 5.1. Representación gráfica de la ruta de señalización de receptores “Toll-Like”, donde se ve la participación de los genes TLR7, TLR8 y CD40LG, extraída de NIH-DAVID (Huang *et al.*, 2009)

Es por ello que estos genes tienen una gran cantidad de estudios asociándolos a diversas enfermedades, entre ellas infecciones por diferentes virus, como el VIH o la bronquiolitis. De hecho, debido a la importancia que está teniendo el SARS-CoV-2 en 2020-2021, ya se está tratando de relacionar la identificación de ARN monocatenario de dicho virus por los receptores TLR7 y TLR8 (Moreno-Eutimio *et al.*, 2020), ya que una mala producción de estos podrían agravar las consecuencias. El hecho de que las mujeres tengan mayor dosis génica del cromosoma X puede impedir que la carga viral sea muy alta, pero los hombres, al solo tener un cromosoma X, sufren más problemas con una infección viral aguda (Conti y Younes, 2020).

Además, el SNP resaltado del gen CD40LG tiene valores en los algoritmos de SIFT y PolyPhen, lo que refuerza más su importancia en el cambio de proteína producido. Este gen está muy estudiado, pudiendo ver que es el gen con la mayor heterogeneidad clínica (es el gen con más asociaciones a enfermedades comparando con el resto de genes que hemos estudiado), viendo que participa, por ejemplo, en la protección frente a infecciones como la malaria (Sabeti *et al.*, 2002), o que una mutación en este pueda producir el síndrome de hiper-IgM (Notarangelo *et al.*, 1996), entre otras enfermedades.

El hecho de poseer la alteración de un SNP concreto incide en la probabilidad de tener dicha enfermedad asociada a ese gen, pero podemos apreciar que más de un gen está asociado a la misma enfermedad, causando lo que se denomina como heterogeneidad génica. Un ejemplo es la bronquiolitis causada por el virus respiratorio sinticial, que se ha relacionado con la mayoría de los genes de los que hemos hablado en los resultados (el gen CD40LG, CYSTR1, TLR7 y TLR8 están relacionados) (Siezen *et al.*, 2009). Si un individuo tiene diferentes genes candidatos con los SNPs asociados a la enfermedad, dicha persona tendrá más probabilidad a contraer la enfermedad.

Relacionando estos datos con la actualidad, aunque el cromosoma X se ha tomado en cuenta a la hora de realizar el GWAS que se está realizando para estudiar el riesgo de contraer el SARS-CoV-2, si tenemos en cuenta los valores de p que ellos consideran para que sean resultados significativos ($p\text{-valor} < 10^{-8}$) (Ganna, 2021), no se ha encontrado significación en ninguno de los SNPs que mencionamos en este estudio. Aun así, debemos mencionar que la cantidad de datos que dicho GWAS tiene con respecto al cromosoma X, es mucho menor al que tiene con los demás cromosomas autosómicos.

Respecto a los datos obtenidos de GTEx Portal, es curioso ver cómo SNPs que pertenecen a otro gen distinto, pueden estar asociados con la expresión del ARNm de alguno de estos genes candidatos dependiendo del lugar donde se expresen. Es por ello que se crearon los análisis eQTLs, viendo así el papel tan importante que tienen los SNPs a la hora de regular la expresión de los genes humanos en general (Lalonde *et al.*, 2011). Esto amplía más el abanico de posibilidades que hay para aumentar la predisposición a contraer una enfermedad, pues ya no solo hay que tener en cuenta los genes candidatos directamente, sino los demás genes que los regulan.

6. CONCLUSIONES

A lo largo de este estudio, podemos sacar las siguientes conclusiones:

- Se encontraron 7 SNPs distribuidos en 6 genes candidatos.
- Destacamos la importancia de los genes CD40LG, TLR7 y TLR8 como principales genes con un gran número de enfermedades asociadas a ellos.
- SNPs intrónicos dentro de los genes candidatos pueden intervenir en la expresión del ARNm alterando el splicing o causando un cambio en la pauta de lectura.
- SNPs intrónicos de otros genes cercanos a los candidatos, regulan la expresión del ARNm que se crea a partir de los genes candidatos mencionados.
- SNPs intergénicos, es decir, que están situados entre dos genes, también intervienen en la expresión del ARNm de los genes candidatos.
- La predisposición a una enfermedad infecciosa aumenta cuanto mayor sea la cantidad de SNPs que afecten a la secuencia de la proteína y a la expresión del ARNm.
- Un estudio de asociación basado en el genotipado de estos SNPs sería interesante en el contexto de enfermedades infecciosas para las cuales haya descrita diferencias importantes de la gravedad en función del sexo.

7. REFERENCIAS

- AGUET, F.; A. N. BARBEIRA; R. BONAZZOLA; A. BROWN; S. E. CASTEL; B. JO; S. KASELA; S. KIM-HELLMUTH; Y. LIANG; M. OLIVA; et al. (2020) "The GTEx Consortium atlas of genetic regulatory effects across human tissues". *Science*, 369(6509), pp. 1318-1330.
- AL-SHEMARI, H.; Y. BOSSÉ; T. J. HUDSON; M. CABALUNA; M. DUVAL; M. LEMIRE; S. VALLEE-SMEDJA; S. FRENKIEL y M. DESROSIERS (2008) "Influence of leukotriene gene polymorphisms on chronic rhinosinusitis". *BMC Medical Genetics*, 9(1), pp. 1-9.
- ALBERT, M.; A. LANKESTER y A. GENNERY (2018) "Primary immunodeficiencies". *The EBMT Handbook: Hematopoietic Stem Cell Transplantation and Cellular Therapies*, . p.663-670, pp. 663-670.
- AMERATUNGA, R. y S. T. WOON (2009) "Customised molecular diagnosis of Primary Immune Deficiency disorders in New Zealand: An efficient strategy for a small developed country". *New Zealand Medical Journal*, 122(1304), pp. 46-53.
- AUGUI, S.; E. P. NORA y E. HEARD (2011) "Regulation of X-chromosome inactivation by the X-inactivation centre". *Nature Reviews Genetics*, pp. 429-442.
- AUTON, A.; G. R. ABECASIS; D. M. ALTSHULER; R. M. DURBIN; D. R. BENTLEY; A. CHAKRAVARTI; A. G. CLARK; P. DONNELLY; E. E. EICHLER; P. FLICEK; et al. (2015) "A global reference for human genetic variation". *Nature*, pp. 68-74.
- BALATON, B. P.; T. DIXON-MCDOUGALL; S. B. PEETERS y C. J. BROWN (2018) "The eXceptional nature of the X chromosome". *Human molecular genetics*, 27(2), pp. 42-49.
- BAMSHAD, M. J.; S. B. NG; A. W. BIGHAM; H. K. TABOR; M. J. EMOND; D. A. NICKERSON y J. SHENDURE (2011) "Exome sequencing as a tool for Mendelian disease gene discovery". *Nature Reviews Genetics*, pp. 745-755.
- BECKER, J.; O. SEMLER; C. GILISSEN; Y. LI; H. J. BOLZ; C. GIUNTA; C.

BERGMANN; M. ROHRBACH; F. KOERBER; K. ZIMMERMANN; et al. (2011) "Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta". *American Journal of Human Genetics*, 88(3), pp. 362-371.

BEHJATI, S. y P. S. TARPEY (2013) "What is next generation sequencing?". *Archives of Disease in Childhood: Education and Practice Edition*, 98(6), pp. 236-238.

BELMONT, J. W.; A. BOUDREAU; S. M. LEAL; P. HARDENBOL; S. PASTERNAK; D. A. WHEELER; T. D. WILLIS; F. YU; H. YANG; Y. GAO; et al. (2005) "A haplotype map of the human genome". *Nature*, 437(7063), pp. 1299-1320.

BERTAKIS, K. D.; R. AZARI; L. J. HELMS; E. J. CALLAHAN y J. A. ROBBINS (2000) "Gender Differences in the Utilization of Health Care Services". *Journal of Family Practice*, 49(2), pp. 147-152.

BIESECKER, L. G.; K. V. SHIANNA y J. C. MULLIKIN (2011) "Exome sequencing: The expert view". *Genome Biology*, 12(9), pp. 128.

BIRNEY, E. y N. SORANZO (2015) "Human genomics: The end of the start for population sequencing". *Nature*, pp. 52-53.

BOCHUD, P. Y.; M. HERSBERGER; P. TAFFÉ; M. BOCHUD; C. M. STEIN; S. D. RODRIGUES; T. CALANDRA; P. FRANCIOLI; A. TELENTI; R. F. SPECK; et al. (2007) "Polymorphisms in Toll-like receptor 9 influence the clinical course of HIV-1 infection". *AIDS*, 21(4), pp. 441-446.

BOTSTEIN, D. y N. RISCH (2003) "Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease". *Nature Genetics*, pp. 228-237.

BROWN, T. (2008) *Genomas / Genome*, Editorial Médica Panamericana., pp. 4-5.

BUSSONE, G. y L. MOUTHON (2008) "Autoimmune manifestations in primary immune deficiencies". *Autoimmunity Reviews*, 8, pp. 332-336.

CAMPBELL, S. J.; P. SABETI; K. FIELDING; J. SILLAH; B. BAH; P. GUSTAFSON; K. MANNEH; I. LISSE; G. SIRUGO; R. BELLAMY; et al. (2003) "Variants of the CD40 ligand gene are not associated with increased susceptibility to tuberculosis in West Africa". *Immunogenetics*, 55(7), pp. 502-507.

CHANG, C. C.; C. C. CHOW; L. C. A. M. TELLIER; S. VATTIKUTI; S. M. PURCELL y J. J. LEE (2015) "Second-generation PLINK: Rising to the challenge of larger and richer datasets". *GigaScience*, 4(1), pp. s13742-015.

CHEN, L.; Q. LI; D. ZHENG; H. JIANG; Y. WEI; L. ZOU; L. FENG; G. XIONG; G. SUN; H. WANG; et al. (2020) "Clinical Characteristics of Pregnant Women with Covid-19 in Wuhan, China". *New England Journal of Medicine*, 382(25), pp. 100.

CHIAL, H. (2008) "Polygenic Inheritance and Gene Mapping | Learn Science at Scitable". *Nature*, 1(1), pp. 17.

CITORES, M. J.; I. RUA-FIGUEROA; C. RODRIGUEZ-GALLEGO; A. DURÁNTEZ; M. I. GARCÍA-LAORDEN; C. RODRÍGUEZ-LOZANO; J. C. RODRÍGUEZ-PÉREZ; J. A. VARGAS y P. PÉREZ-ACIEGO (2004) "The dinucleotide repeat polymorphism in the 3'UTR of the CD154 gene has a functional role on protein expression and is associated with systemic lupus erythematosus". *Annals of the Rheumatic Diseases*, 63(3), pp. 310-317.

CONTI, P. y A. YOUNES (2020) "Coronavirus cov-19/sars-cov-2 affects women less than men: Clinical response to viral infection". *Journal of Biological Regulators and Homeostatic Agents*, 34(2), pp. 339-343.

CONTRERAS, E. (2009) *Mutaciones Genéticas*.

CRAIG VENTER, J.; M. D. ADAMS; E. W. MYERS; P. W. LI; R. J. MURAL; G. G. SUTTON; H. O. SMITH; M. YANDELL; C. A. EVANS; R. A. HOLT; et al. (2001) "The sequence of the human genome". *Science*, 291(5507), pp. 1304-1351.

DAVILA, S.; M. L. HIBBERD; R. H. DASS; H. E. E. WONG; E. SAHIRATMADJA; C. BONNARD; B. ALISJAHBANA; J. S. SZESZKO; Y. BALABANOVA; F. DROBNIEWSKI; et al. (2008) "Genetic association and expression studies indicate a role of Toll-like receptor 8 in pulmonary tuberculosis". *PLoS Genetics*, 4(10).

DAYEM ULLAH, A. Z.; J. OSCANOA; J. WANG; A. NAGANO; N. R. LEMOINE y C. CHELALA (2018) "SNPnexus: Assessing the functional relevance of genetic variation to facilitate the promise of precision medicine". *Nucleic Acids Research*, 46(W1), pp. W109-W113.

DUBOIS, P. C. A.; G. TRYNKA; L. FRANKE; K. A. HUNT; J. ROMANOS; A. CURTOTTI; A. ZHERNAKOVA; G. A. R. HEAP; R. ÁDÁNY; A. AROMAA; et al. (2010) "Multiple common variants for celiac disease influencing immune gene expression". *Nature Genetics*, 42(4), pp. 295-302.

EATON, W. W.; N. R. ROSE; A. KALAYDJIAN; M. G. PEDERSEN y P. B. MORTENSEN (2007) "Epidemiology of autoimmune diseases in Denmark". *Journal of Autoimmunity*, 29(1), pp. 1-9.

ENEVOLD, C.; A. B. OTURAI; P. S. SØRENSEN; L. P. RYDER; N. KOCH-HENRIKSEN y K. BENDTZEN (2010) "Polymorphisms of innate pattern recognition receptors, response to interferon-beta and development of neutralizing antibodies in multiple sclerosis patients". *Multiple Sclerosis*, 16(8), pp. 942-949.

ENEVOLD, C.; T. R. D. RADSTAKE; M. J. H. COENEN; J. FRANSEN; E. J. M. TOONEN; K. BENDTZEN y P. L. C. M. VAN RIEL (2010) "Multiplex screening of 22 single-nucleotide polymorphisms in 7 toll-like receptors: An association study in rheumatoid arthritis". *Journal of Rheumatology*, 37(5), pp. 905-910.

FRUTOS, R. DE (1997) "Elementos transponibles y organización del genoma de eucariotas". *III Simposio Científico en Biología Celular y Molecular*, pp. 175-188.
GANNA, A. (2021) "Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis". *medRxiv*.

GAO, F.; D. CHANG; A. BIDDANDA; L. MA; Y. GUO; Z. ZHOU y A. KEINAN (2015) "XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome". *Journal of Heredity*, 106(5), pp. 666-671.

GARBER, E.; L. SU; B. EHRENFELS; M. KARPUSAS y Y. M. HSU (1999) "CD154 variants associated with hyper-IgM syndrome can form oligomers and trigger CD40-mediated signals". *Journal of Biological Chemistry*, 274(47), pp. 33545-33550.

GILISSEN, C.; H. H. ARTS; A. HOISCHEN; L. SPRUIJT; D. A. MANS; P. ARTS; B. VAN LIER; M. STEEHOUWER; J. VAN REEUWIJK; S. G. KANT; et al. (2010) "Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome". *American Journal of Human Genetics*, 87(3), pp. 418-423.

GILISSEN, C.; A. HOISCHEN; H. G. BRUNNER y J. A. VELTMAN (2012) "Disease gene identification strategies for exome sequencing". *European Journal of Human Genetics*, pp. 490-497.

GRIFFITHS, A. J.; W. M. GELBART; R. C. LEWONTIN y J. H. MILLER (2002) "Modern Genetic Analysis: Integrating Genes and Genomes". *Macmillan*, 1.

HAMOSH, A.; A. F. SCOTT; J. AMBERGER; D. VALLE y V. A. MCKUSICK (2000) "Online Mendelian Inheritance in Man (OMIM)". *Human Mutation*, 15(1), pp. 57-61.

HONG, S. J.; H. BIN KIM; M. J. KANG; S. Y. LEE; J. H. KIM; B. S. KIM; S. O. JANG; H. D. SHIN y C. S. PARK (2007) "TNF- α (-308 G/A) and CD14 (-159T/C) polymorphisms in the bronchial responsiveness of Korean children with asthma". *Journal of Allergy and Clinical Immunology*, 119(2), pp. 398-404.

HOUSTON, F. A.; V. WILSON; C. E. JENNINGS; C. J. OWEN; P. DONALDSON; P. PERROS y S. H. S. PEARCE (2004) "Role of the CD40 locus in Graves' disease". *Thyroid*, 14(7), pp. 506-509.

HOWE, K. L.; P. ACHUTHAN; J. ALLEN; J. ALLEN; J. ALVAREZ-JARRETA; M. RIDWAN AMODE; I. M. ARMEAN; A. G. AZOV; R. BENNETT; J. BHAI; et al. (2021) "Ensembl 2021". *Nucleic Acids Research*, 49(D1), pp. D884-D891.

HOWIE, B.; C. FUCHSBERGER; M. STEPHENS; J. MARCHINI y G. R. ABECASIS (2012) "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing". *Nature Genetics*, 44(8), pp. 955-959.

HUANG, D. W.; B. T. SHERMAN y R. A. LEMPICKI (2009) "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". *Nature Protocols*, 4(1), pp. 44-57.

HUERTA, L. y M. BURKE (2016) "Functional genomics (II): Common technologies and data analysis methods". *EMBL-EBI*, pp. 1-34.

INVERNIZZI, P.; S. PASINI; C. SELMI; M. E. GERSHWIN y M. PODDA (2009) "Female predominance and X chromosome defects in autoimmune diseases". *Journal of Autoimmunity*, pp. 12-16.

JANSSEN, R.; L. BONT; C. L. E. SIEZEN; H. M. HODEMAEKERS; M. J. ERMERS; G. DOORNBOS; R. VAN 'T SLOT; C. WIJMENG; J. J. GOEMAN; J. L. L. KIMPEN; et al. (2007) "Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes". *Journal of Infectious Diseases*, 196(6), pp. 826-834.

JOHNSON, H. D.; D. SHOLCOSKY; K. GABELLO; R. RAGNI y N. OGONOSKY (2003) "Sex differences in public restroom handwashing behavior associated with visual behavior prompts". *Perceptual and Motor Skills*, 97(3 I), pp. 805-810.

KIM, S. H.; J. E. LEE; S. H. KIM; Y. K. JEE; Y. K. KIM; H. S. PARK; K. U. MIN y H. W. PARK (2009) "Allelic variants of CD40 and CD40L genes interact to promote antibiotic-induced cutaneous allergic reactions". *Clinical and Experimental Allergy*, 39(12), pp. 1852-1856.

KIM, S. H.; E. M. YANG; H. J. PARK; Y. M. YE; H. Y. LEE y H. S. PARK (2007) "Differential contribution of the CysLTR1 gene in patients with aspirin hypersensitivity". *Journal of Clinical Immunology*, 27(6), pp. 613-619.

KONO, D. H. y A. N. THEOFILOPOULOS (2017) "Autoimmunity". *Kelley and Firestein's Textbook of Rheumatology*, . p.301-317.e5, pp. 301-317.

LALONDE, E.; K. C. H. HA; Z. WANG; A. BEMMO; C. L. KLEINMAN; T. KWAN; T. PASTINEN y J. MAJEWSKI (2011) "RNA sequencing reveals the role of splicing polymorphisms in regulating human gene expression". *Genome Research*, 21(4), pp. 545-554.

LANDRUM, M. J.; J. M. LEE; M. BENSON; G. R. BROWN; C. CHAO; S. CHITIPIRALLA y E. AL. (2018) "ClinVar: improving access to variant interpretations and supporting evidence". *Nucleic Acids Research*, 46(D1), pp. D1062-D1067.

LESLEY, R.; L. M. KELLY; Y. XU y J. G. CYSTER (2006) "Naive CD4 T cells constitutively express CD40L and augment autoreactive B cell survival". *Proceedings of the National Academy of Sciences of the United States of America*, 103(28), pp. 10717-10722.

LESTER, S. N. y K. LI (2014) "Toll-like receptors in antiviral innate immunity". *Journal of Molecular Biology*, pp. 1246-1264.

LIBERT, C.; L. DEJAGER y I. PINHEIRO (2010) "The X chromosome in immune functions: When a chromosome makes the difference". *Nature Reviews Immunology*, pp. 594-604.

LYON, M. F. (1961) "Gene action in the X-chromosome of the mouse (*mus musculus* L.)". *Nature*, 190(4773), pp. 372-373.

MAJEWSKI, J.; J. SCHWARTZENTRUBER; E. LALONDE; A. MONTPETIT y N. JABADO (2011) "What can exome sequencing do for you?". *Journal of Medical Genetics*, pp. 580-589.

MALHEIROS, D. y M. L. PETZL-ERLER (2009) "Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus". *Genes and Immunity*, 10(6), pp. 547-558.

MANOLIO, T. A. (2010) "Genomewide Association Studies and Assessment of the Risk of Disease". *New England Journal of Medicine*, FEERO, W. G. y A. E. GUTTMACHER (EDS.), 363(2), pp. 166-176.

MANSEGO, M. L.; R. ABELLÁN y F. J. CHAVES (2007) "Tipos de mutaciones y polimorfismos. Posibles aplicaciones en terapia génica en la diabetes.". *avancesendiabetologia.org*, 23(6), pp. 425-431.

MARDIS, E. R. (2011) "A decade's perspective on DNA sequencing technology". *Nature*, pp. 198-203.

MCCLELLAN, J. y M. C. KING (2010) "Genetic heterogeneity in human disease". *Cell*, pp. 210-217.

MEESTER, I.; E. MANILLA-MUÑOZ; R. B. R. LEÓN-CACHÓN; G. A. PANIAGUA-FRAUSTO; D. CARRIÓN-ALVAREZ; C. O. RUIZ-RODRÍGUEZ; X. RODRÍGUEZ-RANGEL y J. M. GARCÍA-MARTÍNEZ (2020) "SeXY chromosomes and the immune system: Reflections after a comparative study". *Biology of Sex Differences*, 11(1), MITCHELL, K. J. y D. J. PORTEOUS (2011) "Rethinking the genetic architecture of schizophrenia". *Psychological Medicine*, 41(1), pp. 19-32.

MORENO-EUTIMIO, M. A.; C. LÓPEZ-MACÍAS y R. PASTELIN-PALACIOS (2020) "Bioinformatic analysis and identification of single-stranded RNA sequences recognized by TLR7/8 in the SARS-CoV-2, SARS-CoV, and MERS-CoV genomes". *Microbes and Infection*, 22(4-5), pp. 226-229.

NG, S. B.; K. J. BUCKINGHAM; C. LEE; A. W. BIGHAM; H. K. TABOR; K. M. DENT; C. D. HUFF; P. T. SHANNON; E. W. JABS; D. A. NICKERSON; et al. (2010) "Exome sequencing identifies the cause of a mendelian disorder". *Nature Genetics*, 42(1), pp. 30-35.

NG, S. B.; D. A. NICKERSON; M. J. BAMSHAD y J. SHENDURE (2010) "Massively parallel sequencing and rare disease". *Human Molecular Genetics*, 19(R2).

NG, S. B.; E. H. TURNER; P. D. ROBERTSON; S. D. FLYGARE; A. W. BIGHAM; C. LEE; T. SHAFFER; M. WONG; A. BHATTACHARJEE; E. E. EICHLER; et al. (2009) "Targeted capture and massively parallel sequencing of 12 human exomes". *Nature*, 461(7261), pp. 272-276.

NIELSEN, C.; D. HANSEN; S. HUSBY; B. B. JACOBSEN y S. T. LILLEVANG (2003) "Association of a putative regulatory polymorphism in the PD-1 gene with susceptibility to type 1 diabetes". *Tissue Antigens*, 62(6), pp. 492-497.

NOTARANGELO, L. D.; M. C. PEITSCH; T. G. ABRAHAMSEN; C. BACHELOT; P. BORDIGONI; A. J. CANT; H. CHAPEL; M. CLEMENTI; S. DEACOCK; G. DE SAINT BASILE; et al. (1996) "CD40Lbase: A database of CD40L gene mutations causing X-linked hyper-IgM syndrome". *Immunology Today*, pp. 511-516.

PESSACH, I. M. y L. D. NOTARANGELO (2009) "X-linked primary immunodeficiencies as a bridge to better understanding X-chromosome related autoimmunity". *Journal of Autoimmunity*, pp. 17-24.

PIERCE, B. A. (2009) *Genética: Un enfoque conceptual*, Editorial Médica Panamericana. pp. 3.5.

PROKUNINA, L.; C. CASTILLEJO-LÓPEZ; F. ÖBERG; I. GUNNARSSON; L. BERG; V. MAGNUSSON; A. J. BROOKES; D. TENTLER; H. KRISTJANSDÓTTIR; G. GRÓNDAL; et al. (2002) "A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans". *Nature Genetics*, 32(4), pp. 666-669.

RAMENSKY, V.; P. BORK y S. SUNYAEV (2002) "Human non-synonymous SNPs: Server and survey". *Nucleic Acids Research*, 30(17), pp. 3894-3900.

RECHTIEN, A. y M. ALTFELD (2019) "Sexual dimorphism in HIV-1 infection". *Seminars in Immunopathology*, pp. 195-202.

ROACH, J. C.; G. GLUSMAN; A. F. A. SMIT; C. D. HUFF; R. HUBLEY; P. T. SHANNON; L. ROWEN; K. P. PANT; N. GOODMAN; M. BAMSHAD; et al. (2010) "Analysis of genetic inheritance in a family quartet by whole-genome sequencing". *Science*, 328(5978), pp. 636-639.

ROSADO, S.; G. PEREZ-CHACON; S. MELLOR-PITA; I. SANCHEZ-VEGAZO; C. BELLAS-MENENDEZ; M. J. CITORES; I. LOSADA-FERNANDEZ; T. MARTIN-DONAIRE; N. REBOLLEDA y P. PEREZ-ACIEGO (2008) "Expression of human leukocyte antigen-G in systemic lupus erythematosus". *Human Immunology*, 69(1), pp. 9-15.

SABETI, P.; S. USEN; S. FARHADIAN; M. JALLOW; T. DOHERTY; M. NEWPORT; M. PINDER; R. WARD y D. KWIATKOWSKI (2002) "CD40L association with protection from severe malaria". *Genes and Immunity*, 3(5), pp. 286-291.

SALZER, U.; T. HAGENA; D. B. WEBSTER y B. GRIMBACHER (2008) "Sequence analysis of BIRC4/XIAP in male patients with common variable immunodeficiency". *International Archives of Allergy and Immunology*, 147(2), pp. 147-151.

SANCHEZ, E.; J. L. CALLEJAS-RUBIO; J. M. SABIO; M. A. GONZALEZ-GAY; J. JIMENEZ-ALONSO; L. MICÓ; E. DE RAMÓN; M. CAMPS; A. SUAREZ; C. GUTIERREZ; et al. (2009) "Investigation of TLR5 and TLR7 as candidate genes for susceptibility to systemic lupus erythematosus". *Clinical and Experimental Rheumatology*, . v. 27, p.267-271, pp. 267-271.

SAWALHA, A. H.; J. B. HARLEY y R. H. SCOFIELD (2009) "Autoimmunity and Klinefelter's syndrome: When men have two X chromosomes". *Journal of Autoimmunity*, pp. 31-34.

SCHOTT, E.; H. WITT; K. NEUMANN; S. TAUBE; D. Y. OH; E. SCHREIER; S. VIERICH; G. PUHL; A. BERGK; J. HALANGK; et al. (2007) "A Toll-like receptor 7 single nucleotide polymorphism protects from advanced inflammation and fibrosis in male patients with chronic HCV-infection". *Journal of Hepatology*, 47(2), pp. 203-211.

SCHURZ, H.; M. SALIE; G. TROMP; E. G. HOAL; C. J. KINNEAR y M. MÖLLER (2019) "The X chromosome and sex-specific effects in infectious disease susceptibility". *Human genomics*, pp. 2.

SCHUURHOF, A.; L. BONT; C. L. E. SIEZEN; H. HODEMAEKERS; H. C. VAN HOUWELINGEN; T. G. KIMMAN; B. HOEBEE; J. L. L. KIMPEN y R. JANSSEN (2010) "Interleukin-9 polymorphism in infants with respiratory syncytial virus infection: An opposite effect in boys and girls". *Pediatric Pulmonology*, 45(6), pp. 608-613.

SCULLY, E. P.; J. HAVERFIELD; R. L. URSIN; C. TANNENBAUM y S. L. KLEIN (2020) "Considering how biological sex impacts immune responses and COVID-19 outcomes". *Nature Reviews Immunology*, 20(7), pp. 442-447.

SHERRY, S. T.; M. H. WARD; M. KHOLODOV; J. BAKER; L. PHAN; E. M. SMIGIELSKI y K. SIROTKIN (2001) "DbSNP: The NCBI database of genetic variation". *Nucleic Acids Research*, 29(1), pp. 308-311.

SIEZEN, C. L. E.; L. BONT; H. M. HODEMAEKERS; M. J. ERMERS; G. DOORNBOS; R. VANT SLOT; C. WIJMENGA; H. C. V. HOUWELINGEN; J. L. L. KIMPEN; T. G. KIMMAN; et al. (2009) "Genetic susceptibility to respiratory syncytial virus bronchiolitis in preterm children is associated with airway remodeling genes and innate immune genes". *Pediatric Infectious Disease Journal*, 28(4), pp. 333-335.

SILVA, L. K.; R. E. BLANTON; A. R. PARRADO; P. S. MELO; V. G. MORATO; E. A. G. REIS; J. P. DIAS; J. M. CASTRO; P. F. C. VASCONCELOS; K. A. B. GODDARD; et al. (2010) "Dengue hemorrhagic fever is associated with polymorphisms in JAK1". *European Journal of Human Genetics*, 18(11), pp. 1221-1227.

SOUSSI, T.; B. LEROY; M. DEVIR y S. ROSENBERG (2019) "High prevalence of cancer-associated TP53 variants in the gnomAD database: A word of caution concerning the use of variant filtering". *Human Mutation*, 40(5), pp. 516-524.

SOUYRIS, M.; J. E. MEJÍA; J. CHAUMEIL y J. C. GUÉRY (2019) "Female predisposition to TLR7-driven autoimmunity: gene dosage and the escape from X chromosome inactivation". *Seminars in Immunopathology*, pp. 153-164.

STREISINGER, G.; Y. OKADA; J. EMRICH; J. NEWTON; A. TSUGITA; E. TERZAGHI y M. INOUE (1966) "Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday.". *Cold Spring Harbor symposia on quantitative biology*, 31, pp. 77-84.

STRUNZ, T.; F. GRASSMANN; J. GAYÁN; S. NAHKURI; D. SOUZA-COSTA; C. MAUGEAIS; S. FAUSER; E. NOGOCEKE y B. H. F. WEBER (2018) "A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver". *Scientific Reports*, 8(1), pp. 1-11.

SUDMANT, P.; T. RAUSCH; E. GARDNER; R. H.- NATURE y U. 2015 (2015) "An integrated map of structural variation in 2,504 human genomes". *nature.com*, 526(7571), pp. 75-81.

SUTTNER, K.; M. DEPNER; M. WETZKE; N. KLOPP; E. VON MUTIUS; T. ILLIG; T. SPARWASSER y M. KABESCH (2010) "Genetic variants harbored in the forkhead box protein 3 locus increase hay fever risk". *Journal of Allergy and Clinical Immunology*, pp. 1395-1399.

SUZUKI, A.; R. YAMADA; X. CHANG; S. TOKUHIRO; T. SAWADA; M. SUZUKI; M. NAGASAKI; M. NAKAYAMA-HAMADA; R. KAWAIDA; M. ONO; et al. (2003) "Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis". *Nature Genetics*, 34(4), pp. 395-402.

VASER, R.; S. ADUSUMALLI; S. N. LENG; M. SIKIC y P. C. NG (2016) "SIFT missense predictions for genomes". *Nature Protocols*, 11(1), pp. 1-9.

VELEZ, D. R.; S. FORTUNATO; P. THORSEN; S. J. LOMBARDI; S. M. WILLIAMS y R. MENON (2009) "Spontaneous preterm birth in African Americans is associated with infection and inflammatory response gene variants". *American Journal of Obstetrics and Gynecology*, 200(2), pp. 209.e1-209.

VISSERS, L. E. L. M.; J. DE LIGT; C. GILISSEN; I. JANSSEN; M. STEEHOUWER; P. DE VRIES; B. VAN LIER; P. ARTS; N. WIESKAMP; M. DEL ROSARIO; et al. (2010) "A de novo paradigm for mental retardation". *Nature Genetics*, 42(12), pp. 1109-1112.

VOIGHT, B.; H. KANG; J. DING; C. PALMER; ... C. S.-PL. y U. 2013 (2013) "Correction: the metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits". *PLoS Genetics*, 9(4).

XUE, Y.; Y. CHEN; Q. AYUB; N. HUANG; E. V. BALL; M. MORT; A. D. PHILLIPS; K. SHAW; P. D. STENSON; D. N. COOPER; et al. (2012) "Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing". *American Journal of Human Genetics*, 91(6), pp. 1022-1032.